DISSERTATION

# CYP2D6 POLYMORPHISMS AND PESTICIDE EXPOSURE IN PARKINSON'S DISEASE

Submitted by

Susan H. Hariri

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring, 2003

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 3092674

UMI®

UMI Microform 3092674

Copyright 2003 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

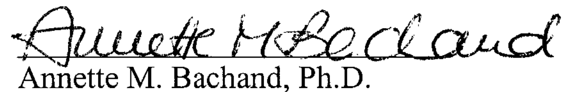
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

COLORADO STATE UNIVERSITY

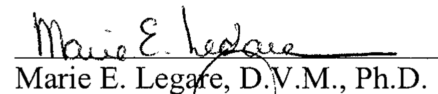
April 9, 2003

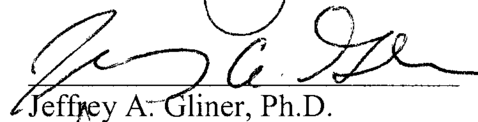
WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY SUSAN H. HARIRI ENTITLED "CYP2D6 POLYMORPHISMS AND PESTICIDE EXPOSURE IN PARKINSON'S DISEASE" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

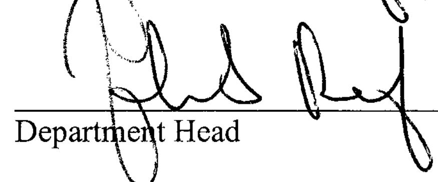

Annette M. Bachand, Ph.D.

Thomas J. Keefe, Ph.D.


Marie E. Legare, D.V.M., Ph.D.


Jeffrey A. Gliner, Ph.D.

Advisor – John S. Reif, D.V.M., M.Sc.


Department Head



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## ABSTRACT OF DISSERATION

## CYP2D6 POLYMORPHISMS AND PESTICIDE EXPOSURE IN PARKINSON'S DISEASE

The objective of this dissertation was to develop and examine new methods to estimate gene-environment interaction (GEI) effects in disease etiology using Parkinson's disease (PD) as a model. We conducted a case-only study of GEI effects in PD, evaluated the reproducibility of a published CYP2D6 assay and designed a novel CYP2D6 assay for use with buccal DNA, and, designed a simulation study to determine the relative accuracy of case-only and case-control risk estimates of GEI.

In a case-only study of 21 Caucasian PD patients, we examined interaction effects of pesticide exposure and functional polymorphisms in the CYP2D6 gene associated with defective pesticide metabolism in 7-10% of Caucasians. Exposure was ascertained via a questionnaire administered in personal interviews, at which time self-collected DNA samples were obtained from patients using buccal swabs. Several subjects reported pesticide exposure; however, no CYP2D6 poor metabolizers (PM) were identified, thus GEI evaluation was not possible. However, using two historical control groups with characteristics similar to our subjects, we found a statistically significant association between gardening and PD (OR=5.83, 95% CI 1.70, 21.93) and an increased but not statistically significant risk for ever versus never pesticide exposure ($OR_{combined}$ = 1.52, 95% CI 0.50, 4.46; $OR_{ControlGroup1}$=1.39, 95% CI 0.43, 4.33; $OR_{ControlGroup2}$=1.67, 95% CI 0.51, 5.24).

We demonstrated failure of a published polymerase chain reaction (PCR) assay for simultaneous detection of five most common CYP2D6 PM alleles in Caucasians to



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

produce results using a non-blood source of DNA and designed a new PCR assay for use with buccal cells to simultaneously detect CYP2D6*4 and *5, the two most common PM alleles in Caucasians.

We compared accuracy of case-only and case-control risk estimates of Type 2 GEI effects in a simulation study using populations with known risks. Case-only estimates were shown to approximate known risks more consistently and accurately under most conditions provided gene and exposure of interest are independent in the population. We demonstrated the influence of sample size on accuracy of both estimates, and showed that inadequate study size can lead to such large random error as to result in inaccurate estimates despite efforts to eliminate bias.

Susan Hariri
Environmental and Radiological Health Sciences
Colorado State University
Fort Collins, CO 80523
Spring 2003



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## ACKNOWLEDGEMENTS

I wish to gratefully acknowledge the numerous individuals who directly and indirectly contributed to shaping this body of work. I thank Dr. John Reif for his mentorship and assistance throughout the frequently shifting course of my project, and other members of my committee for providing invaluable guidance at various stages of my dissertation. I am particularly grateful to Drs. Marie Legare, Annette Bachand, and Thomas Keefe for their expert direction, contribution, assistance, and support during the completion of my dissertation.

I thank Dr. Bill Hanneman for graciously allowing me full access to his laboratory and for his generous contribution to supplies whenever necessary. I am grateful to Dr. Marie Legare, from whom I received laboratory training, for her friendship and good humor as I blundered along the way. I thank Dr. Annette Bachand for providing the conceptual framework for the simulation part of this dissertation. I would also like to thank her for her impeccable statistical guidance, constant encouragement, valuable friendship, genuine empathy, and tough criticism, all of which helped me to be a better scientist, and, more importantly, to believe in myself. I thank Dr. Keefe for his gentle support and patience while he imparted expert statistical advice and guidance. I equally appreciate his friendship and kind personal advice. I would also like gratefully acknowledge the countless hours of scientific and emotional support I received from my fellow students. A special thanks to Alison Hinckley whose friendship and eternal optimism allowed me to regain enough perspective to endure many difficult situations.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Equally valuable to the guidance of those mentioned above was the constant emotional support of my family. I thank my parents, Reza Hariri and Mahvash Shahegh Hariri, without whose encouragement and support none of this would have been possible. In particular, I am indebted to my husband, Chris Bennett, in whose unwavering love and support, prodigious computer expertise, and peaceful presence I could always take shelter.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## TABLE OF CONTENTS

Abstract of Disseration ..... iii
Acknowledgements ..... v
TABLE OF CONTENTS ..... vii
List of Tables ..... xi
List of Figures ..... xvi
Chapter 1 ..... 1
1.1 Background and objectives ..... 1
1.2 Gene-environment interactions ..... 3
1.3 Gene-environment interactions in the etiology of Parkinson's disease ..... 3
1.4 Comparison of case-control and case-only designs to estimate gene-environment interactions ..... 5
1.5 Dissertation overview ..... 6
Chapter 2 ..... 9
2.1 Parkinson's disease ..... 9
2.1.1 Disease overview ..... 9
2.1.2 Clinical features ..... 10
2.1.3 Diagnosis ..... 10
2.1.4 Mechanisms of disease ..... 11
2.1.4.1 Oxidative stress ..... 11
2.1.4.2 Mitochondrial dysfunction ..... 12
2.1.5 Descriptive epidemiology ..... 13
2.1.5.1 Incidence and prevalence ..... 13
2.1.5.2 Geographical distribution ..... 14
2.1.5.3 Temporal variation ..... 15
2.1.6 Endogenous risk factors ..... 15
2.1.6.1 Age ..... 16
2.1.6.2 Race and Gender ..... 16
2.1.6.3 Genetic risk factors ..... 16
2.1.6.3.1 Twin studies ..... 17
2.1.6.3.2 Family studies ..... 17
2.1.6.3.2.1 Alpha-synuclein ..... 17
2.1.6.3.2.2 Parkin gene ..... 18
2.1.6.3.3 Susceptibility genes (detoxification enzymes) ..... 19
2.1.6.3.3.1 Monoamine oxidase B (MAO-B) ..... 19
2.1.6.3.3.2 Glutathione-S-transferase (GST) ..... 20
2.1.6.3.3.3 Cytochrome P450 enzymes ..... 20
2.1.6.3.3.3.1 Cytochrome P450 2D6 (CYP2D6) ..... 21
2.1.7 Exogenous (environmental) risk factors ..... 23
2.1.7.1 Infectious agents ..... 23
2.1.7.2 Head trauma ..... 24
2.1.7.3 Diet ..... 24
2.1.7.4 Smoking ..... 25
2.1.7.5 Metals ..... 25
2.1.7.6 Industrial chemicals ..... 26



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.1.7.7 Rural residence and well water consumption ........................................26
2.1.7.8 Agricultural occupation ........................................27
2.1.7.9 MPTP ........................................27
2.1.7.10 Pesticides ........................................28
2.1.7.10.1 Mechanism of action of pesticides in PD ........................................30
2.1.7.10.2 Classes of pesticides ........................................33
2.2 Molecular techniques ........................................33
2.2.1 CYP2D6 ........................................33
2.2.1.1 Background and evolution ........................................33
2.2.1.2 Structure ........................................35
2.2.1.3 Function ........................................36
2.2.1.3.1 Polymorphisms ........................................37
2.2.2 Sample/DNA collection ........................................39
2.2.2.1 Buccal cells ........................................39
2.2.2.1.1 Advantages ........................................40
2.2.2.1.2 Disadvantages ........................................40
2.2.3 Genotyping ........................................40
2.2.3.1 Polymerase chain reaction ........................................41
2.3 Gene-environment interactions ........................................43
2.3.1 Introduction ........................................43
2.3.2 Definition ........................................44
2.3.3 Biological mechanisms ........................................45
2.3.4 Models of interaction ........................................46
2.3.5 Scale of measurement ........................................48
2.3.6 Epidemiologic approaches ........................................51
2.3.6.1 Case-control studies ........................................51
2.3.6.1.1 Case-control design ........................................51
2.3.6.1.2 Sample size and power required to detect gene-environment interactions (GEI) in case-control studies ........................................52
2.3.6.1.3 Analysis of gene-environment interactions (GEI) in case-control studies ........................................54
2.3.6.1.4 Advantages of case-control studies ........................................55
2.3.6.1.5 Disadvantages of case-control studies ........................................56
2.3.6.2 Case-only studies ........................................56
2.3.6.2.1 Case-only study design ........................................57
2.3.6.2.2 Principle of the case-only design ........................................58
2.3.6.2.3 Sample size for case-only studies ........................................59
2.3.6.2.4 Analysis of gene-environment interaction (GEI) in a case-only study ........................................59
2.3.6.2.5 Advantages of case-only design ........................................59
2.3.6.2.6 Disadvantages of the case-only design ........................................60
2.3.6.3 Design considerations for studies to detect gene-environment interaction effects ........................................60
Chapter 3 ........................................63
3.1 Introduction ........................................63
3.2 Methods ........................................68



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.2.1 Study population ........ 68
3.2.2 Diagnostic criteria ........ 69
3.2.3 Subject recruitment ........ 69
3.2.4 Subject interview ........ 70
3.2.4.1 Personal information ........ 71
3.2.4.2 Occupational information ........ 72
3.2.4.2.1 Pesticide exposure ........ 72
3.2.4.3 DNA sample collection ........ 73
3.2.5 Genotyping ........ 73
3.3 Results and analysis ........ 73
3.4 Discussion ........ 77
3.4.1 Study design ........ 79
3.4.2 Study size and power ........ 80
3.4.3 Case recruitment ........ 82
3.4.4 Exposure assessment ........ 82
3.5 Summary and future research direction ........ 83
Chapter 4 ........ 88
4.1 Introduction ........ 88
4.2 Subjects, materials, and methods ........ 91
4.2.1 Subjects ........ 91
4.2.2 DNA Source ........ 91
4.2.3 Buccal DNA Extraction ........ 92
4.2.4 Buccal DNA concentration ........ 92
4.2.5 CYP2D6 Genotype determination ........ 92
4.2.5.1 Long distance and multiplex polymerase chain reaction ........ 92
4.2.5.1.1 Preamplification ........ 92
4.2.5.1.2 Multiplex allele-specific PCR ........ 94
4.2.5.1.3 Results ........ 94
4.2.5.2 Long PCR of CYP2D6 fragment containing *4 mutation ........ 95
4.2.5.3 CYP2D6 *4 PCR assay ........ 96
4.2.5.4 Results ........ 96
4.2.5.4.1 Validation and DNA Sequencing ........ 97
4.2.5.4.2 Sequencing Results ........ 98
4.3 Discussion ........ 101
Chapter 5 ........ 106
5.1 Introduction ........ 106
5.2 Methods ........ 109
5.2.1 Construction of populations ........ 111
5.2.2 Sampling from each population ........ 112
5.2.3 Determination of relative contribution of population parameters to accuracy of GEI estimates ........ 113
5.2.4 Analysis ........ 114
5.2.4.1 Case-control estimates ........ 114
5.2.4.3 Case-only estimate ........ 115
5.2.4.3 Control-only estimate ........ 116
5.3 Results ........ 116



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

5.3.1 Evaluation and comparison of accuracy of case-only and case-control interaction estimates ........ 118
5.3.1.1 Evaluation and comparison of case-only and case-control study designs given weak interaction effect ........ 120
5.3.1.2 Relative contribution of population parameters to accuracy of GEI estimates ........ 121
5.3.1.3 Effect of violation of the independence and Type 2 assumptions in the sample on accuracy of case-only and case-control estimates ........ 125
5.3.1.4 Effect of violation of the independence assumption in the population on accuracy of case-only and case-control estimates ........ 132
5.3.2 Odds ratio for exposure alone ........ 133
5.4 Discussion ........ 134
5.4.1 Strengths of the simulation study ........ 135
5.4.2 Limitations of the simulation study ........ 137
5.5 Conclusions ........ 138
Chapter 6 ........ 185
Appendices ........ 211
Appendix A Consent Form ........ 212
Appendix B Questionnaire ........ 215
Appendix C Relative Accuracy of Case-only and Case-control Estimates of Gene-Environment Interactions ........ 253



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## LIST OF TABLES

Table 2-1. Two by four table for gene-environment interaction analysis in a case-control study. ........ 58

Table 3-1. Demographic characteristics of the study population. ........ 85

Table 3-2. Summary of risk factors in study population. ........ 86

Table 3-3. Genotype analysis and phenotype prediction in study population. ........ 87

Table 3-4. Crude odds ratios of PD risk by selected exposures using historical controls for comparison. ........ 87

Table 4-2. Oligonucleotide primer sequences with position numbers according to Kimura et al. (1989). All primers are located in CYP2D6 with (-) and (+) indicating positions in the 5' and 3' flanking regions, respectively ........ 104

Table 4-3. Newly designed oligonucleotide primer sequences with position numbers according to Kimura et al. (1989) within the CYP2D6 gene ........ 104

Table 4-3. Genotype analysis and phenotype prediction for DNA samples presented in Figures 4-1 to 4-3 ........ 105

Table 5-1. Summary of input parameters for each hypothetical population. ........ 140

Table 5-2. Relative standard errors of estimates for percentage of estimates within 10, 30, and 50% ranges of the "true" odds ratios for the Typical population based on 1000 sample replicates in the simulation study (SIM=3 or COR=3). ........ 141

Table 5-3. Replicates with missing frequencies. ........ 142

Table 5-4. Summary of the simulation study results for gene-environment interaction with respect to case-only and case-control designs in the Typical Population: Percentage of estimated SIM or COR values that fall within specified ranges of "true" SIM or COR value (SIM =3 or COR = 3). ........ 143

Table 5-5. Summary of the simulation study results for gene-environment interaction with respect to case-only and case-control designs in the Typical Population: Percentage of estimated averages from exponentiated coefficients [ln(SIM) or ln(COR)] that fall within specified ranges of "true" exponentiated coefficient of SIM or COR (beta=1.10). ........ 144

Table 5-6. Summary of the simulation study results for gene-environment interaction with respect to case-only and case-control designs in Population 1: Percentage of estimated SIM or COR values that fall within specified ranges of "true" SIM or COR value (SIM =1.5 or COR = 1.5) ........ 145



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5-7. Summary of the simulation study results for gene-environment interaction with respect to case-only and case-control designs in Population 32: Percentage of estimated SIM or COR values that fall within specified ranges of "true" SIM or COR value (SIM =6 or COR = 6). 146

Table 5-8. Summary of the simulation study results for testing the independence assumption (COOR) in Typical Population: Percentage of estimated OR values that fall within specified ranges of "true" COOR (COOR=1). 147

Table 5-9. Summary of the simulation study results for testing the independence assumption (COOR) in Population 1: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1). 147

Table 5-10. Summary of the simulation study results for testing the independence assumption (COOR) in Population 32: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1). 148

Table 5-11. Summary of the simulation study results for testing Type 2 gene-environment interaction assumption ($OR_{Ge}$) in Typical Population: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1). 148

Table 5-12. Summary of the simulation study results for testing Type 2 gene-environment interaction assumption ($OR_{Ge}$) in Population 1: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1). 149

Table 5-13. Summary of the simulation study results for testing Type 2 gene-environment interaction assumption ($OR_{Ge}$) in Population 32: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1). 149

Table 5-14. Summary of the simulation study results for effects of exposure alone ($OR_{Eg}$) in Typical Population: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=2). 150

Table 5-15. Summary of the simulation study results for effects of exposure alone ($OR_{Eg}$) in Population 1: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=2). 150

Table 5-16. Summary of the simulation study results for effects of exposure alone ($OR_{Eg}$) in Population 32: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=2). 151

Table 5-17. Summary of the simulation study results for effects of multiplicative interaction ($OR_{GE}$) in Typical Population: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=6). 151

Table 5-18. Summary of the simulation study results for effects of multiplicative interaction ($OR_{GE}$) in Population 1: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=3). 152



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5-19. Summary of the simulation study results for effects of multiplicative interaction ($OR_{GE}$) in Population 32: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=12). ........ 152

Table 5-20. Summary of the simulation study results for gene-environment interaction with respect to case-only and case-control designs in the Population 33: Percentage of estimated SIM or COR values that fall within specified ranges of "true" SIM or COR value (SIM=1.2 or COR=1.2). ........ 153

Table 5-21. Summary of the simulation study results for gene-environment interaction with respect to case-only and case-control designs in the Population 33: Percentage of estimated averages from exponentiated coefficients [ln(SIM) or ln(COR)] that fall within specified ranges of "true" exponentiated coefficient of SIM or COR (beta=.18). ........ 154

Table 5-22. Summary of the simulation study results for effects of multiplicative interaction ($OR_{GE}$) in Population33: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1.5). ........ 155

Table 5-23. Summary of the simulation study results for effects of multiplicative interaction ($OR_{GE}$) in Population 33: Percentage of exponentiated coefficients [ln($OR_{GE}$)] that fall within specified ranges of "true" exponentiated coefficient (beta=0.41). ........ 155

Table 5-24. Summary of the simulation study results for testing the independence assumption (COOR) in Population 33: Percentage of estimated OR values that fall within specified ranges of "true" COOR (COOR=1). ........ 156

Table 5-25. Summary of the simulation study results for testing Type 2 gene-environment interaction assumption ($OR_{Ge}$) in Population 33: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1). ........ 156

Table 5-26. Summary of the simulation study results for effects of exposure alone ($OR_{Eg}$) in Population 33: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=2). ........ 157

Table 5-27. Relative contribution of design parameters, number of subjects, study type, and independence and Type2 assumptions (and all 1 order interactions) to the accuracyof Type 2 gene-environment interaction estimates........ 158

Table 5-28. Relative contribution of design parameters, number of subjects, and the independence and Type2 assumptions (and 1 order interactions) to the accuracyof case-only estimates (COR) of Type 2 gene-environment interaction estimates...... 159

Table 5-29. Relative contribution of design parameters, number of subjects, and the independence and Type2 assumptions (and 1 order interactions) to the accuracyof case-control estimates (SIM) of Type 2 gene-environment interaction estimates... 160



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5-30. Relative contribution of design parameters, number of subjects, and the independence and Type2 assumptions (and all 1 order interactions) to the accuracyof case-control estimates (COR) of Type 2 gene-environment interaction estimates. 161

Table 5-31. Relative contribution of design parameters, number of subjects, and the independence and Type2 assumptions (and all 1 order interactions) to the accuracyof case-control estimates (SIM) of Type 2 gene-environment interaction estimates... 164

Table 5-32. Summary of the simulation study results for the Typical population as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" COR and number of estimates within categories of the independence and Type 2 assumptions. .......... 167

Table 5-33. Summary of the simulation study results for the Typical population as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" SIM and number of estimates within categories of the independence and Type 2 assumptions .......... 169

Table 5-34. Summary of the simulation study results for Population 1 as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" COR and number of estimates within categories of the independence and Type 2 assumptions. .......... 171

Table 5-35. Summary of the simulation study results for Population 1 as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" SIM and number of estimates within categories of the independence and Type 2 assumptions .......... 173

Table 5-36. Summary of the simulation study results for Population 32 as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" COR and number of estimates within categories of the independence and Type 2 assumptions. .......... 175

Table 5-37. Summary of the simulation study results for Population 32 as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" SIM and number of estimates within categories of the independence and Type 2 assumptions .......... 177

Table 5-38. Summary of the simulation study results for gene-environment interaction with respect to case-only and case-control designs in the Population 34: Percentage of estimated SIM or COR values that fall within specified ranges of "true" SIM or COR value (SIM=3 or COR=3) .......... 179

Table 5-39. Summary of the simulation study results for gene-environment interaction with respect to case-only and case-control designs in the Population 35: Percentage of estimated SIM or COR values that fall within specified ranges of "true" SIM or COR value (SIM=3 or COR=3) .......... 180



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5-40. Summary of the simulation study results for testing the independence assumption (COOR) in Population 34: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1). ........................................................ 181

Table 5-41. Summary of the simulation study results for testing the independence assumption (COOR) in Population 35: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1). ........................................................ 181



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## LIST OF FIGURES

Figure 2-1. Chromosome 22. ....................................................................................35

Figure 2-2. CYP2D6 exon positions.....................................................................................36

Figure 2-3. CYP2D6 polymorphism map..............................................................................38

Figure 4-1. Allele-specific PCR analysis of CYP2D6 mutation *4 (B) using the wild-type primer. Lane designations are as follows: "M" indicates a 100 bp molecular weight marker (Gibco BRL); numerals indicate sample ID as presented in Table 4-3; "+/-" indicate positive and negative control, respectively. ................................................99

Figure 4-2. Allele-specific PCR analysis of CYP2D6 mutation *4 (B) using the mutant primer. Lane designations are as follows: "M" indicates a 100 bp molecular weight marker (Gibco BRL); numerals indicate sample ID as presented in Table 4-3; "+/-" indicate positive and negative control, respectively. .............................................100

Figure 4-3. Allele specific PCR analysis of CYP2D6 mutation*4 (B) using wild-type (left lanes) and mutant (right lanes) primers for each sample. A 100 bp molecular weight marker (Gibco BRL) is presented in Lane 1. Lanes 10 and 11 show a positive and a negative control, respectively. .......................................................100

Figure 5-1. Effect of sample size on accuracy of case-only and case-control estimates from the Typical population ( a: n=100; b: n=300; c: n=700). Numerals with lines indicate percent of 1000 replicates included (fitted logistic model converged or OR < 15). .........................................................................................................182

Figure 5-2. Effect of sample size on accuracy of case-only and case-control estimates from the Typical population when the independence assumption is violated by a magnitude of 1.5 ( a: n=100; b: n=300; c: n=700). Numerals with lines indicate percent of 1000 replicates included (fitted logistic model converged or OR < 15). .........................................................................................................183

Figure 5-3. Effect of sample size on accuracy of case-only and case-control estimates from the Typical population when the independence assumption is violated by a magnitude of 3 ( a: n=100; b: n=300; c: n=700). Numerals with lines indicate percent of 1000 replicates included (fitted logistic model converged or OR < 15). .........................................................................................................184



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER 1

# INTRODUCTION

## 1.1 Background and objectives

Disease etiology is increasingly recognized as a complex process involving multiple causes. In particular, the development of chronic diseases depends not on one identifiable source alone but the interaction of multiple exogenous and endogenous factors.[1-5] Mounting evidence in favor of the multifactorial explanation of disease etiology combined with explosive advances in the field of molecular biology has recently shifted epidemiologic research toward incorporation of genetic assays into studies of disease.[6-10] Moreover, growing interest in determining the relative and combined contribution of genetics and environment in the pathogenesis of most chronic conditions along with the ability to investigate genetic contributions of disease has led to the creation of a new hybrid field, genetic epidemiology.[3,5,11] Although a burgeoning area of research, genetic epidemiology is still a nascent discipline with few well-designed studies to support the gene-environment interaction hypothesis with respect to chronic diseases.[5] More studies are needed in order to enhance understanding of the influence of interactions between suspected risk factors in disease causation and to help target preventive and therapeutic interventions.

One condition that exemplifies the multifactorial etiology of chronic diseases, and thus is a good prototype for improving such studies, is Parkinson's disease (PD).[12-16] Nearly two centuries of exhaustive research into a cause for the disease has implicated various exogenous and endogenous risk factors for PD.[13-15,17,18] However, most investigations have been divided along environmental and genetic lines.[15] Moreover,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

results from epidemiologic studies in each area are often contradictory, leading to ambiguous interpretations.[13,19] Today, the etiology of this increasingly common and debilitating disease of middle and late life continues to vex scientists despite successful elucidation of the pathophysiologic mechanism of the disease.[16]

Although inconclusive, evidence from years of environmentally and genetically focused studies have encouraged research direction toward a more integrated paradigm for the study of PD etiology.[14-18,20] Taken together, the results suggest that PD results from modulation of a genetic predisposition by certain environmental toxins which can then affect the onset of the disease.[17] Yet despite growing interest in the impact of interactions between genetic factors and environmental chemicals in the occurrence of Parkinson's disease, there remains a paucity of research in this area.

The overall goal of this dissertation was to develop and examine methods by which to estimate separate as well as joint contributions of genetic and environmental risk factors to further the understanding of gene-environment interaction effects in the development of complex diseases in general and Parkinson's disease in particular. More specifically, we were interested in developing an epidemiologic study to examine the effects of pesticide exposure (a highly suspected risk factor in PD) in conjunction with a defect in CYP2D6 (a gene responsible for the elimination of pesticides from the body) on the development of PD. To that end, the specific objectives of the project included 1) development of a rapid and accurate molecular assay to detect the gene defect of interest, 2) use of the case-only study design to detect effects of interaction between the defective genotype and pesticides, and 3) evaluation of the case-control design versus the case-only design, two epidemiologic study designs commonly used to study gene-environment



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interactions. Each component is presented and discussed in separate chapters, and hypotheses are explored in detail through specific aims.

## 1.2 Gene-environment interactions

All chronic diseases have a multifactorial etiology involving genetic and environmental influences. Moreover, two or more risk factors in the same causal pathway often do not act independently but rather interact with each other to cause disease.[21] In epidemiologic studies, an interaction is defined in relation to an underlying biological mechanism and the effect is measured using statistical concepts. Interactions can involve many combinations of factors, including environment-environment and gene-gene interactions.[21] However, this study is restricted to the special case of gene-environment interactions. Described as "genetic control of sensitivity to the environment"[8], gene-environment interactions are a frequent phenomenon in the study of genes with common variants that can increase a person's susceptibility to disease in the presence of an environmental exposure. Assessment of interaction is model-dependent; therefore, different interaction models have been proposed to explain most known biological mechanisms.[6,7,10,22,23] The appropriate model for any given study is one that best reflects the biological basis of the disease and risk factors under investigation.[21] Depending on the selected model, interactions effects must be statistically evaluated either on a multiplicative or additive scale. Again, choice of scale must be made based on an explicitly defined scientific principle.[23]

## 1.3 Gene-environment interactions in the etiology of Parkinson's disease

After decades of comprehensive inquiry into potential environmental risk factors for PD, certain patterns have emerged. Epidemiologic studies of exogenous factors have



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

implicated certain occupational categories such as farming and steel and metal work, as well as certain lifestyles including rural residence and well-water consumption, in increased risk of PD.[24,25] More importantly, however a host of environmental chemicals are suspected to be associated with PD.[25-34] Of the countless suspected toxins, pesticides have been of particular interest, partly because certain pesticides exhibit the strongest and most consistent association with the disease, but more importantly, because they share a strikingly similar chemical structure with the molecule MPTP, a confirmed cause of parkinsonism in humans.[25,35-45]

Although investigations of external causes of PD have been rivaled by equally intensive searches for a possible endogenous influence, a genetic basis for the disease has also remained elusive. While family history has been repeatedly shown to be a strong predictor for an increased risk of Parkinson's disease in numerous epidemiologic studies, research on families and twin pairs with PD has yielded largely negative findings.[46-48] The apparent incongruities are suggestive of either low penetrance of causative genes or, more likely, a multifactorial etiology.[14] Thus, recent studies have focused on discovering gene mutations that may increase susceptibility to an environmental neurotoxin, particularly genes encoding for enzymes responsible for detoxification of chemicals.[15,36] Of the several suspected susceptibility genes, Cytochrome 450 debrisoquine hydroxylase (CYP2D6) is of particular interest because it is involved in the first of a multi-step process of detoxification of foreign organic compounds including pesticides and the parkinsonism-inducing MPTP.[49-53] Moreover, CYP2D6 is suspected to be involved in the metabolism and processing of neurotransmitters such as dopamine, the depletion of which is responsible for PD.[53,54]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## 1.4 Comparison of case-control and case-only designs to estimate gene-environment interactions

The case-control design is appropriate for epidemiologic studies of rare diseases with long latency periods, a category into which most chronic diseases can be placed.[21] Moreover, case-control studies allow investigation of multiple risk factors simultaneously while being more cost- and time- efficient than other epidemiologic designs. These are among the reasons that the case-control design is the most commonly used design in gene-environment interaction studies. In such studies, subjects are selected on the basis of disease status, and non-diseased or control subjects are used as a referent group against which risk estimates for cases with disease can be estimated.[55]

Selection of appropriate controls in case-control studies, a historically difficult task, becomes even more so in studying insidious diseases with long latencies between disease inception and presence of identifiable signs. To alleviate this problem, alternative study designs based on the traditional case-control model have been developed to investigate gene-environment interactions.[21] One such method is the case-only design whereby the association between an environmental exposure and a genotype is examined among case subjects only (affected subjects with a given disease).[2,9,56-58]

The case-control study and its derivative designs have been well researched and documented, both with respect to study methodology as well as to analysis. [1,22,23,55,59-74] However, since it is impossible to compare the estimated versus true risks in the absence of the latter, epidemiologic study designs are developed and evaluated based on theoretical principles. Comparing the performance of different models with respect to key assumptions using various possible scenarios can further elucidate our understanding



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of theoretically-based models, which in turn, can have a large influence on policy changes.

Simulation studies can be designed to perform such comparisons. Hypothetical populations with pre-specified parameters created using available information to resemble possible real populations can be the basis for the simulation study. "True" disease risk can be easily calculated for each entire population. Multiple samples of different sizes can then be drawn from each population and risk estimates computed using both case-control and case-only methods. Subsequently, variability in the estimates calculated for each of many repeated samples from the different populations can be evaluated. Additionally, the case-control study can be compared to the case-only method in relation to variability of estimates given different population characteristics.

### 1.5 Dissertation overview

This project was intended as an integrative approach to the study of gene-environment interactions in chronic diseases using Parkinson's disease as a model. Specifically, the dissertation involved three components, including development of a genetic assay to detect a polymorphism in the CYP2D6 gene resulting in defective metabolism of pesticides, examination of the joint effects of pesticide exposure and deficient CYP2D6 enzyme activity in the causation of PD, and selection and evaluation of an appropriate study design for evaluating the interaction of said risk factors in association with the disease.

Chapter 2 provides a detailed and comprehensive literature review covering all components of the dissertation. Chapter 3 is a summary of the pilot study to assess the



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

effects of gene-environment interaction in the pathogenesis of Parkinson's disease. The objective of this study was to investigate the following hypothesis:

- Interaction between exposure to pesticides and a mutation in the CYP450 2D6 gene resulting in poor xenobiotic metabolism increases susceptibility to Parkinson's disease.

The fourth chapter is an examination of the molecular techniques employed to determine CYP2D6 genotype of the patients studied. The hypotheses examined in this project were as follow:

- Buccal cells are an adequate source of DNA for detection of polymorphisms in the CYP2D6 gene using a multiplex polymerase chain reaction technique (PCR) to simultaneously detect mutations in the five PM-associated allele groups CYP2D6*3, *4, *6, *7, and *8.
- PCR is an accurate and reliable method for detection of single nucleotide polymorphisms (SNP) in the CYP2D6 gene.

Chapter 5 contains an evaluation of case-control and case-only designs to detect gene-environment interactions using a simulation model. In this study, we investigated the following hypotheses:

- Odds ratio estimates of gene-environment interactions obtained from case-control and case-only studies are precise measures of the actual effect of such interactions in the population
- Estimates based on the case-only study design demonstrate less variability from the true risk than those obtained from case-control studies.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The sixth and final chapter provides a synthesis and discussion of findings and recommendations for future research endeavors.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER 2

# LITERATURE REVIEW

This chapter is intended as a comprehensive review of the literature pertaining to topics discussed in the three chapters that follow. The subsequent chapters are presented in manuscript format, requiring some repetition of materials covered in this literature review.

## 2.1 Parkinson's disease

### 2.1.1 Disease overview

In 1817, Dr. James Parkinson published an essay on "the shaking palsy", describing in detail the disease that today bears his name.[75] Nearly two centuries after Parkinson's observations, the etiology of this common debilitating neurodegenerative disease remains largely unclear. Despite intensive scientific research, problems and limitations inherent in the study of Parkinson's disease (PD) have impeded progress toward elucidation of putative causes. The elusive nature of PD has sparked much scientific debate over whether it is caused by genetic or environmental factors, and research direction has vacillated between the two theories for decades. Although the exact cause of PD remains enigmatic, together, results from various studies have provided important clues toward a multifactorial etiology hypothesis.[12,76] It is largely accepted that PD is caused by a combination of age, genetics, and environmental factors, and the latest theories emphasize the synergistic effects of exogenous chemicals with genetic susceptibility traits in PD etiology.[12,76] However, despite recent advances in molecular genetic technology enabling investigation of gene-environment interactions in



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the pathogenesis of PD, no data are currently available to explain interactions between risk factors in the development of this devastating disease.

In addition to its overwhelming effects on the victims, PD exerts a heavy economic burden on society. The long duration of the illness and eventual dependence of patients to long-term care make it an extremely costly disease with an estimated annual societal cost upward of $20 billion in the United States.[19,77]

### 2.1.2 Clinical features

The disease is characterized by resting tremor, rigidity, bradykinesia, gait disturbance, and postural instability.[78] Symptoms appear only after 80% of pigmented cells in the substantia nigra and 80% of the dopamine content in the striatum are lost.[78] A cardinal histological feature of PD is the presence of inclusions known as Lewy bodies in the remaining neurons.[78] PD is a progressive disorder, with cognitive and motor deficits such as depression, dementia, and postural deformity manifesting in advanced stages.[14,79]

### 2.1.3 Diagnosis

Currently, no biomarker or diagnostic test is available to detect PD; hence diagnosis is based on clinical criteria.[13,80,81] PD is often not distinguished from parkinsonism, a general term used to describe syndromes with similar clinical and pathological features.[14,78] Lack of standardized diagnostic criteria and confirmatory tests frequently leads clinicians to confuse it with other common but benign movement disorders in the elderly.[13,80] Insidious onset and a long latency period further make diagnosis difficult, resulting in misdiagnosis and under-diagnosis of the disease.[13,29]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

### 2.1.4 Mechanisms of disease

A number of processes have been proposed as contributing factors in the development of PD, including oxidative stress, mitochondrial defects, aging, excitotoxicity, apoptosis, inflammation, and protein aggregation. To date, none of these putative mechanisms have unequivocally proven to increase susceptibility to PD. Nonetheless, two mechanisms have emerged as strong candidates and are discussed briefly below.

#### 2.1.4.1 Oxidative stress

Neurons are unable to replicate and have high metabolic activity, making them susceptible to the effects of oxidative processes. Thus, the influence of oxidative stress has been investigated in the etiology of many neurodegenerative diseases.[82] Oxidative stress was first promoted as an underlying mechanism for PD after the observation that MPTP forms oxygen free radicals, and has since been of particular interest for two reasons. First, dopamine metabolism results in the formation of hydrogen peroxide ($H_2O_2$) and superoxide radicals in the substantia nigra with the rate of formation of these compounds being proportional to dopamine turnover.[82-85] Second, neuromelanin found in pigmented neurons contains large amounts of iron produced by auto-oxidation of dopamine.[86] Hydrogen peroxide reacts with the reduced iron in the neuromelanin, generating the highly toxic hydroxyl radical.[87] Supporting this hypothesis, results of laboratory studies have shown a decrease in dopamine accompanied by parkinsonian syndrome in rats receiving intranigral iron injections.[83,84] Post-mortem studies in humans further support the role of oxidative damage by demonstrating selective damage to neurons containing the highest levels of neuromelanin in PD patients.[83] Evidence from



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

epidemiologic studies also implicates oxidative stress as a contributing cause of the disease.[83,88,89] For example, case-control studies have shown higher levels of lipid peroxidation products as well as higher nigral iron levels in the substantia nigra of PD patients than among control brains.[83] However, although these findings are strongly suggestive of the involvement of oxidative processes in the development of PD, they do not provide definitive evidence of a causal role. If involved in PD causation, oxidative stress likely acts in coordination with other environmental and possibly genetic factors.[90]

**2.1.4.2 Mitochondrial dysfunction**

Since oxidative damage and mitochondrial dysfunction are synergistic phenomena that can both result from and precipitate the other,[91] and if oxidative stress is assumed to exert an influence in PD,[84,92] then it follows that mitochondrial damage may also be involved in Parkinson's disease. In addition, the parkinsonism-inducing neurotoxin MPP+ has been shown to impair mitochondrial oxidative phospohorylation by the inhibition of complex I enzymes of the electron transport chain, leading to further suspicion of mitochondrial damage as another possible neurotoxic mechanism in the development of PD.[93] Demonstration of classic signs of PD including selective death of nigrostriatal dopaminergic neurons, Lewy body formation, and behavioral changes resembling PD in rats chronically dosed with rotenone, a high affinity mitochondrial complex I inhibitor, has recently intensified speculation about the role of mitochondrial injury in PD.[94,95] Moreover, a decrease in mitochondrial function has been observed in PD patients providing further evidence of a link.[94,95] However, whether mitochondrial dysfunction is a secondary process or a contributing cause of PD remains unknown.[94] Since mitochondrial DNA are maternally inherited, a disease process resulting from



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

mitochondrial damage would be expected to show a maternal pattern of inheritance. Yet epidemiologic studies have failed to show a maternal inheritance pattern of PD, undermining the laboratory findings.[94,96] However, given the current evidence, the significance of this process as a contributing if not sufficient cause of the disease deserves further inspection.

### 2.1.5 Descriptive epidemiology

Descriptive data on PD have been difficult to obtain due to diagnostic inaccuracies and discrepancies in case ascertainment.[13,14,27] Despite the limitations, however, certain consistent patterns have emerged as discussed in the following sections.

#### 2.1.5.1 Incidence and prevalence

Parkinson's disease is a common disorder of the brain and a growing national problem. The disease affects over a million people in the United States and is second only to Alzheimer's disease in frequency as a neurodegenerative disorder.[19,78,79] Estimates of the prevalence of PD range from 150-300/100,000 population, with an estimated 20 new cases per 100,000 persons annually.[14,97] PD is the seventh leading diagnosis for the 45-65 age group and the most frequently reported diagnosis for patients 65 and over. More people suffer from PD than multiple sclerosis, muscular dystrophy, and amyotrophic lateral sclerosis combined.[98] Given the increase in life expectancy in the general population and the growing numbers of the elderly, it is reasonable to expect a steady and significant increase in PD during the next century. In fact, neurodegenerative diseases (including PD) are predicted to exceed cancer as the second most common cause of death in the elderly by the year 2040.[78] Despite its significant impact, however, the



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

disease frequently goes unrecognized among both physicians and the general public, resulting in inaccurate and incomplete incidence and prevalence rates.[13]

### 2.1.5.2 Geographical distribution

Frequency of PD is marked by international variations, even after adjustment for many inconsistencies.[99,100] Estimates of disease prevalence range from 31 per 100,000 population in Libya to 657 per 100,000 population in Argentina [46]. Regional differences, based on studies conducted in North America, indicate a gradient from the northwest to the southeast for both the US and Canada.[101,102] Both genetic differences and exposure to different environmental risk factors are plausible explanations for the observed geographic variations.[102]

Using the assumption that Lewy bodies reflect preclinical disease, a multi-national study found similar prevalence rates in under-developed countries as in industrialized western countries, indicating a universal propensity for developing PD.[12,13] Yet, international comparisons have consistently shown PD to be less common in non-industrialized countries.[99] These findings support the role of industrialization by suggesting the involvement of exogenous toxins in the pathogenesis of PD.[13,14] Among industrialized countries, moreover, it has repeatedly been shown that rural residence is more strongly associated with risk of PD compared to urban living, suggesting that agricultural practices and/or use of agricultural chemicals in developed countries contribute to the risk of PD.[27,28,103] However, discrepancies in health care quality, access, and disease reporting across regions and countries make any such conclusions untenable.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 2.1.5.3 Temporal variation

Studies using mortality data have indicated that PD has increased in older age groups (>75) and has decreased in younger age groups (<65), suggesting a temporal variation.[13] However, data from other similar studies have failed to show any change in temporal trends in recent decades.[13] Therefore, no conclusions can be drawn with respect to variation in incidence over time.[20] More studies of large cohorts over long periods of time would be necessary to confirm observations based on mortality information alone however. Moreover, prospective studies would provide other very important clues to the etiology of PD, particularly with respect to the influence of environmental factors. Unfortunately, difficulties inherent in the study of PD have made conduct of longitudinal investigations largely infeasible. Hence, variability in incidence and prevalence of PD over time is poorly understood. Information obtained from a 50-year (1935-1988) population-based study of incidence of PD in Olmstead, Minnesota suggests an increase in incidence from 9.2 per 100,000 persons annually from 1935-1944 to 16.3 per 100,000 persons from 1975-1984.[20] In this investigation, effort was made to minimize variability in diagnostic criteria by assigning a single neurologist to classify cases. However, while standardization of case selection did serve to strengthen the results of the study, it was unable to address uncertainties with respect to improvements in record keeping or other competing disease risks. More importantly, the question of whether PD incidence has differed over longer periods of time remains unresolved.

### 2.1.6 Endogenous risk factors

PD is largely recognized as an idiopathic disease; however, less common variants exist that should be differentiated.[12,15] Familial PD accounts for approximately one third



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of cases and can follow an autosomal dominant inheritance pattern.[15] Young-onset PD is independent of age and also appears to be related to hereditary factors.[15]

#### 2.1.6.1 Age

PD is a chronic, debilitating, and age-related disease.[78] Occurrence is rare before 40, increases sharply after 50 and continues to rise for the remainder of life.[14,104] However, although age is the single most consistent factor related to PD, normal age-related neuronal damage is insufficient to explain development of the disease.[12,13]

#### 2.1.6.2 Race and Gender

The disease has a predilection for the male gender, in whom it occurs 20-30% more frequently than in females.[13] Various hypotheses have been proposed to explain the apparent discrepancy in the frequency of PD between genders, including a protective role of female sex hormones, and greater male exposure to extrinsic risk factors such as industrial toxicants.[18] Population-based studies have not revealed any significant associations with race, ethnic origin, social class, or education level internationally.[13] However, in the US, prevalence and mortality are slightly higher among whites than in African Americans.[105]

#### 2.1.6.3 Genetic risk factors

Interest into studies of a genetic basis for PD has waxed and waned over time, and historically, less significance has been placed on the influence of genetic risk factors in the etiology of PD than environmental factors.[77] However, increasingly sophisticated molecular techniques have renewed research interest toward genetic factors, with many studies suggesting heredity as a determinant of PD. Familial clustering of PD has been observed in some large kindred, suggesting an autosomal dominant pattern of inheritance



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and providing evidence for an endogenous cause.[12-15] Genetic studies have confirmed that certain variants of PD, as well as some causes of parkinsonism are inherited.[106,107] However, only a small fraction of cases (10% or less) represent the hereditary form.[14] Overall, genetic study results are mixed, and genetic factors alone are inconclusive evidence for the causal mechanism of PD.[12,76,108]

**2.1.6.3.1 Twin studies**

Several twin studies have been conducted to further examine the role of heredity in causing idiopathic PD.[15,19,47,109,110] The objective of these studies was to determine whether concordance between monozygotic and dizygotic twins is the same or different.[15,18] Similar concordance in both sets would argue against a genetic influence. Almost all studies, including a large one of 31,000 WWII veteran twins found concordance to be virtually identical in typical sporadic cases (onset >50) while different among twins with young onset disease.[29,47,103,109,110] These results confirm the role of heredity in rare young onset cases while providing strong evidence against a hereditary component among typical cases.

**2.1.6.3.2 Family studies**

Although not all of the familial cases follow a Mendelian pattern of inheritance, an autosomal dominant pattern has been identified among several large families with PD.[12,14]

**2.1.6.3.2.1 Alpha-synuclein**

Extensive studies of one large Italian and three smaller Greek families with the autosomal dominant form of PD led to the discovery of the first gene responsible for this rare form of PD in 1997.[12,14] The gene was localized to chromosome 4q21-q23, and the



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

specific mutation, a single base pair change at position 209 of the gene encoding for a brain protein called alpha-synuclein was later identified.[12,14] Although the gene mutation was associated with an atypical form of PD characterized by younger age at onset, the findings stimulated research into genetic causes of the more typical sporadic form. Subsequent investigations in both familial and sporadic cases using PCR techniques failed to identify the alpha-synuclein gene in even a single patient, however, suggesting that the mutation rarely, if ever, is responsible for typical (idiopathic) PD.[12,14,15]

#### 2.1.6.3.2.2 Parkin gene

More recently, investigations of several families with a rare autosomal recessive early-onset form of PD has led to the discovery of a genetic mutation in the parkin gene on the long arm of chromosome 6.[111] Further studies of patients with the more common sporadic form of PD also found mutations in the parkin gene. [111] However, these mutations were present only in patients with age of onset at 20 years or younger but without any family history of PD. [111] Although restricted to familial and non-familial PD with an unusually early age at onset, the parkin gene provides a promising lead into understanding of the underlying mechanisms of neurodegeneration in both rare and more common sporadic forms of the disease. Results of basic research on the gene suggest that the product of the mutated versions of the gene can lead to pathologic protein accumulations within neurons by causing a defect in the ubiquitin pathway.[77] More research using molecular and animal models is currently underway to uncover other possible genes associated with PD.[100,111]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

### 2.1.6.3.3 Susceptibility genes (detoxification enzymes)

Certain genes clearly play an important role in the pathogenesis of some forms of PD.[76,111,112] Thus far, however, genes identified as directly resulting in development of PD have been limited to specific kindred with rare familial forms of the disease and cannot be generalized to the most common type (idiopathic PD) which has stubbornly failed to reveal any direct hereditary or autosomal dominant patterns and is more likely characterized by genetic heterogeneity.[76,113] Although equivocal, evidence from studies of the genetic basis of idiopathic PD does indicate some degree of hereditary influence on disease causation however.[76] This has led research priority toward investigations of genes that can indirectly confer susceptibility to PD.[112,114] Among such susceptibility genes are those that encode for enzymes responsible for metabolism of drugs and other environmental chemicals. [112,113] Genetic mutations resulting in defective detoxification enzyme production are hypothesized to increase the risk of PD through decreased hepatic metabolism of xenobiotics, thereby increasing their delivery and toxicity to the central nervous system.[111] Thus, individuals with the susceptibility gene are at higher risk of developing PD if exposed to environmental toxins. Three most likely candidate susceptibility genes associated with PD are monoamine oxidase B (MAO-B), glutathione-S-transferase (GST), and Cytochrome P450 2D6.[16,96,115-118]

#### 2.1.6.3.3.1 Monoamine oxidase B (MAO-B)

MAO-B gene polymorphisms have been a prominent focus of investigation for several reasons. First, cerebral MAO-B enzyme is involved in the metabolism of nigrostriatal dopaminergic neurons, the loss of which is a hallmark of PD.[119] Additionally, the parkinsonism-inducing molecule MPTP is activated to its neurotoxic



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

metabolite by MAO-B.[14,92,118,120] Relevance of this enzyme in the pathogenesis of PD is further evidenced by reduction of MAO-B activity by cigarette smoke which has repeatedly been shown to have a protective effect against the disease. Moreover, the MAO-B inhibitor selegiline has been clinically shown to slow progression of the disease.[14] Early studies measured MAO-B activity in platelets and were subject to bias, leading to dubious results. [14] Unfortunately, results of genotypic investigations to date have also been inconsistent with most studies failing to demonstrate an association. [14,119]

#### 2.1.6.3.3.2 Glutathione-S-transferase (GST)

Glutathione-S-transferase (GST) enzymes are of interest due to their integral role in xenobiotic biotransformation. GSTs are involved in conjugation, a process whereby compounds and/or their metabolites combine with a readily available endogenous chemical, forming a product that is more easily excreted.[49,114] GST enzymes are also important because reduced glutathione levels have been observed in the substantia nigra of patients with PD.[121] Several classes of GST enzymes have been genotyped for polymorphisms using PCR methods. Mutations in GSTM1 and GSTP1[112,122,123] have been reported to result in poor metabolism of toxins and an increased risk for PD, but results have been inconsistent and more studies are needed to confirm an association.[116,124,125]

#### 2.1.6.3.3.3 Cytochrome P450 enzymes

The largest enzyme system and one that plays a central role in detoxification of most xenobiotics is the cytochrome P450 family of enzymes. Because they play a major role in the body's defense against exposure to external chemicals, P450 enzymes have received the most attention in relation to PD. Although there are 35 different P450



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

enzymes in humans, only six have been studied in relation to PD: CYP1A1, CYP2C9, CYP2C19, CYP1A2, CYP2E1, and CYP2D6.[12,17,27,111,112,114,126-128] Moreover, CYP2D6 is the only one of the six P450 genes investigated that has shown a convincing association with PD.[108,129]

#### 2.1.6.3.3.3.1 Cytochrome P450 2D6 (CYP2D6)

Since 1985, more than 50 investigations have been conducted examining the relationship between CYP2D6 and PD.[128] The CYP2D6 gene codes for debrisoquine 3-hydroxylase and metabolizes more than 40 drugs including antidepressants, antiarrythmics, and antihypertensive drugs.[53] The gene has been of particular interest to PD researchers partly due to speculation about the involvement of CYP2D6 in metabolism and processing of neurotransmitters based on evidence of it binding to GBR-12935, a high affinity probe for the dopamine transporter.[130] A more intriguing feature of the Cyp2D6 with respect to PD however is its detoxification of the parkinsonism-inducing neurotoxin MPTP as well as many pesticides including organophosphates and organochlorines.[130] Moreover, CYP2D6 is characterized by high genetic heterogeneity and is one of few proteins present in functional form in the human brain (raising the possibility that it catalyzes reactions in the central nervous system).[36,131,132] Collectively, evidence from experimental and epidemiologic studies provide a strong argument for an association between CYP2D6 and PD.[55,132-139] Genetic heterogeneity in CYP2D6 results in two distinct phenotypes: extensive and poor metabolizers (EMs and PMs).[51-53,140,141] The distribution of alleles varies across ethnic groups. For example, the PM phenotype, the result of an autosomal recessive inheritance of nonfunctional (null) alleles of the CYP450 2D6 gene, and thought to confer susceptibility to PD, occurs in 7-10% of the



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

white population of Europe and the United States, but in less than 1% of Asians.[50,53,142] The association between CYP2D6 and MPTP was first reported in 1985, leading to subsequent phenotypic studies of the PM phenotype in relation to PD. Although most reports confirmed an increased risk for PD in poor metabolizers, myriad limitations associated with the use of pharmacokinetic assays have rendered doubt in the results of these studies.[133] Results from more reliable DNA-based genetic studies have been commensurate with earlier findings, however.[143] Two early British studies have shown mutations in this gene to double the risk of PD.[55,143] Subsequent studies have confirmed an increased risk,[55,131,134,135,143] but others have failed to show an association,[128,144-148] thus, no unequivocal conclusion has been reached on the role of the gene in PD.[128] Ambiguities in results of available research have led to re-evaluation of available studies using meta-analytic techniques.[108,129] Results of a meta-analysis conducted on 7 phenotypic and 11 genotypic studies of CYP2D6 and PD gave an odds ratio of 1.32 (95% CI, 0.98,1.78) The results were of borderline statistical significance (P = 0.074); thus, the meta-analysis was unable to either support or refute an association between CYP2D6 polymorphism and risk of PD.[108] Another meta-analysis using 8 phenotypic and 7 genotypic studies showed a statistically significant combined odds ratio of 1.48 (95% CI, 1.10,1.99).[129] However, the odds ratio for phenotypic studies alone was not statistically significant at 1.05 (95% CI, 0.63, 1.77) while the odds ratio for the genotypic studies alone was 1.67 (95% CI 1.11, 2.50).[129] Moreover, since significant results obtained from the meta-analysis of genotypic studies were due primarily to a single large study, the authors concluded that there was no association between CYP2D6 and PD, and recommended performance of other large genotypic studies.[129]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In light of the contradictory results, additional well designed studies are necessary to draw definitive conclusions.[108,132] Recent improvements in molecular technology enabling characterization of the CYP2D6 gene locus as well as identification of specific point mutations and deletions resulting in the PM phenotype have recently facilitated the search for answers to the nature and extent of CYP2D6 involvement in the pathogenesis of PD. Moreover, laboratory methods for detection of greater than 99% of poor metabolizer genotypes have been described in the literature.[50,142,149,150] However, methods currently available have not been optimized or validated for accuracy. Furthermore, most have failed to be reproduced in other laboratories [55]. Hence, a technique for rapid and accurate identification of CYP2D6 polymorphisms clearly remains a priority.

### 2.1.7 Exogenous (environmental) risk factors

While genetic forms of Parkinson's disease have recently been identified, these PD-related genes are absent in idiopathic PD, the most common form of the disease. The accidental discovery of MPTP, a simple molecule (discussed below) capable of inducing most if not all signs and symptoms of PD has since the mid 1980's led to a tremendous resurgence of interest in an environmental cause of the disease. Numerous case-control studies since have suggested a myriad of potential extrinsic risk factors and the most important ones are discussed below.

#### 2.1.7.1 Infectious agents

Following the great pandemic of von Economo's encephalitis lethargica from 1917-1928, there appeared a phenomenon heretofore unknown which later became referred to as postencephalitic parkinsonism.[104] As very little was known about



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Parkinson's disease at the time, some erroneously hypothesized that the disease only occurs in those exposed to the agent responsible for von Economo's encephalitis. [104] Today, however, it is understood that the disease known as Parkinson's disease is different from postencephalitic parkinsonism and unrelated to von Economo's encephalitis. [104] However, more recent laboratory studies in mice have implicated the neurovirulent influenza A virus in the etiology of PD based on the accumulation of antigens to the virus in the substantia nigra. [12,54] Moreover, there have been occasional reports of other cases of viral encephalitis resulting in parkinsonism, raising the possibility of exposure to such agents predisposing an individual to PD. However, seroepidemiologic studies have been unable to produce any evidence of an infectious etiology of the disease, leaving the influence of infectious agents ambiguous. [13]

**2.1.7.2 Head trauma**

The association between head trauma and PD is not well understood. [12,13] Yet, theories have been proposed attributing severe head trauma leading to concussion and possible reorganization of the central nervous system in response to peripheral nerve injury to increased risk of PD. [12] In the absence of credible epidemiologic evidence, however, head trauma can only be considered as an aggravating rather than a causative factor in PD.[12]

**2.1.7.3 Diet**

In light of accumulating evidence of the role of oxidative nerve cell injury in the etiology of PD, dietary factors having a potentially negative (increasing oxidative neuronal death) or positive (antioxidant properties) influence on oxidative processes have been of great interest to researchers. [12-14] Animal fat consumption has been associated



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with increased risk of PD. [12,13] Neuroprotective factors have also been identified, including the antioxidant Vitamins E and Vitamin C.[35] Alcohol consumption, an important factor in many disease processes has not been shown to have any effect on PD.[151] Results on the effects of dietary nutrient intake are based on few poorly designed studies and therefore must be viewed with caution. [12,14,15,18]

**2.1.7.4 Smoking**

Cigarette smoking has been repeatedly associated with a decreased risk of PD in a dose-dependent manner, with heaviest smokers having the lowest risk.[18] Moreover, experimental animals data have shown a protective effect of nicotinine against toxin-induced parkinsonism.[12] Several hypotheses have been formulated to explain the biochemical basis for the protective effects of cigarette smoking. Mechanisms proposed include inhibitory effects of nicotine and/or other components in cigarette smoke on monoamine oxidase B, a neurotransmitter responsible for catabolism of dopamine as well as activation of MPTP. [12] Other theories involve the reduction in free radical formation due to release of carbon monoxide from cigarette smoke. [12] However, although many potential biases leading to erroneous conclusions of a protective effect of cigarettes have been ruled out, the role of a possible psychological predisposition leading to particular personality characteristic or other smoking-associated behavior is still in question. [18]

**2.1.7.5 Metals**

Due to their ability to form free radicals in the body, thereby enhancing oxidative damage, a number of metals have been investigated as possible risk factors for PD.[18] Free radical forming metals explored include copper, zinc, and iron, all of which have shown to be associated with increased risk of PD in several epidemiologic studies.[12,14]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Another metal, manganese is known to cause a parkinsonian syndrome in humans, but the mechanism of neuronal degeneration is obscure. [12,14] Overall, although some studies have found an association between copper, iron, aluminum, lead, and manganese, other studies have failed to show any relationship. [12,14]

### 2.1.7.6 Industrial chemicals

As mentioned previously, Parkinson's disease appears to be more prevalent in industrialized countries, an observation that could provide important clues to the etiology of PD, such as to occupational-linked toxicants or other industrial chemical waste. Several ecologic studies have investigated areas of high disease frequency with respect to environmental risk factors in an effort to identify general risk trends. Using various techniques to estimate prevalence, such as total antiparkinsonian drug sales or county-specific PD mortality rates, these studies found increased rates of disease in areas with vegetable farming, wood pulp mills, and steel alloy industries, [133,152] paper, copper, and iron, and other chemical related industries,[153] and more generally, in rural areas.[154,155] Collectively, these results provide support for the hypothesis that exposure to agricultural and other industrial chemicals is associated with risk of PD.

### 2.1.7.7 Rural residence and well water consumption

Many studies have shown rural residence to increase the risk of PD.[24,25,102,156] Results of a meta-analysis of 16 case-control studies of the association between rural living and PD showed a combined odds ratio of 1.56 (95% CI: 1.17, 2.07) for all studies and a combined odds ratio of 2.17 (95% CI: 1.54, 3.06) for studies done in the United States. [25]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Although results vary widely between studies of rural residence, they have led to investigations of more specific factors related to a rural lifestyle. Well water consumption and pesticide exposure have emerged as components of rural life most frequently associated with PD.[18,103] Nineteen case-control studies examining the association between exposure to well water and PD were subjected to a meta-analysis.[25] Although there was significant heterogeneity among the studies, the combined odds ratio was 1.26 (95% CI: 0.96, 1.64) for all studies and 1.44 (95% CI: 0.92, 2.24) for studies conducted in the United States.[25] Presence of environmental contaminants, such as heavy metals and pesticides in well water is thought to be responsible for the observed increase risk.[12,14,18]

#### 2.1.7.8 Agricultural occupation

Farming and other jobs requiring exposure to certain metals and chemicals have been associated with an elevated risk of PD in many case-control studies. However, as with other factors, study results vary leading to uncertainty about the observed associations. [12-14,16,18]

#### 2.1.7.9 MPTP

In 1983, a contaminated batch of synthetic heroin caused an outbreak of parkinsonism in Northern California.[14,18,46] The toxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) was eventually identified as the responsible agent, becoming the first confirmed exogenous cause of parkinsonism. [14,18,46] MPTP is capable of inducing acute and irreversible PD-like symptoms in humans, as well as in monkeys.[14] Although the signs and symptoms resemble those of idiopathic PD in most aspects, MPTP-induced parkinsonism differs from classic PD since it does not result in formation



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of Lewy bodies in either humans or other animals.[14] MPTP is toxic because it is converted (by monoamine oxidase B) into 1-methyl-4-phenylpyridinium ion (MPP+), a neurotoxic molecule which is selectively imported across the blood-brain barrier into dopaminergic neurons by the plasma membrane dopamine transporter.[14]

**2.1.7.10 Pesticides**

Discovery of MPTP as a confirmed cause of parkinsonism naturally raised questions about the significance of other environmental neurotoxins in PD. Pesticides are one set of environmental chemicals that have been of particular interest in the pathogenesis of PD. The group of chemicals known collectively as pesticides is among the most poisonous substances that are routinely dispersed throughout the environment.[36,41,49] Pesticides have been in widespread usage for decades, and use of these chemicals is unlikely to abate in the near future.[49] An obvious source of occupational hazard in the agricultural industry where they are routinely used in large quantities, pesticides also pose a significant health risk to the general population through their use in home gardening and domestic pest control.[14] Furthermore, everyone is at risk of exposure to certain types of pesticides to varying degrees through contaminated food, water, fumigation, and aerial spraying.[49] Acute neurotoxicity of the organophosphate class of pesticides has been well established, and chronic exposure leading to a host of neuropsychiatric disorders is suspected.[36] Moreover, the mechanism of acute neurotoxicity of organophosphate pesticides is well understood.[49,157] Chronic neurological sequelae have also been extensively studied though the exact mechanism is less clear.[49] Given the ubiquity of pesticides in the environment, it is important to further investigate their role in neurodegenerative diseases like PD.[157]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Numerous epidemiologic studies conducted in various countries have implicated exposure to pesticides as a risk factor for PD.[26,122,158-167] Moreover, several investigations have been able to show an association between certain classes of pesticides.[26,158,159] In a study of PD patients in Calgary, Semchuck et al. were able to demonstrate that occupational herbicide use was a significant (threefold) predictor of PD risk, even after adjusting for potential confounders.[158] A population-based case-control study of PD patients in Detroit showed significant association (after adjusting for confounders) of occupational exposure to herbicides (odds ratio, 4.10; 95% CI:, 1.37,12.24) and insecticides (odds ratio, 3.55; 95% CI:, 1.75, 7.18) but not with fungicides.[26] In another study of young-onset Parkinson's disease, Butterfield et al. showed both exposure to both insecticides and herbicides to be significantly associated with PD (odds ratios 5.75 and 3.22, respectively).[159]

Pesticides encompass a large group of diverse chemicals, however, and studies to date have been unable to identify any specific agents.[167] Moreover, other epidemiologic studies have failed to show a relationship between pesticides and PD.[105,121,126,132,168-171] An argument against the role of pesticides centers on the fact that PD prevalence does not seem to have increased commensurate with increasing use of pesticides.[41] On the other hand, epidemiologic evidence of an increased risk of PD among farmers and residents in rural areas strengthens counter arguments. [41] Numerous epidemiologic studies have attempted to link pesticide exposure, particularly organophosphates, to PD.[14] The large majority of these studies have used a population-based case-control design in which exposure was determined through questionnaire responses.[36] Inadequate exposure assessment and poor recall have made identification of specific chemicals difficult or



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

impossible in most of the studies.[36] Moreover, variations in case ascertainment, incomplete or inaccurate exposure assessment, and use of prevalence rather than incidence data have caused unavoidable errors in most studies. Nevertheless, despite methodologic constraints, more than half of the studies conducted internationally have shown statistically significant associations between exposure to pesticides and PD.[26,106,122,158,159,161-164,166,167] Moreover, a recent meta-analysis of 14 peer-reviewed case-control studies examining the association between pesticides and PD confirmed a positive association.[25] The number of cases used in the 14 studies ranged from 38 to 224, and the number of controls ranged from 38 to 464.[25] One study showed a negative association while 2 showed no association. The remaining 11 studies reported positive associations, 4 of which had estimated odds ratios that were statistically significant.[36] The overall results of the meta-analysis showed a combined odds ratio of 1.85 (95% CI: 1.31, 2.60) for all studies and an odds ratio of 2.16 (95% CI: 1.95,2.39) for studies conducted in the United States.[25] In addition, recent animal studies have further emphasized the potential role of specific pesticides such as rotenone, paraquat, and maneb as risk factors for PD.[38,88] Clearly, the weight of the evidence from both observational and experimental studies suggests that pesticides have a considerable contribution to the development of PD. However, given the discrepant findings, it is reasonable to postulate that pesticides likely increase risk of PD in combination with other risk factors.

**2.1.7.10.1 Mechanism of action of pesticides in PD**

There are two primary mechanisms by which to explain the association between pesticides and Parkinson's disease: direct neurotoxicity and metabolism modulation.[36]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Extensive investigations of MPTP as the prototype neurotoxin have been of great use for the elucidation of the direct neurotoxic action of pesticides. [54] MPTP is activated by monoamine oxidase B to MPP+, which is then selectively transported into dopaminergic cells in the substantia nigra.[54,130,132] There, MPP+ inhibits complex I of the mitochondria, thus disrupting mitochondrial function and resulting in oxidative injury and cell death.[54] Evidence of both mitochondrial dysfunction and oxidative stress has been reported in PD, suggesting the role of other neurotoxins whose modes of action resemble MPTP.[29,54,93,172] Extending the MPTP paradigm to pesticides was an obvious next step in investigating extrinsic causes of PD because of the multiple ways in which pesticides resemble MPTP. The pesticide cyperquat is identical to MPP+.[54] Rotenone binds the same site on complex I as MPTP.[54] Another widely used pesticide, paraquat, has a similar structure to MPTP and induces mitochondrial damage although it promotes cell death via redox cycling. [35,36] Many other pesticides are confirmed mitochondrial toxins including chlordane, heptachlor, endosulfan, ethaphos, DDT, permethrin, and cyhalothrin.[36] Dieldrin is non-selectively desctructive to dopaminergic neurons in cell culture. Pesticides capable of oxidative damage include glyphosate, dinitrophenols, cyanide, benzonitriles, and rotenone.[36] This evidence, compelling on its own, is further supported by repeated epidemiologic observations of an association between pesticides and PD.[25,26,106,122,158-167] However, it does not provide a complete explanation for two important reasons. [36] First, it fails to explain how pesticides may selectively target dopaminergic neurons. Secondly, not all pesticides are considered to be mitochondrial toxins.[36] For example, organochlorines are a large class of pesticides that exert acute toxicity not by causing mitochondrial damage but through inhibition of sodium



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

channels.[37] Organophosphates and carbamates, two large groups of chemicals that belong to another important class of pesticides, also have a different mechanism of action.[49] Although well established as neurotoxins, these chemicals act by disrupting acetylcholine levels through inhibition of the enzyme acetylcholinesterase and are not involved in mitochondrial dysfunction.[49]

An alternate mechanism for the observed effects of pesticides on PD is that pesticides may modulate the metabolism of other neurotoxins by either inhibiting or inducing xenobiotic metabolizing enzymes.[36] In the body, the disposition of pesticides and other neurotoxins involves a multi-step process. The sequence can be divided into discrete phases, each of which requires the activation of different classes of enzymes such as cytochrome P450 and glutatione S transferase.[49] Laboratory studies have demonstrated the induction of cytochrome P450 enzymes by DDT and other organochlorines pesticides.[36] Epidemiologic study results using biomarkers to show increased marker activity in workers exposed to organochlorines have been consistent with laboratory findings.[36] On the other hand, organophosphates and pyrethrins are known inhibitors of cytochrome P450 enzyme activity.[36] Other pesticides such as the herbicide tridiphane and the fungicides captan and captofol inhibit glutathione transferase enzymes.[36]

In addition to modulating exogenous toxins, it has been hypothesized that pesticides may have a role in altering metabolism of endogenous compounds.[36] The potential significance of this as a mechanism for PD is related to the fact that dopamine can breakdown into toxic metabolites such as tetrahydroisoquinoline. [36] However, very few studies have explored this concept to date.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 2.1.7.10.2 Classes of pesticides

Pesticides can be functionally classified into three groups: insecticides, herbicides, and fungicides.[41] Cyperquat and paraquat are two herbicides that have been particularly suspected due to their structural similarity to MPP+, the neurotoxic metabolite of MPTP, even though paraquat has not been demonstrated to cross the blood-brain barrier as MPP+ clearly does.[54] However, an epidemiologic study of Taiwanese farmers who used paraquat and other herbicides showed a 4.74-fold increased risk of PD as compared to farmers who used herbicides other than paraquat who had a 2.17-fold increased risk.[167] Another pesticide associated with PD in laboratory studies of rats is rotenone, a plant-derived inhibitor of Complex I oxidative phosphorylation, widely used as both a household insecticide and as an agent for eradication of fish in fishery management programs. [173,174] The toxic mechanism of rotenone is identical to the aforementioned molecule MPP+ which accumulates in the mitochondria of dopaminergic neurons where it inhibits complex I enzyme activity.

Although studies of specific pesticides have been reported, most epidemiologic studies, constrained by limitations of estimating historic exposure, have classified exposure to pesticides as a general class rather than trying to examine exposure to specific chemicals. [25,77]

## 2.2 Molecular techniques

### 2.2.1 CYP2D6

#### 2.2.1.1 Background and evolution

Drug metabolizing enzymes (DMEs), commonly referred to as "liver detoxification systems" are a group of enzymes best known for their integral role in the breakdown of



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

drugs and other environmental chemicals toward successful elimination from the body.[52,175] Although primarily recognized in relation to exogenous compounds, DMEs have many endogenous compounds as their substrates as well.[52] Moreover, DMEs are involved in other important life functions including cell division and electrolyte balance.[52] In fact, not only does the existence of these enzymes long predate the manufacture of exogenous drugs, DMEs seem to have been present before the divergence of animals and plants from eukaryotic organisms, leading to evolutionary theories that DMEs co-evolved as a function of the interaction of plants and animals.[52] This theory is further supported by an explosion of new animal cytochrome P450 genes, a class of DMEs, about 400 million years ago when animals first came onto land and began to eat plants. Assuming that plants mounted a defense system and developed genes that could produce toxic compounds in response to this attack, it is hypothesized that animals likewise responded by evolving new genes to adapt to the constantly changing plants.[52]

DMEs can be classified into Phase I and Phase II enzymes.[52] The cytochrome P450 class of mixed function oxygenases represents most of the Phase I DMEs.[52] These monooxygenase proteins catalyze many reactions involved in the metabolism of man-made drugs and environmental chemicals and pollutants, as well as numerous and diverse endogenous chemicals such as cholesterol, steroids and other lipids.[53] CYP enzymes are involved in the initial phase of chemical biotransformation by catalyzing a variety of different reactions resulting in the oxidation of the compounds.[53] The resulting metabolites are more polar, making them less toxic and easier to eliminate from the body.[49]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Phase II DMEs include glutathione transferase enzymes and are responsible for conjugation of the Phase I intermediates to water-soluble derivatives, thus completing the detoxification cycle.[176]

The gene encoding debrisoquine hydroxylase (CYP2D6), a member of the cytochrome P450 superfamily of enzymes, was the first polymorphic gene locus identified to result in defective metabolism of chemicals among this class of enzymes.[49] Since this discovery, CYP2D6 polymorphisms have been studied intensively in relation to both interindividual variability in metabolism of exogenous drugs and chemicals and increased susceptibility to certain diseases.[175]

### 2.2.1.2 Structure

Through a combination of linkage mapping, in situ hybridization, and somatic cell hybrid analysis, the CYP2D6 gene has been localized to the long arm of human chromosome 22q 13. 1 where it occurs as part of a cluster of genes including CYP2D7 and CYP2D8, two highly homologous but inactive pseudogenes which appear to occur ubiquitously across individuals and ethnic groups.[53] CYP2D8 is a true pseudogene with many frame-disrupting mutations but CYP2D7 has only a single insertion leading to a frame disruption.[53]

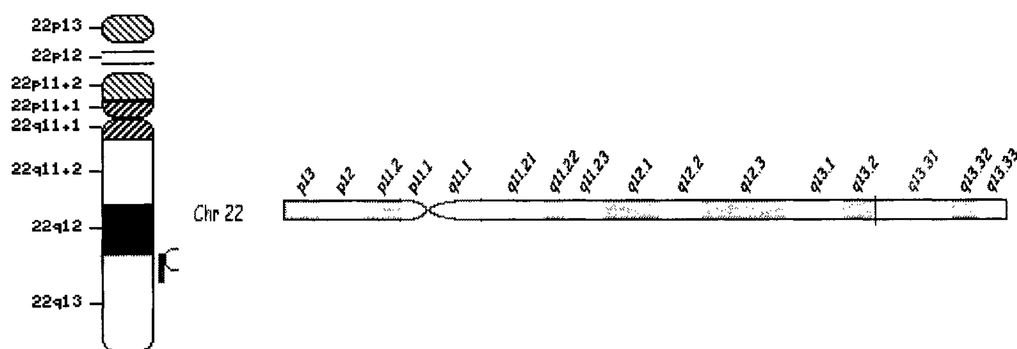



**Figure 2-1. Chromosome 22.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The normal CYP2D6 gene contains nine exons within a total of 4378 base pairs as seen below.[51] However, CYP2D6 has a large number of functionally different isoforms and is characterized by high polymorphisms.[51]

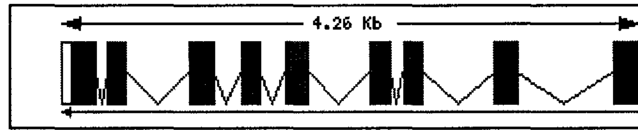



**Figure 2-2. CYP2D6 exon positions.**

### 2.2.1.3 Function

Since its purification from human liver in the 1980's, the CYP2D6 protein, which localizes in the endoplasmic reticulum, has been demonstrated to serve as a substrate for numerous exogenous and endogenous compounds.[53] Although structurally diverse, the majority of substrates for CYP2D6 are small molecules with a basic nitrogen atom.[53] Interaction of the substrate with the enzyme is through electrostatic charge between the nitrogen in the substrate and a negatively charged aspartate residue in the active site of the protein.[53]

A wide range of activity in CYP2D6 in human populations has now been well described, with interindividual metabolic rates estimated to differ up to 10,000 fold.[177] The gene has been of particular importance in the area of pharmacogenetics (study of interindividual differences in response to drugs) because of its involvement in the metabolism of at least 20% of all commonly prescribed drugs.[175] Hence, aberrant CYP2Dr6 expression can potentially produce a variety of deleterious pharmacologic and clinical consequences.[175] Moreover, their demonstrated impact on numerous chronic diseases has led to broader ecogenetic (study of interindividual differences in response to



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

all environmental chemicals and physical agents) investigations of CYP2D6 allelic variants.[175]

#### 2.2.1.3.1 Polymorphisms

CYP2D6 is highly polymorphic in the population.[53] The gene qualifies as polymorphic on the definitional basis that a gene is polymorphic if it results in two or more different phenotypes, and, presumably, genotypes, neither of which is rare (occurs at a frequency of less than 1 percent).[178] Genetic variability in the CYP2D6 gene results in a bimodal distribution representing either a normal or decreased ability to metabolize the enzyme's substrates.[53] Alleles resulting in the poor metabolizer phenotype are inherited as an autosomal recessive trait.[53] It is estimated that 7-10% of Caucasians inherit two non-functional alleles, and are thus poor metabolizers (PM) of xenobiotics, with the rest of the Caucasian population falling into the extensive metabolizer category.[177] There is no evidence of polymorphisms in the two pseudogenes located near the CYP2D6 gene, thereby restricting allelic variation to the functional gene alone. [53,149] Although the reason for the observed polymorphisms in the CYP2D6 gene is unclear, one mechanistic hypothesis is that CYP2D6 is located near the immunoglobulin in a genetically unstable region of chromosome 22. [53] Regardless of the reason for the observed variation, numerous allelic variants at the CYP2D6 locus have been described, several of which result in gene-inactivating mutations. [53] The molecular basis of the enzyme deficiency has been well characterized with more than 50 different mutations in the gene sequence described to date. [53]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

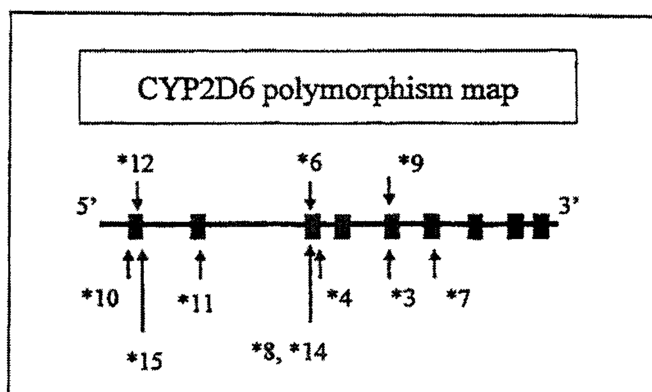



**Figure 2-3. CYP2D6 polymorphism map**

However, more than 99% of all PM phenotypes among Caucasians can be traced to five point mutations and a deletion of the entire CYP2D6 gene. [51,59] The most common gene inactivating allelic variants are *3 (single nucleotide deletion in Exon 5 resulting in frameshift) and *4 (several silent single nucleotide substitutions and a point mutation in the consensus sequence for the splice site at the intron 3/exon 4 splice junction). [132,134] Together, with the *5 (gene deletion) variant, these alleles account for greater than 95% of all poor metabolizers in the Caucasian population. [132]

One of few polymorphic genes encoding for enzymes that are present in functional form in the brain, CYP2D6 is also suspected to catalyze reactions in the central nervous system. [114,137] In particular, it is hypothesized that the enzyme is involved in the metabolism of dopamine neurotransmitters. [128] Although this theory remains to be proven, together with evidence that CYP2D6 is involved in the metabolism of several chemicals that have strongly been associated with Parkinson's disease, it implicates the gene in the etiology of Parkinson's disease.[51,55,108,128,129,132-137,139,143,179] Specifically, CYP2D6 polymorphisms leading to the PM phenotype are presumed to increase risk of PD. [108,128,129] Relevance of poor CYP2D6 expression in the association between environmental chemicals and PD depends on the extent to which CYP2D6 is the major



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

route of metabolism of the toxic compound of interest, as well as the range and concentration of that compound.[53]

### 2.2.2 Sample/DNA collection

Genomic DNA can be routinely collected from numerous biological samples such as peripheral blood, tissue, dried blood spots, and buccal cells. [180] However, although methods used to extract and process DNA from various biological specimen are largely equivocal with respect to laboratory time and effort, the least invasive of all collection techniques are preferable to minimize discomfort in subjects from whom the samples are obtained. [150,180] Moreover, collection of blood, which requires medically trained staff and special facilities, is often refused due to medical, cultural, religious, or other personal reasons. [150,180]

#### 2.2.2.1 Buccal cells

Oral epithelial cells, an easily accessible source of genomic DNA, have recently been promoted as a safer and less invasive alternative to traditional collection methods requiring venous blood.[180,181] DNA derived from epithelial cells has recently been reported as adequate for use in genotyping procedures including polymerase chain reaction. Moreover, epithelial cells can be either self-collected at the subject's convenience at home or workplace or can be obtained from subjects in a clinic setting. Although the buccal sample collection is advised under the direction of an interviewer or other staff involved in the study, the supervisory person need not be a medically trained technician. [180,181]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 2.2.2.1.1 Advantages

As discussed in the previous section, buccal cells are a more suitable source of DNA for genetic epidemiology studies for several reasons. Epithelial cells are easily accessible, constantly exfoliated from the oral mucosa, and can be collected with little or no discomfort by gentle scraping of the inside of the cheek or by oral rinsing. [180,181] Moreover, buccal cell collection eliminates exposure to blood products, does not require presence of medically trained professionals, and, thus, is more cost and time effective for sampling of large populations. [180,181]

#### 2.2.2.1.2 Disadvantages

Although buccal cells have been successfully used as a source of DNA in some genotyping assays, they are unsuitable for amplification of large DNA fragments due to excessive shearing of the genomic DNA. [150] Thus, buccal cell-derived DNA is only appropriate for PCR products of less than 500 base pairs and cannot be used in molecular assays that require long PCR products, such as CYP2D6.[150]

### 2.2.3 Genotyping

Traditionally, xenobiotic metabolizing enzyme activity has been determined using phenotyping methodology. [177,182] The phenotype technique allows prediction of enzyme activity through administration of a probe drug (a substrate to the enzyme of interest) and subsequent measurement of the ratio of unchanged drug to metabolite by either high-performance liquid chromatography or gas chromatography.[177] More recently, however, new molecular genetic techniques provide alternatives to phenotyping methods to predict interindividual enzyme activity by allowing direct determination of the genotype. [177] Genotyping is superior to phenotyping methods for several reasons. First, genotyping



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

provides an unequivocal genetic basis for drug metabolism based on exact allelic inheritance. [177] Moreover, this technique has higher accuracy, is cheaper, faster, and less invasive, and is not subject to the confounding effects of drug-drug interactions and variable metabolic processes among individuals that may skew the results from phenotypic studies. [177] Molecular techniques are currently used to identify variability in a variety of susceptibility genes. [177] Most methods such as polymerase chain reaction (PCR) are simple to perform and easily adaptable in a molecular biology laboratory. [183] However, some susceptibility genes have unique characteristics that represent a challenge to their genotyping. [150] CYP2D6 is one such gene.

#### 2.2.3.1 Polymerase chain reaction

The technique known as polymerase chain reaction (PCR) is an in vitro laboratory method for the enzymatic amplification of nucleic acid (DNA) sequences. [183] Amplification increases the amount of DNA, thus enabling better detection of segments of the target sequence of interest. Before performing the polymerase chain reaction, a nucleic acids (DNA) sample must be collected and isolated from a biological source such as tissue, blood, or other body fluid. Next, a set of short (20-30 bases) pieces of single stranded DNA called oligonucleotide primers must be designed to specifically hybridize (anneal with) either end of the DNA sequence to be detected. Once the target DNA is isolated and appropriate primers are designed, a heat-stable enzyme with DNA polymerase activity can amplify the DNA in the presence of two or more primers, deoxynucleoside triphosphates, and magnesium. In addition, it is necessary to add a reaction buffer providing proper pH, salts and other cofactors to the final mixture. A single PCR reaction cycle consists of heating the mixture to a temperature sufficient to



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

break the bonds between the two bases that make up each rung of the molecule, thus separating the two complementary strands of the target DNA. The solution is next cooled to a temperature that allows the primers in the mixture to hybridize (anneal with) the single stranded target DNA. The last step in the cycle is to control the reaction at a temperature whereby the DNA polymerase enzyme can stick to and extend the primers in order to copy the initial two strands of DNA into four. This cycle is repeated (e.g.,, 30 cycles) in an instrument called a thermal cycler capable of changing temperature rapidly, uniformly, and automatically, until DNA is amplified to an adequate amount to be detected. [183]

Depending on the nature of the target DNA to be amplified, the reaction may require increased sensitivity and specificity. [142] For example, if the gene of interest has high homology with another gene or genes, with the exception of only a few base pairs, the sequence to be detected could be identical to sequences on the homologous genes. [142] Thus, design of primers specific only to the target DNA becomes a difficult challenge. In such a case, a modified PCR protocol called a nested PCR can be used to increase specificity of the target sequence. Nested PCR involves two rounds of amplification using two sets of primer pairs. In the first round, a shorter segment of the gene of interest (ideally an area with no or little resemblance to other gene segments) is amplified using the outermost primers in an effort to reduce the overlap of DNA sequence with other homologous genes. The amplified piece is then subjected to another round using the internal primer pair designed to more specifically amplify the desired sequence. [142]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## 2.3 Gene-environment interactions

### 2.3.1 Introduction

The advent of new molecular techniques and the Human Genome Project have opened a Pandora's box of genetic information, heralding a new era in the field of genetics and molecular biology.[3,5,8] Although the vast amount of data generated will undoubtedly provide valuable information regarding the contribution of genes in the pathogenesis of disease, each piece of genetic information will also introduce doubts regarding the significance to human health. For example, although several genes have been discovered that directly lead to certain conditions such as breast cancer, the genes identified account for a negligible proportion of all breast cancers.[184,185] It is increasingly obvious that the identification of genes alone is insufficient to explain the genesis and progression of most diseases of multifactorial origin.[8,68,186] Widespread availability of genetic data has inevitably led to a stronger alliance between genetics and epidemiology, thus, underscoring the importance of assessing the combined effects of genetic and environmental risk factors in the development of disease.[8,68]

At present, there is no standard or universal method for investigations of gene-environment interactions as determinants of complex diseases.[3,5] Accuracy criteria to determine precision and validity of gene-environment interaction study designs have not been examined to a great extent. Various approaches have been considered and used, none of which has been adequately scrutinized with respect to study efficiency as measured by a reduction in random error relative to power.[56,62,187] For example, the effects of variations in study design leading to increased efficiency such as selection of a target population likely to be genetically susceptible or exposed to the risk factor of



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interest, or who have or will likely develop the disease remain to be addressed in relation to gene-environment interaction studies.[187] Efforts are underway for more in depth assessments of the efficiency of each study design.[56,59-64,67,68,187] However, pending further information availability, it is largely agreed that the choice of study design should be carefully considered on a case by case basis, depending on multiple factors such as the disease itself, type of environmental exposure, nature of susceptibility gene, frequencies of each in the population of interest, suspected magnitude of interaction, and amount of information available. [187]

### 2.3.2 Definition

The phenomenon known as interaction can be described in several ways. Statistically, interaction is considered to be present if two risk factors do not act independently to cause disease. [188] Statistical interaction is equivalently termed "heterogeneity of effect" or "effect modification" and described as heterogeneity in the magnitude of a primary effect measure resulting from the influence of the value of a third variable responsible for modification of the effect. Independence is measured statistically on either an additive or multiplicative scale depending on the underlying model. Biologically, interaction is viewed as a mechanistic function of a disease process. [6,7,186] Conceptually, interaction has been described as the "heterogeneity of effect",[21,188] "discordance between prediction and observation"[72], or more descriptively, as "the coparticipation of two factors in a single causal mechanism" and/or component causes in competing causal mechanisms.[6] Likewise, the specific case of interaction referred to as gene-environment interaction may be defined in two ways: biological and statistical. A biological explanation of gene-environment interaction could simply be stated as the



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

"genetic control of sensitivity to the environment".[8] Alternatively, the stricter statistical definition can be defined as "a different effect of an environmental exposure on disease risk in persons with different genotypes" or similarly, "a different effect of a genotype on disease risk in persons with different environmental exposures."[8]

### 2.3.3 Biological mechanisms

Experimental studies using animal models have demonstrated that interactions are fundamental properties of most genetic systems.[8] Pharmacogenetic studies have confirmed these findings by showing different responses to exposure to environmental chemicals among individuals with differing genotypes.[8] However, genes can affect disease outcome differently, depending on whether they act directly or indirectly. Therefore, it is important to distinguish between genetic alterations that can play a direct role in disease causation and variations on gene products that indirectly affect disease risk. For example, mutations have been identified in certain genes that are directly responsible for development of diseases such as breast cancer.[189] More commonly, however, genes responsible for metabolic processes in the body and known to have multiple allelic variants can increase an individual's susceptibility and thus indirectly contribute to disease causation.[9,190] Given the right combinations, allelic variants in susceptibility genes can interact with various environmental risks encompassing a broad spectrum of factors including either physical (e.g., temperature, radiation) or chemical (e.g., pesticides, hydrocarbons) exposures, biological agents (e.g., viruses, bacteria), lifestyle-related factors (e.g., injury, rural living), and even behaviors (e.g., illicit drug use, late age at first pregnancy) to increase the chances of developing disease.[6]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

### 2.3.4 Models of interaction

The underlying biological mechanism leading to disease state necessarily varies depending on the nature of the interactions in question. [10] Therefore, different statistical models of interaction have been proposed for several biologically plausible models of gene-environment interactions.[6,7,10,65]

In the first biological model (Type 1 interaction), both exposure and genotype must be present for the occurrence of disease.[65] The relationship between phenylketonuria (PKU) gene, diet (risk factor), and mental retardation (disease) illustrates the Type 1 model of interaction. Individuals homozygous for the recessive PKU gene lack the enzyme to convert phenylalanine to tyrosine resulting in abnormally high blood levels of phenylalanine that lead to mental retardation. However, since food is the source of phenylalanine, this outcome can be prevented through dietary restrictions than can reduce the amount of phenylalanine in the blood.[191]

In the second biological model (Type 2 interaction), disease is caused by environmental exposure alone and presence of a particular genotype can exacerbate the effect of a given environmental risk factor. Hence, genotype only plays a role in the context of interaction with the exogenous risk factor and is neither a necessary nor a sufficient causal factor. This model is applicable to xenobiotic metabolizing genes that are known to occur in various defective forms which can differentially affect an individual's susceptibility to a given disease but that cannot alone produce disease.[65] An example of the Type 2 model is development of squamous cell carcinoma of the lungs in smokers with a genetic deficiency in metabolism of carcinogenic chemicals in cigarette smoke caused by a polymorphism in the glutathione-S-transferase M1 gene.[192] Another



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

possible example of a Type 2 GEI model, and the basis for this dissertation, is the relationship between pesticide exposure and Parkinson's disease in individuals with a polymorphism who have a reduced capacity to metabolize pesticides due to a polymorphism in the CYP2D6 gene which encodes for the enzyme responsible for detoxification of these compounds.

The third biological model (Type 3 interaction) represents the opposite of the previous scenario. In this case, presence of genotype alone can cause disease. Environmental exposure serves to exacerbate the effect of genotype but singularly has no effect on disease state. This scenario could occur in the presence of a genotype resulting definitely in disease but that can be manifested to varying degrees depending on an exposure that may intensify severity.[65] An example of the Type 3 model is the relationship between porphyria variegata, an autosomal dominant disorder resulting in skin abnormalities such as blisters and exposure to barbiturates. Barbiturates do not usually lead to acutely adverse effects in the general population. However, in persons with porphyria variegata, exposure to barbiturates can lead to paralysis or even death.[193]

Lastly, another model can be envisaged in which both exposure and genotype alone have some effect on disease risk, but where presence of both risk factors increases (or decreases) the chance of disease more than the sum of the presence of either alone. For example, the risk of chronic obstructive pulmonary disease is increased in individuals with either α-1-antitrypsin deficiency or those who smoke but the risk is greater among those with both risk factors than the sum of the risk for smokers or those with the α-1-antitrypsin deficiency.[6,7]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

### 2.3.5 Scale of measurement

Statistically, interactions can be measured on two different scales: multiplicative and additive. Mathematical notation is used to define each scale and to illustrate the difference between the two.

Let $RR_{11}$ = relative risk of disease given presence of both risk factors,

$RR_{01}$ = relative risk of disease given one risk factor, and

$RR_{10}$ = relative risk of disease given other risk factor.

Under the additive model, the absence of interaction can be stated as:[21]

$$RR_{11}-1 = (RR_{01}-1) + (RR_{10}-1)$$

Synergistic and antagonistic interactions under the same (additive) model are expressed as: [21]

$$RR_{11} > RR_{01} + RR_{10} - 1 \text{ (synergistic) and}$$

$$RR_{11} < RR_{01} + RR_{10} - 1 \text{ (antagonistic).}$$

Comparatively, under the multiplicative assumption, the absence of interaction, synergistic interaction and antagonistic interaction respectively are stated as: [21]

$$RR_{11} = RR_{01} \text{ x } RR_{10},$$

$$RR_{11} > RR_{01} \text{ x } RR_{10}, \text{ and}$$

$$RR_{11} < RR_{01} \text{ x } RR_{10}.$$

Equivalently, effect modification on a multiplicative scale can be detected by stratification of the study group into categories based on the effect modifiers in question, followed by calculation of the stratum-specific effect measures which can then be compared to each other to determine the presence of heterogeneity (effect measure is not equal across strata) or homogeneity (effect measures are equal across strata) of effect.[21]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Values of all relative risks can generally be estimated using odds ratios from case-control studies.

Applying the above principle to the more specific case of gene-environment interaction, such interaction can be said to exist if the joint effect of gene and environment differs from either the product of the relative risks of individual factors on multiplicative scale or the sum of the excess relative risks for environment and gene on additive scale.[9]

The appropriate scale of measurement in studies of interactions has been a matter of great controversy. Rothman favors the use of the additive scale of measurement as a fixed reference point.[21,188] He states that the additive model is the only meaningful one based on the epidemiologic concept that two factors are not independent if they are component causes in the same sufficient cause or in competing causes. He follows by asserting that in such cases, additivity of the excess relative risks is the state of no interaction. Therefore, two independent factors that are part of different non-competing causal mechanisms will have a perfectly additive relationship whereas factors that are part of the same or competing causal pathways will have a greater or less than additive effect (i.e. departure from an additive scale).[21] This explanation has been challenged by others who believe the choice of scale should be determined on a case-by-case basis depending on the underlying pathophysiologic mechanism of the disease. For example, it is mostly accepted that an additive model is more appropriate in studies of multistage diseases where two factors act at the same stage, whereas studies of diseases with risk factors that act at different stages would require the use of a multiplicative model of interaction.[6,7] In the absence of information regarding the biological mechanism of



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

disease, the choice of scale has been suggested to be arbitrary. Although not fully accepting the selection of scale as completely arbitrary, most proponents of using an additive scale do concede that the choice of scale depends on many factors, particularly the goals of the study. [21] Rothman himself agrees that selection of model should be based on the objectives of the study, and suggests that a multiplicative scale, due to its simplicity relative to an additive scale, is a reasonable choice in studies aimed at predicting the presence of interaction since "description and prediction need not be linked to biologic inference" and "evaluation of statistical interaction, as a step to finding a simple, accurate predictive function, need not be tied to specific biologic models".[188]

In case-control studies, departures from multiplicativity can be assessed by calculating the interaction effect, referred to as the synergy index on a multiplicative scale (SIM) and defined as the joint odds ratio for the exposure and genotype divided by the product of the odds ratio for the effect of exposure alone and the odds ratio for the genotype alone as follows:

$$SIM = OR_{GE}/(OR_{Ge} \times OR_{Eg}) \qquad (2\text{-}1)$$

where

$OR_{GE}$ = the odds ratio of disease for those with both the susceptibility genotype and environmental exposure relative to those without either risk factor,

$OR_{Ge}$ = the odds ratio of disease, among the unexposed, for those with the susceptibility genotype relative to those without the genotype, and,

$OR_{Eg}$ = the odds ratio of disease, among the genetically non-susceptible, for the exposed relative to the unexposed.[194]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A SIM of greater than 1 indicates a greater than multiplicative effect between the exposure and genotype.[9]

### 2.3.6 Epidemiologic approaches

Methods used to study gene-environment interactions have been rapidly evolving in the post Human Genome Project era. [3,5] As a result, traditional heritability studies using linkage techniques have become less useful in the study of genetic determinants of disease due to their inability to account for interactions. [3,5] In their place, population-based epidemiologic studies that can simultaneously measure extrinsic and intrinsic risk factors as well as the interaction of these factors have emerged as a more comprehensive alternative study design.[3,5] Case-control studies along with new variants based on the case-control design are most conducive to and thus, commonly used in the evaluation of gene-environment interactions. [8,59]

#### 2.3.6.1 Case-control studies

The case-control study design is the most frequently used design in the study of gene-environment interaction.[9] Additionally, as more genetic markers become identified, case-control studies will increasingly be used to define the potential role of genetic factors in the causal pathway of many diseases. However, the case-control method was originally designed to evaluate the main effects of risk factors. [23,65] Thus, it is important to re-evaluate and adapt the standard case-control design for studies of gene-environment interaction. [23,56,64]

##### 2.3.6.1.1 Case-control design

A traditional design in many epidemiologic investigations, case-control studies have been described in full detail elsewhere.[21] Briefly, this method uses controls (non-



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

diseased subjects, ideally selected from the same source population from which cases were selected) as the referent group against which risk estimates for the diseased groups can be calculated. Controls are typically selected from a population of individuals who are free of the disease of interest. Adjustment for confounding can be accomplished using either stratification, including individual matching (e.g., machted case-control pairs), or multivariate statistical techniques. An important design consideration is the fact that the case-control method is a model-dependent concept. As such, any interactions explored in this type of study must be defined explicitly as either multiplicative or additive.[9] The choice of interaction model depends on many factors including overall goals of the study. Recall that it is generally accepted that most gene-environment interaction investigations aimed at describing the etiologic mechanisms of disease can be assessed on a multiplicative scale.[188]

#### 2.3.6.1.2 Sample size and power required to detect gene-environment interactions (GEI) in case-control studies

Statistical power and sample size determination are two critical elements in the design of case-control studies.[23,62,64,195] Power of the study, defined as the probability of rejecting the null hypothesis ($H_0$) when it is false, depends on the sample size.[188] Traditional case-control methods were developed based on evaluating main effects without consideration to the influence of possible interactions.[9,23,65] However, recognition of gene-environment interaction effects as central to disease processes and, thus, epidemiologic studies has led to a re-examination of power and sample size calculations in case-control studies.[9,23,56-58,62,65,67,69,70,194-198] The consensus reached through these investigations is that for case-control studies in which the study objective is to determine gene-environment interaction



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

as well as main effects, sample size requirements must often be increased to assure adequate statistical power when the objective of a study is estimation of interaction effects.[9,23,59] As shown by Smith and Day,[23] the following parameters must be specified to calculate sample sizes for case-control studies of gene-environment interactions: 1) prevalence of exposure ($P_E$); 2) prevalence of genotype ($P_G$); 3) odds ratio for exposure alone ($OR_{Eg}$); 4) odds ratio for genotype alone ($OR_{Ge}$); 5) the odds ratio for gene-environment interaction (SIM); 6) the ratio of cases to controls; and, 7) probabilities for Type I and II errors ($\alpha$ and $1-\beta$). The most important factors in determining sample size in a case-control study of gene-environment interaction are frequency of genotype, frequency of exposure, and SIM.[59]

The power to detect gene-environment interactions has theoretically been shown to be proportional to the magnitude of interaction.[59,64] Moreover, power is enhanced when both genetic and environmental risk factors are common in the population.[9,59,64] If, however, one of the risk factors under study in a gene-environment model is rare, the number of study subjects must increase substantially in order to achieve adequate power to detect an interaction.[9] Smith and Day showed that in order to achieve the same power, sample size for a study of interaction effects (when prevalences of genotype and exposure are low in the population) must be four times as large as for a study of main effects [23]. Moreover, simulation studies have been conducted to demonstrate that study size for a case-control study can be prohibitively large given certain combinations of odds ratios and frequencies of the two factors under investigation.[56,59] Therefore, the objective of a study (importance of interaction effects) is a critical methodological consideration toward achieving adequate study power.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 2.3.6.1.3 Analysis of gene-environment interactions (GEI) in case-control studies

Recall that gene-environment interaction effects, assuming both genetic susceptibility and environmental exposure are dichotomous, can be estimated by calculating the SIM (Equation 2-1).[194] Of course, in addition to determining the joint effects of genotype and environmental factor on disease risk, case-control studies allow the evaluation of individual effects of each risk factor on disease while adjusting for potential confounding factors. For example, in studies where no interaction effect is detected, assessment of the main effects can provide valuable information about the individual contributions to disease.

Potential confounding in the study can be detected and controlled for using either of two different approaches: stratification by the covariate suspected of confounding the effects or by multivariate logistic regression analysis, a binomial parametric model suitable for case-control studies where the outcome variable is dichotomous.

In a logistic model, the quantity $\Pi(x)$ represents the expected probability of the outcome and is expressed as:

$$\Pi(x) = \frac{e^{\beta_o + \beta_1 x}}{1 + e^{\beta_o + \beta_1 x}}$$

where

x = the risk factor of interest.

$\beta_1$ = the coefficient of risk factor x, and,

$\beta_0$ = the intercept parameter.

$\Pi(x)$ is transformed from S-shaped to linear in x using the logit transformation as follows:



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\ln\frac{\Pi(x)}{1-\Pi(x)} = \frac{\frac{e^{\beta_o+\beta_1 x}}{1+e^{\beta_o+\beta_1 x}}}{1-\frac{e^{\beta_o+\beta_1 x}}{1+e^{\beta_o+\beta_1 x}}} = \ln(e^{\beta_o+\beta_1 x}) = \beta_0 + \beta_1 x$$

Confounding variables can be added to this univariate model and their effect on the risk factor of interest can be evaluated.

The equation for a standard fitted logistic regression model for genotype ($\mathcal{G}$), environmental exposure ($\mathcal{E}$) and for gene-environment interaction ($\mathcal{GE}$) can be expressed as:

$$\text{logit } P(D=1) = \beta_0 + \beta_1\mathcal{G} + \beta_2\mathcal{E} + \beta_3\mathcal{GE} \tag{2-3}$$

where

$\beta_1 = \ln(OR_{Ge})$,

$\beta_2 = \ln(OR_{Eg})$,and,

$\beta_1+\beta_2+\beta_3 = \ln(OR_{GE})$.[62,199]

Departures from gene-environment interactions on a multiplicative scale (SIM) can be evaluated using the exponentiated value of $\beta_3$. This can be shown using the following equation:

$$SIM = OR_{EG} / (OR_{Ge} * OR_{Eg}) = \left(\frac{e^{\beta_1+\beta_2+\beta_3}}{e^{\beta_1}e^{\beta_2}}\right) = e^{\beta_3}. \tag{2-4}$$

A value of $e^{\beta_3}$ greater than one represents synergistic multiplicative interaction and a value of $e^{\beta_3}$ less than one shows an antagonistic multiplicative interaction.[62,69].

### 2.3.6.1.4 Advantages of case-control studies

The case-control method is particularly suitable for study of rare diseases with long latency such as Parkinson's disease.[21] Moreover, the study design allows



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

evaluation of multiple risk factors and the interaction of these factors, essential components of a gene-environment interaction study. [59] Additionally, case-control studies are more cost and time effective and generally require smaller sample sizes than cohort studies. [21]

#### 2.3.6.1.5 Disadvantages of case-control studies

Selection of an appropriate control group has traditionally been one of the biggest disadvantages of the case-control study design. [21] This shortcoming is especially problematical when the disease of interest has a long latency period. Since gene-environment interactions are a prominent feature of most chronic diseases with long latency periods, selection of proper controls while maintaining internal validity has become a daunting obstacle for studies of this type. [21]

Evaluation of interactions in case-control studies present additional unique challenges.[63] Although this study design is usually preferred over the cohort design due to its smaller sample size requirement,[21] several studies have suggested that the efficiency of the method becomes compromised when trying to assess gene-environment interactions. [9,23,62,63] Specifically, given small sample sizes, gene-environment interactions can be detected using a case-control design only if the environmental and the genetic factors are common.[9,59]

### 2.3.6.2 Case-only studies

In order to address the important limitations of the case-control study design arising from improper selection of controls and prohibitively large sample size requirements, the case-only design has recently been proposed as a more efficient method to evaluate the interaction effects of genetic and environmental risk factors in disease



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

causation.[57,194,198] Since its introduction, this approach has been used to detect interaction between environmental factors and genetic markers in a variety of diseases of multifactorial origin (e.g., association between spontaneous abortion and polymorphism in human estrogen receptor gene in women with estrogen-receptor positive breast tumors.[9]) The underlying rationale for the design as well as several important advantages and limitations are discussed in detail in the following sections.

#### 2.3.6.2.1 Case-only study design

The case-only model follows the same epidemiologic principles of case selection as for any other case-control study.[2,9,57,194] As discussed above, this variant design differs from the case-control method from which it was derived in one important way: the association between an environmental exposure and a genotype is examined among case subjects only (affected subjects with a given disease).[9] Cases are categorized into either a pseudocontrol or pseudocase group depending on absence or presence of susceptibility genotype respectively. [9,194] Prevalence of environmental exposure (i.e., risk factor) is then compared among the two groups, with the pseudocontrols serving as the referent group. Odds ratios and confidence intervals are computed using either standard crude analyses or multivariate models to adjust for covariates.[194,197] An important limitation is that the case-only design is valid only if the investigators can assume that genotype occurs independently of the environmental exposure in the population.[58,194] If this assumption is met, the case-only odds ratio (COR), the odds ratio relating exposure to genotype, is equal to the interaction effect (SIM) as measured in a case-control study under a multiplicative model.[194]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 2.3.6.2.2 Principle of the case-only design

The case-only design is based on the observation that when exposure and genotype occur independently among the general population, the odds ratio relating the exposure to the genotype among controls in a case-control study is one.[2,23,194]

To describe the reasoning behind the case-only study design, we begin by constructing a 2 x 4 table as shown in Table 2-1 below:

**Table 2-1. Two by four table for gene-environment interaction analysis in a case-control study.**

| Genotype | Exposure | Cases | Controls | Odds Ratio (OR) |
|---|---|---|---|---|
| + | + | a | b | ah/bg ($OR_{GE}$) |
| + | - | c | d | ch/dg ($OR_{Ge}$) |
| - | + | e | f | eh/fg ($OR_{Eg}$) |
| - | - | g | h | 1 |

Using notation from Table 2-1, it can be shown that SIM is equal to the OR for association between genotype and exposure among cases (COR) divided by the OR for association between genotype and exposure among controls:

$$\text{SIM} = \frac{\frac{ah}{bg}}{\frac{eh}{fg} x \frac{ch}{dg}} = \frac{ah}{bg} x \frac{fg}{eh} x \frac{dg}{ch} = \frac{ag}{ce} \div \frac{bh}{fd}.$$

Thus, if the odds ratio for association between genotype and exposure among controls equals 1, which is the case when exposure and genotype act independently, then SIM is equal to the OR for association between genotype and exposure among cases (COR), and departure from multiplicative independence can be detected using only cases.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 2.3.6.2.3 Sample size for case-only studies

Assuming independence of exposure and genotype and presence of background risk unrelated to either exposure or genotype, Yang et al have developed a method by which to estimate the required number of cases in a case-only design to detect gene-environment interaction based on expected prevalence of exposure and frequency of gene in the population, independent risks of gene and exposure, gene-environment interaction effect, and type I and II errors.

#### 2.3.6.2.4 Analysis of gene-environment interaction (GEI) in a case-only study

If the exposure and genotype are independent, a case-only odds ratio (COR) can be calculated from a 2 x 2 table of genotype versus exposure among cases. Alternatively, COR can be estimated by fitting the logistic regression model below to the cases:

$$\text{logit } P(G=1) = \beta_0 + \beta_1 \mathcal{E} \tag{2-6}$$

where G= presence of genotype.[69] In this model, COR = $e^{\beta_1}$. Confidence intervals can also be obtained using the same formulas as for case-control studies given a sufficiently large sample size. Similarly, fitting the logistic regression model logit $P(G=1) = \eta_0 + \eta_1 \mathcal{E}$ to the controls alone would allow evaluation of the independence assumption.[69]

#### 2.3.6.2.5 Advantages of case-only design

The strongest advantage of the case-only study to the case-control study is that, under the independence assumption, use of the former model yields higher precision for estimation of gene-environment interactions than traditional case-control studies.[2,57,194,198] The increase in precision results from a decrease in the standard errors due to the elimination of control group variability without compromising power to detect the interaction effect.[2,57,194,198]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 2.3.6.2.6 Disadvantages of the case-only design

Although it has been shown theoretically to have higher efficiency than its parent design, the case-only design has several restrictions preventing its widespread substitution for the original design.[9,69,70] In fact, this method was designed and promoted specifically as an alternative approach to estimation of the magnitude of interactions among various genes and environmental exposures.[194] The two most important disadvantages of this design are that the main effects of either genetic or environmental factors cannot be estimated and interaction is assessed on a multiplicative scale.[194] The latter limitation is not critical in gene-environment interaction studies because it is biologically plausible to assume that most such models exhibit extreme departures from multiplicativity.[194] However, the case-only method is limited to screening for gene-environment interactions when the individual effects of the risk factors have been previously estimated.[194]

### 2.3.6.3 Design considerations for studies to detect gene-environment interaction effects

As described in Section 2.3.6.1.2, it is largely accepted that in order to achieve adequate power, sample sizes for case-control studies of gene-environment interaction must be larger than sample sizes for studies of main effects alone.[23] However, Khoury et al. (based on the model developed by Hwang et al.) contend that, while conventional wisdom regarding increase in sample size is valid under a simple multiplicative model, it is not necessary to increase the size of a study for main effects in three special cases of genotype and environment interactions (Types 1-3 models of interaction discussed in Section 2.3.4). This contention is predicated on the difference in magnitude of



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interaction between a simple multiplicative model and the special cases of GEI, especially when considered in relation to main effects. Specifically, since, by definition, at least one or both of the odds ratios for main effects in the three biological models must be equal to one, the authors state that a departure from the multiplicative model of interaction would necessarily be very large proportional to the odds ratios for main effects alone. Therefore, they conclude that presence of large interaction effects increases study precision, obviating the need to increase sample size beyond that required for detection of main effects.[59,65] They further suggest that since at least one of the odds ratios for main effect is one (depending on type of model), and because the effect of interaction is so large as to obscure the effect of the other variable that independently contributes to disease, that it is sufficient to only consider the interaction effect in determining sample size.

The model developed by Hwang et al. has come under strong criticism for its overemphasis on the effect for interaction and for its lack of consideration to the possible role such a large effect may play on the influence of main effects.[62] It has been argued that presence of large interaction effects results in changes to other variables (such as the influence of main effects) in the model used to determine sample size. Thus, contribution of main effects to the model varies depending on presence or absence of interaction and cannot be ignored. Moreover, counter to the arguments set forth by Khoury et al., the a larger interaction effect will more likely exert a bigger influence on the sample size.[62]

Using their own formulae, Lubin and Gail [64] show that calculations of sample size and power depend on a covariance matrix under the $H_o$ that is a function of $\beta_o$, $\beta_1$, and $\beta_2$. The formulae are designed to use the maximum likelihood estimates (MLE) of $\beta_1$ and $\beta_2$



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

when the alternate hypothesis ($H_A$) is true. This method allows power (i.e. probability of rejecting the null when $H_A$ is true) to be determined using values most likely to be observed if $H_A$ is true, thus allowing more accurate estimations of sample size and power.[64]

The Hwang method has also been criticized with respect to the impact of differential misclassification on power and sample size of case-control studies. Theoretical applications of the method indicate that even small errors in correct classification of environmental or genetic factors result in biased interaction parameters, thus requiring a substantial increase in sample size.[60,61,63] In light of the apparent limitations of the case-control study to detect gene-environment interactions, the case-only study design has been proposed and validated as an alternative approach to traditional case-control studies.[58,70,194,198]



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER 3

# CASE-ONLY STUDY OF GENE-ENVIRONMENT INTERACTION EFFECTS IN PARKINSON'S DISEASE

## 3.1 Introduction

Idiopathic Parkinson's disease, characterized by resting tremor, muscle rigidity, slow body movements, and loss of balance, is a debilitating and increasingly common adult neurodegenerative disorder.[12-16] Although identification of risk factors that contribute to the pathogenesis of Parkinson's disease has remained a largely unmet challenge despite nearly two centuries of research, accumulating evidence from both observational and experimental research suggests that pesticides play an important role in the neurodegenerative process which ultimately results in idiopathic Parkinson's disease.[14]

Numerous epidemiologic studies have demonstrated an association between pesticides and PD,[25,26,106,122,158-167] with some being able to further narrow the association to specific classes of herbicides and insecticides.[26,158,159] Epidemiologic studies to date have been unable to link individual pesticides to PD. However, experimental studies have succeeded in indirectly implicating specific chemicals.[38,88] For example, carbamate fungicides enhance the neurotoxicity of the known parkinsonism inducing compound 1-methyl-4-phenylpyridinium (MPP+), the active metabolite of the chemical contaminant 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP).[36] Paraquat, a commonly used herbicide, the structure of which is almost identical to MPP+, has also been studied in relation to PD.[35] However, although one study showed a reduction of catecholamines in the midbrain of mice treated with paraquat, implying that it plays a role in the



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

neurodegenerative process, the potential involvement of this herbicide in the development of PD has not yet been clearly defined.[35,36] Several other pesticides have also been investigated in experimental studies as potential causative agents in PD. For example, higher concentrations of dieldrin, an organochlorine pesticide, have been found in the brains of PD patients post-mortem.[37] Additionally, mitochondrial impairment accompanied by parkinsonian-like symptoms have been shown to occur in rats treated with rotenone, a rodenticide.[43,174] However, the precise role of pesticides in the pathogenesis of PD has continued to elude researchers, leading to the hypothesis that pesticides most likely act in combination with certain genetic factors to cause PD.[12,54,76,84]

Structural variations in genes resulting in two or more common and distinct forms (alleles) are referred to as genetic polymorphisms.[178] Polymorphisms often alter the function of the gene leading to one or more different phenotypes. [178] Several polymorphic candidate genes have been investigated in gene-environment interaction studies of PD. [111,112] The genes of interest can be broadly classified according to the following characteristics: 1) genes involved in dopamine metabolism, such as dopamine receptors DRD2 and DRD4, the dopamine transporter DAT, dopamine metabolizing monoamine oxidases MAOA and MAOB, and catechol-O-methyltransferase (COMT); 2) genes involved in pathways of mitochondrial metabolism $tRNA^{Glu}$ and ND2 /id; 3) genes that provide cellular defense against oxidative damage related to the metabolism of endogenous toxic brain chemicals, such as detoxification enzymes glutathione-S-transferase M1 (GSTM1), NAD(P)H:quinone oxidoreductase 1 (NQO1), and (NRH):quinone oxidoreductase 2 (NQO2); 4) genes responsible for the detoxification of xenobiotics, including Cytochrome-P450 enzymes, N-acetyltransferase 2 (NAT2), and



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

glutathione transferase enzymes; and 5) other genes with a putative causal role in PD development, such as the lipoprotein gene APOE and estrogen receptors.[112] However, the results of most gene association studies have been inconsistent. A recent meta-analysis of fourteen individual gene polymorphisms in PD found significant associations only in polymorphisms involving NAT2, MAOB, GSTT1, and $tRNA^{Glu}$.[112] That study excluded Cytochrome-P450 enzyme polymorphisms for which separate meta-analyses had been performed.[108,129]

In gene-environment interaction studies of PD in which pesticides are the environmental risk factor under investigation, genetic variations leading to deficient metabolism in Phase I and Phase II detoxification enzymes, both of which can directly impact pesticide toxicity, have been of great interest. The Cytochrome P450 superfamily of mixed function oxygenases is the largest and most frequently studied group of Phase I biotransformation enzymes in PD. Numerous studies have been performed on various CYP450 enzyme polymorphisms including CYP1A1,[200-202] CYP2C9, CYP2C19, CYP1A2, and CYP2E1[128] with largely negative results.[112,128] However, among the many potential CYP450 genes, the highly polymorphic CYP2D6 gene is most closely linked to PD for several reasons. CYP2D6, one of few proteins occurring in functional form in the human brain, is thought to be involved in the metabolism of neurotransmitters such as dopamine. Moreover, CYP2D6 is responsible for the Phase I biotransformation of the only known parkinsonism-inducing neurotoxin MPP+, as well as many other pesticides. At least twelve distinct PM alleles have been identified that result in absent enzyme activity. [142] However, the PM phenotype is attributable to only five of the twelve mutations in more than 99% of Caucasians.[142] Moreover, of the five most common PM



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

alleles, a single mutation (CYP2D6*4) accounts for more than 95% of PM alleles in this population. CYP2D6 alleles representing the poor metabolizer phenotype are a result of autosomal recessive inheritance. It is estimated that 7-10% of Caucasians inherit two non-functional alleles due to CYP2D6 polymorphisms and are, thus, poor metabolizers (PM) of xenobiotics, with the rest of the Caucasian population falling into the extensive metabolizer category.[177] However, despite intensive scrutiny, relevance of this gene in the pathogenesis of PD remains undetermined, warranting further investigations.

In gene-environment interaction studies involving susceptibility genes, the ability to detect an association increases when a higher proportion of the study population is exposed. [130] Therefore, classic case-control studies of the association between CYP2D6 and PD must be designed either to ensure large enough sample sizes to avoid a Type 2 error or to select only patients with known exposure to the neurotoxin in question. [56,60,62,64] When the main research objective is determination of interaction effects, study efficiency can be significantly increased without a commensurate increase in sample size by utilizing the case-only study design. [58,197] A variation on the traditional case-control study, the case-only method follows the same principles as the parent design with the exception that the study is limited to cases only, thereby eliminating the need for a control group. [9,57,58,194,198] This method was developed as a more efficient technique to detect gene-environment interaction when the individual contributions of each risk factor are not of interest. [9,57,58,194,198] The case-only model follows the same epidemiologic principles of case selection as for any other case-control study. [9,57,58,194,198] This variant design differs from the case-control method from which it was derived in one important way: the association between an environmental exposure and a genotype is examined



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

among case subjects only (affected subjects with a given disease).[194] Cases are categorized into either a pseudocontrol or pseudocase group depending on absence or presence of susceptibility genotype, respectively.[9,194] Prevalence of environmental exposure (i.e., risk factor) is then compared among the two groups, with the pseudocontrols serving as the referent group.[9,194]

The goal of this study was to evaluate the use of the case-only study design to assess the magnitude of the combined effects of gene-environment interaction in PD using pesticide exposure as the environmental risk factor and CYP2D6 as the genetic risk factor. The specific aims of the project were to:

- Enlist cooperation from neurologists in private practice and recruit patients into study
- Develop structured questionnaire to assess pesticide exposure and apply to patient population
- Evaluate accuracy and degree of recall for past exposures to pesticides and environmental chemicals in patients with PD
- Determine CYP2D6 genotype among subjects using DNA isolated from buccal swabs
- Determine magnitude of the effect of interaction between pesticides and CYP2D6 in Parkinson's disease etiology
- Assess use of case-only study design in estimation of the magnitude of gene-environment interaction in the development of Parkinson's disease.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## 3.2 Methods

The case-only approach was used in this investigation because the primary purpose of the study was to evaluate the magnitude of association between genetic and environmental factors with respect to PD. This study design was selected because it is suitable for the investigation of a rare disease, such as PD, with respect to multiple sources of exposure while significantly reducing the need for as large a sample size as would be required for a conventional case-control study. [194,197] Moreover, by eliminating the control group, the case-only design reduces risk of spurious findings resulting from recall bias (caused by improper selection of controls) leading to differential misclassification, a historically problematic concern in studies of this type. [194]

The following assumptions had to be met in order to apply the case-only method to this study: 1) independence between pesticide exposure and the CYP2D6 genotype; 2) existence of background risk for PD (unrelated to either exposure or genotype); and 3) a multiplicative joint effect between genotype and environment (versus an additive effect).[197]

### 3.2.1 Study population

Low population prevalence of pesticide exposure leading to lack of study precision has resulted in ambiguous findings in most previous studies. [12-14] Since rural residence has repeatedly been shown to increase risk of PD, [82,103,161,164,170,203] this study attempted to increase efficiency by selecting a largely rural population at increased risk of pesticide exposure in two Colorado counties. Larimer and Weld counties in Northeastern Colorado were selected as the study site because they cover a predominantly rural area with a high proportion of Caucasian population, the only ethnic



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

group in whom the frequency of CYP2D6 PM genotype has been determined. [51] Urban metropolitan areas (>50,000) are also present in the selected counties, thus maintaining adequate representativeness among the study population. Eligibility was limited to living Caucasian residents (without dementia) of either Larimer or Weld counties with a confirmed diagnosis of idiopathic PD. Only primary respondents (no proxies) were used in an effort to minimize risk of exposure misclassification. [120] Eligibility criteria are described in the following section.

### 3.2.2 Diagnostic criteria

Patient eligibility was determined by participating neurologists using the study eligibility guidelines provided to each physician. The eligibility criteria required presence of two or more of the following cardinal signs: resting tremor, bradykinesia, disturbance of posture or equilibrium, or muscle rigidity. [160] Cases with the following diagnoses were excluded: essential tremor, any other apparent causes of parkinsonism (e.g., progressive supranuclear palsy, normal pressure hydrocephalus, multi-infarct dementia, Shy-Drager syndrome), encephalitis, Alzheimer's disease, head trauma, tumor, drug- or chemical-induced parkinsonism, young-onset PD (<50 years), [54,78] and dementia.

### 3.2.3 Subject recruitment

All individual and group neurology practices in Larimer and Weld counties were identified and invited to participate in the study. A total of ten neurologists at four independent clinics were approached. One clinic consisting of two neurologists refused to participate. The remaining three clinics agreed to cooperate by recruiting patients into the study at their own discretion. Clinic staff was instructed to select patients that met the study eligibility criteria provided, but the ultimate decision regarding eligibility and



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

presence or absence of dementia was left to the neurologists. One small practice with two neurologists recruited patients by mail. Eligible patients received a package directly from the neurology clinic including a cover letter from their physician, description of the study objectives, informed consent forms, and information on how to contact the investigators if interested in participating (Appendix 3-1). The remaining two clinics chose to recruit eligible patients during office visits by providing the same information verbally through a nurse who instructed the patients to call if interested. We were subsequently contacted by eighteen patients who were informed by mail, all of whom agreed to participate in the study. Three patients who were given information in person during their doctor's visit contacted us, and all three agreed to participate. None of the clinics agreed to provide us with total numbers of eligible patients; thus, we were unable to determine patient response rates. However, since almost all participants were those who received letters, and since the practice from which the letters were mailed had a much smaller patient base than the other two participating clinics, we concluded that notifying patients by mail rather than in person was more effective in eliciting participation.

### 3.2.4 Subject interview

Study participants were required to sign a written consent form before the interview and DNA sample collection. Participants were interviewed in person by the study investigator. Although subjects were informed briefly about the general objectives of the study, underlying hypotheses were not explicitly revealed in order to minimize potential respondent bias. A detailed structured questionnaire was developed to collect a



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

variety of exposure and other information (Appendix 3-2). To validate responses, some questions were repeated during the interview in a different way.

#### 3.2.4.1 Personal information

Demographic data obtained included date of birth, place of birth, gender, ethnic background, marital status, education level attained, occupation, and annual family income. A complete residential history beginning with their first ever residence up to the present was obtained. The name of the city, county, and state, dates and duration, water supply, and description of type of community (i.e., rural, agricultural, suburban, urban) with approximate population size was collected for each address. Presence of a farm or farmland on or adjacent to each place of residence was determined, along with use of specific types of pesticides on each farm or farmland. Other residential information obtained included presence of a garden, types of pesticides used on the garden, if any, presence of animals inside or outside the home, and types of insecticides used in the home or on animals.

Information about family history included the following: patient's birth order, parent's ancestry and heritage, country of birth of each parent, living status of each parent, number and gender of siblings, living status of siblings, number and gender of children, living status of children, and cause of death for each deceased parent, sibling, or child. Additionally, family history of PD was established by determining if any parent, sibling, child or other relative had ever had a diagnosis or showed any signs of PD.

A detailed personal health history was also obtained including number of amalgam fillings and number of times the subject underwent general anesthesia before



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PD diagnosis, age when symptoms first appeared, age at diagnosis, and smoking history. Finally, exposure to chemicals in home activities or hobbies was assessed.

#### 3.2.4.2 Occupational information

A detailed lifetime work history including job titles, type of industry, dates of employment, details of tasks performed, workplace exposure to chemical agents (including duration, timing, and peaks), specific types of chemicals used, use of protective equipment, and percentage of time wearing protective gear, was collected for all jobs held for at least one year from age fifteen to one year prior to diagnosis. The questionnaire was designed to obtain information on a variety of potentially neurotoxic environmental exposures including heavy metals, wood preservatives, solvents, gases and vapors, glues, paints, and lacquers, exhaust fumes, carbon monoxide, and, particularly, pesticides.

##### 3.2.4.2.1 Pesticide exposure

Exposure to all classes of pesticides has been studied in relation to PD, and although specific pesticides, such as paraquat, rotenone, and organophosphates, have shown the strongest association among all pesticides, other chemical groups cannot be ruled out as potential risk factors. Therefore, information on exposure to all pesticides was collected. Attempts were made to determine the specific class, type, and when possible, name of chemicals used. Respondents were asked to separately recall insecticide, herbicide, fungicide, and fumigant use. Additionally, subjects were presented with names of the most commonly used chemicals in each pesticide category and asked to recall the use of each individually. Start and stop dates for all jobs involving pesticide use and details about the nature of the work, such as use of protective equipment and



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

method of application, were recorded. The specific purpose for pesticide application was determined for each group of pesticides reported, such as whether the pesticides were used on a farm, and if so, what type of farm.

#### 3.2.4.3 DNA sample collection

DNA specimens were collected at the time of the interview using a simple, non-invasive self-collection protocol as described in previous studies. [142,149] Briefly, subjects were directed by the interviewer to self-collect epithelial cells by firmly rolling the buccal brushes provided (MasterAmp Buccal Swabs, Epicentre Technologies, Madison, Wisconsin) on the inside of the cheek, approximately 20 times on each side, after thoroughly rinsing the mouth. Samples were transported to the laboratory at room temperature where they were processed within 24 hours and stored at -70°C.

### 3.2.5 Genotyping

Following extraction of genomic DNA from collected samples, nested PCR assays were used to detect the two most common CYP2D6 null alleles responsible for more than 95% of the poor metabolizer (PM) phenotypes in Caucasians. [142,149] All procedures, described in detail in the previous chapter, were performed at Colorado State University Department of Environmental Health toxicology/molecular biology laboratory.

## 3.3 Results and analysis

Twenty-one Caucasian patients with a neurologist-confirmed diagnosis of idiopathic PD were studied in this case-only study. Questionnaire responses were first edited and checked for accuracy and completeness before being coded and recorded



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

electronically. All variables were checked for expected distributions, as well as for any unusual values, during the data entry process.

The mean age of cases was 77 years (range, 66 to 91 years). Mean age at diagnosis was 69 years (range, 51 to 82 years), with a mean estimated duration of disease of 8 years (range, <1 to 21 years). There were twice as many males enrolled (14) as females (7), a much higher proportion than is estimated for the general population. Male and female cases did not differ significantly in mean age (77 years vs. 75 years). Although age at PD diagnosis in females was similar to their male counterparts (69 years vs. 70 years), the reported duration of diagnosis was higher in males (9 years) than in females (5 years). A large proportion of subjects had attained some college education (33.3%) with education level ranging from some high school to post graduate/professional degrees. More than 95% of cases reported an annual family income over $21000, more than 90% were either currently married or widowed, and all were white of Northern European descent. Table 3-1 provides a detailed summary of the descriptive data.

Exposure to pesticides as well as other risk factors suspected in PD was evaluated in all cases as summarized and presented in Table 3-2. As can be seen, a large proportion of subjects (76.2%) resided in a rural area for at least one year, and 71.4% drank well water on the residence. Moreover, 66.7% reported at least one year of residence on a farm on which all 14 reported working at the time of residence. Eighty-one percent of subjects maintained a garden for one year or longer. Approximately half (52.4%) of the subjects reported use of pesticides for home, lawn, or garden; 28.5% reported professional application of pesticides. Nineteen percent of subjects were farmers by



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

occupation. Exposure to chemical fertilizers was reported in 47.6% of subjects as compared to 33.3% who used natural fertilizers. Reported occupational exposure to other chemical agents included carbon monoxide (27.3%), mercury (18.2%), gasoline (36.4%), and solvents (9.1%). Only 33.3% reported smoking at least 100 cigarettes during their lifetime, and none was a current cigarette smoker. Less than 5% reported use of other tobacco products. Eighty-one percent had undergone anesthesia, and 61.9% had received amalgam fillings prior to PD diagnosis. No subjects had siblings with PD; however, two subjects (9.5%) reported a parent with PD, and 33.3% reported another relative (grandparent) with PD.

Additionally, subjects were assessed for the presence of the two most common CYP2D6 gene polymorphisms resulting in the poor metabolizer (PM) phenotype (CYP2D6*4 and CYP2D6*5). Since both CYP2D6*4 and *5 PM alleles are recessive, results must be interpreted as follows: samples containing two wild-type alleles (homozygous for the wild-type allele) are classified as expressing the normal or extensive metabolizer (EM) phenotype, samples containing one wild-type and one mutated allele (CYP2D6*4) indicate a heterozygous genotype and are likewise categorized as expressing the normal (EM) phenotype, samples with two mutated (CYP2D6*4) alleles (homozygous for the mutated allele) are considered to express the PM phenotype and are thus poor metabolizers. Finally, samples lacking either a wild-type or a mutated allele were considered to represent the CYP2D6 *5 genotype resulting in the poor metabolizer phenotype due to the complete deletion of the CYP2D6 gene.

Genotype analysis revealed either heterozygous or homozygous for the wild-type allele indicating all patients tested were CYP2D6 normal extensive metabolizers (EMs)



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(Table 3-3). Hence, calculation of appropriate risk estimates (odds ratios) to determine the effects or magnitude of the interaction between pesticides and CYP2D6 PM individuals was not possible.

Our data suggest a high proportion of potential exposure to pesticides among the study group. However, due to the absence of a control group for this study, we were unable to directly estimate the risk of pesticide exposure for this study. Therefore, we attempted to evaluate the risk of pesticide exposure among our subjects indirectly (21 PD patients) by using a historical control group with similar characteristics for comparison. Searching the literature, we found a recently published death certificate-based study of pesticides and PD among rural residents in an agricultural area of California similar to our study area.[204] However, exposure information among the control group was not described explicitly, thus, not allowing any analysis. Likewise, none of the 10 remaining studies conducted in the United States proved suitable for comparison, either because the study population was urban and, therefore, not similar to our study population, or due to inadequate information available on the control subjects.

We identified a 1994 Canadian study conducted by Hertzman et al. in a horticultural region of British Columbia that provided adequate exposure information for the control group.[162] Although the study was not conducted in the United States, the geographically defined agricultural and stable rural population described in the Hertzman study shared enough similarities to our subjects to warrant a comparison. Moreover, exposures were classified exactly as was done in our study, further facilitating comparison between the studies. Two control groups were used in the Hertzman et al. study, the first group consisting of 124 individuals between the ages of 45 and 80 years



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

who were randomly selected from electoral rolls, and the second, a group of 121 patients with chronic cardiac disease recruited from physician's offices. Analyses were performed using EpiInfo 2002, a public domain software program from the Centers for Disease Control and Prevention (CDC). The following exposure categories were examined: 1) pesticides, 2) farming, 3) gardening, and 4) well water consumption. All exposures were categorized as dichotomous variables and crude (unstratified) univariate analysis was performed to calculate odds ratios with 95% confidence intervals (Table 3-4). Comparisons were made between our subjects and all controls as well as between each control group separately. None of the exposure categories was found to be statistically significantly associated with PD with the exception of gardening based on Control Group 1 (OR=5.83, 95% CI: 1.70, 21.93). The odds ratios for ever versus never pesticide exposure were increased overall (OR= 1.52, 95% CI: 0.50, 4.46) as well as for each control group ($OR_{ControlGroup1}$=1.39, 95% CI: 0.43, 4.33; $OR_{ControlGroup2}$=1.67, 95% CI: 0.51, 5.24) and the 95% confidence intervals were heavily skewed to the right, suggesting a possible association between pesticides and PD. However, all confidence intervals included one (no statistically significant associations), possibly due to lack of sufficient power; thus, no conclusions can be made from this analysis.

### 3.4 Discussion

Structural variations, or polymorphisms, in genes encoding for detoxification enzymes cannot lead to development of PD by themselves. Rather, such genetic differences can only increase susceptibility to PD in the presence of pesticide exposure. Therefore, it is important to study interactions between the two potential risk factors rather than studying each alone. Surprisingly, despite compelling evidence of a complex



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and multifaceted etiology of PD involving both environmental and genetic factors, most studies have been divided along genetic and environmental lines, with few investigations evaluating the interaction between susceptibility genes such as CYP2D6 and environmental factors such as pesticide exposure. Only a single study has been published to date has attempted to examine the interaction between a Phase I enzyme (CYP2D6), two Phase II enzymes (GSTM1 and GSTT1), pesticides, and PD.[127] The 1998 hospital-based case-control study by DePalma et al. recruited 100 PD patients from an outpatient neurology clinic affiliated with a university hospital in Parma, Italy. [127] Two hundred controls were enrolled from other outpatient clinics at the same hospital. [127] Patients were interviewed to determine exposure to environmental risk factors. Rural residence, well water consumption, exposure to copper-sulfate, and exposure to pesticides were significantly associated with PD. [127] DNA samples were obtained from subjects although the source of DNA was not specified. [127] No significant associations were found for CYP2D6 genotype alone or for interaction of pesticides and CYP2D6. [127] A significant association was reported for the interaction of CYP2D6 and solvent exposure based on a limited number of subjects (5 cases, 14 controls). [127] A significant association was reported with the GSTT1*0 genotype but nothing was mentioned in regards to GSTT1*0 interactions with environmental risk factors.[127]

An important weakness in the DePalma et al. study was the lack of an adequate number of subjects resulting in a significant reduction in statistical power, which in turn prevented meaningful analysis. Demographic information on study population was not provided, raising the possibility that study efficiency may have also been compromised if



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

very few subjects recruited were exposed to pesticides and/or had the CYP2D6 genotype among the selected population.

We attempted to expand and improve on the study of DePalma et al. by selecting our subjects from a population with a high prevalence of agricultural pesticide exposure. Our population also consisted only of Caucasians in whom the CYP2D6 poor metabolizer genotype is reported to occur with 7-10% frequency.

### 3.4.1 Study design

PD is a relatively rare disease with a long latent period between subclinical damage and the appearance of clinical symptoms,[48] and, as such, is highly suitable to the classic case-control study design. For purposes of this study, however, the case-only design, a recently developed modification to the traditional case-control design model was used. The case-only design has been shown to be a more efficient and cost effective method for investigating gene-environment interactions because it produces more precise estimates of interaction than the case-control design.[194,197] The method differs from the traditional case-control approach in that the study population is comprised only of subjects with the disease of interest, and uses an internal control group for analysis. [197] However, the case-only design has two major limitations; 1) it does not allow evaluation of an association between individual variables and disease; and 2) possible joint additive effects cannot be ascertained by this method. [197]

Our study provides an example of the first limitation of a case-only study described above. Specifically, we were unable to evaluate the effects of gene-environment interaction due to inadequate sample size resulting in the absence of the susceptibility genotype in all of our subjects. The frequency of the all PM phenotypes in



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the Caucasian population has been estimated to be between 7-10%.[177] Thus, we would expect 1-2 individuals from our group of 21 cases to have the PM phenotype. However, this expectation was not borne out by the analysis. Moreover, due to the limitation of the case-only design, we were unable to assess the individual effects of each risk factor alone, preventing any direct measurement of effect in this study despite the fact that information regarding other environmental exposures was collected, the results of which could have been used to examine their potential contribution in PD had a traditional case-control design been used. The case-only design did not allow direct analysis of the detailed exposure data collected since there were no controls available. However, we used a historical control group from the literature to evaluate the effects of environmental exposure on PD. The analysis showed a significant association between gardening and PD;no significant associations were found with exposure to pesticides, farming, or well water consumption (Table 3-4).

### 3.4.2 Study size and power

Based on the estimated PD prevalence rate of 150/100,000, and population estimates for Larimer and Weld counties from 1997-8, the expected number of Caucasian PD cases (as ascertained by ICD-9 codes) over a 2-year period was approximately 600. We attempted to enroll 30 cases based on the calculations that follow.

Yang et al. developed a method to estimate the required number of cases in a case-only design to detect gene-environment interaction.[197] Sample size can be calculated using their method if the following parameters are specified: prevalence of exposure ($P_e$), prevalence of genotype ($P_g$), relative risk for exposure alone ($OR_{Eg}$),



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

relative risk of genotype alone ($OR_{Ge}$), and the estimated effect of gene-environment interaction ($OR_I$), as well as probabilities for type I and type II errors ($\alpha$ and $\beta$).

The population frequency of the CYP2D6 poor metabolizer genotype was estimated at 7% based on the reported prevalence of 7-10% among Caucasians.[51,140] Because subjects were selected from rural areas and, thus, were at higher risk of pesticide exposure, we assumed an exposure prevalence of 20% in the study population. Pesticide exposure has been shown to increase the risk of PD by greater than 2-fold in numerous studies; therefore, we assumed the relative risk of exposure of 2 in our calculations. Although studies of CYP2D6 polymorphisms indicate an increased risk of PD among poor metabolizers, due to the controversial and unsettled nature of the available data, we assumed the relative risk of genotype to be unity in order to obtain the most conservative estimate. Currently, no estimate of the effect and magnitude of interaction between pesticide exposure and CYP2D6 polymorphisms exists from which to estimate relative risks. Assuming that our hypothesized biological model of gene-environment interaction is multiplicative and large, we postulated an odds ratio of 10 for interaction ($R_I$). Using the above values, from the table provided by Khoury et al. based on the equation given by Yang et al., a sample size of 30 would be required for a study with 70% power at the 5% level of significance.

Only 21 patients were enrolled during the limited study period due to case recruitment restrictions discussed in the following section. This value was below the target of thirty cases. The major limitation in evaluating gene-environment interaction was the lack of any cases with the CYP2D6 poor metabolizer polymorphism, resulting in uniformly normal phenotype in all subjects and preventing analysis.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

### 3.4.3 Case recruitment

There are no centralized registries or databases from which to obtain a PD patient list. Moreover, patient records in neurologists offices cannot be accessed without prior consent from the patient. Despite these limitations, we attempted to recruit subjects using an accurate and consistent case definition of idiopathic PD in order to minimize potential diagnostic bias. We approached all neurologists in the two-county area for participation in the study through their identifying and enrolling eligible patients based on the uniform diagnostic criteria we provided. Four out of the five clinics agreed to participate under the condition that method of patient recruitment would be left to the discretion of each clinic. Three clinics chose to recruit eligible patients on an ongoing basis at the time of the patient's appointment while one clinic sent letters to all their eligible patients. Almost all (18 out of 21) patients enrolled were those who received letters from the single clinic although the other three clinics were much larger. Overall, with the exception of one neurologist who was genuinely interested in the research objectives and by whom most of the cases were referred, the physicians contacted were reluctant to participate and unwilling to spend extra time or effort to collaborate in the research study. Hence, we were unable to recruit the targeted number of cases.

### 3.4.4 Exposure assessment

The strength and credibility of epidemiologic studies of environmental risk factors rely largely on valid and precise exposure assessment, yet myriad methodological issues can result in obscure and inaccurate exposure assessment as described in numerous epidemiologic studies.[14] Lack of standardized, validated methods can further complicate classification of exposures. Nevertheless, several major methods are currently used to



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

assess environmental exposure, all with unique benefits and limitations.[120] Three major exposure assessment approaches were considered for this study: the self-report technique, use of an expert industrial hygienist, and use of a job exposure matrix (JEM). The use of an industrial hygienist was considered infeasible given budget restrictions in this study. The JEM method was also disregarded due to the very low specificity resulting from false-positive exposure classification associated with the technique.[120] Hence, exposure information for this study was obtained by the self-report method. A comprehensive questionnaire was developed to assess exposure to specific categories and classes of pesticides as well as to individual chemicals used as pesticides. Further, the questionnaire was designed to determine the timing and duration of each exposure in detail. However, although all cases interviewed were non-demented PD patients, the exposures reported were too far back in time for respondents to recall with any precision. Subjects were able to recall the pesticide categories (i.e., type of application) to which they were exposed but were unable to identify individual chemicals when prompted. Likewise, there was poor recall of dates and durations of pesticide use. Our findings indicate that the self-report method may result in incomplete and possibly inaccurate exposure assessment for studies of diseases with long latency period such as PD. Therefore, the self-report method is recommended only in conjunction with another form of exposure assessment, such as the use of an industrial hygienist.

### 3.5 Summary and future research direction

The results from this study provide important information regarding the challenges involved in studies of multifactorial diseases of long duration. Importantly, our experience underscores the significance of ensuring adequate sample size and power



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to detect the desired association. Specifically, our results suggest that the success of such studies depends largely on the full cooperation of clinical neurologists willing to recruit an adequate number of subjects. Moreover, incompleteness of the exposure information obtained directly from subjects in this study indicates that exposure may perhaps be best determined using two or more assessment methods for cross-verification.

The findings from this pilot project present an opportunity for future studies to develop improved study designs with regard to exposure assessment and patient recruitment methods. Importantly, larger studies are necessary to examine, not only the association of CYP2D6 and pesticides, but also other gene-environment and gene-gene interactions. For example, interactions between other Phase I biotransformation enzymes such as CYP1A1, CYP2C9, CYP2C19, CYP1A2, and CYP2E1 and various environmental chemicals including pesticides remain unresolved. Additionally, other susceptibility genes including those encoding for Phase II biotransformation enzymes such as GSTT1 (responsible for detoxification of free radicals from nicotine metabolites), as well as genes involved in oxidative damage and mitochondrial dysfunction, should be examined. MAO-B is another important polymorphic gene responsible for metabolizing nigrostriatal dopaminergic neurons, as well as transformation of MPTP to the active metabolite MPP+. Lastly, further studies of autosomal dominant genes that occur in a small proportion of PD cases including alpha-synuclein and parkin could perhaps help to elucidate other underlying mechanisms contributing to development of PD



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 3-1. Demographic characteristics of the study population.**

| Variable | Categories | Number | Percent |
|---|---|---|---|
| Age | 61-70 | 3 | 14.3 |
| | 71-80 | 10 | 47.6 |
| | 81-90 | 7 | 33.3 |
| | >91 | 1 | 4.8 |
| Gender | Male | 14 | 67.7 |
| | Female | 7 | 33.3 |
| Race/ethnicity | White | 21 | 100 |
| Marital status | Married | 13 | 61.9 |
| | Widowed | 7 | 33.3 |
| | Divorced | 1 | 4.8 |
| Education | Some high school | 2 | 9.5 |
| | High school graduate | 5 | 23.8 |
| | Some college | 7 | 33.3 |
| | College graduate | 1 | 4.8 |
| | Graduate/professional | 4 | 19.1 |
| | Vocational training | 1 | 4.8 |
| Income | <$20000 | 1 | 4.8 |
| | $21000-40000 | 10 | 47.6 |
| | $41000-80000 | 10 | 47.6 |
| Occupation | Farmer | 4 | 19.1 |
| | Pesticide applicator | 1 | 4.8 |
| | Teacher/college professor | 3 | 14.3 |
| | Homemaker | 4 | 19.1 |
| | Office worker | 5 | 23.8 |
| | Metal worker | 1 | 4.8 |
| | Appliance repairman | 2 | 9.5 |
| | Artist | 1 | 4.8 |
| Age at PD diagnosis | 51-60 | 4 | 19.1 |
| | 61-70 | 7 | 33.3 |
| | 71-80 | 8 | 38.1 |
| | > 81 | 2 | 9.5 |
| Duration of diagnosis | < 5 years | 6 | 31.6 |
| | 6-10 years | 9 | 42.9 |
| | 11-15 years | 2 | 9.5 |
| | > 16 years | 4 | 19.1 |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 3-2. Summary of risk factors in study population.**

| Variable | | | Number | Percent |
|---|---|---|---|---|
| Drank well water at residence | | | 15 | 71.4 |
| Lived in rural area (1 year or more) | | | 16 | 76.2 |
| Lived on farm (1 year or more) | | | 14 | 66.7 |
| Worked on farm (1 year or more) | | | 14 | 66.7 |
| Kept garden (1 year or more) | | | 17 | 81. |
| Farmer by occupation | | | 4 | 19.1 |
| Used pesticides (1 year or more) | Home/lawn/garden | | 11 | 52.4 |
| | Professional application | | 6 | 28.5 |
| Used protective gear during application (n=17, 81%) | Gloves | | 7 | 41.2 |
| | Mask | | 1 | 5.9 |
| | None | | 10 | 58.8 |
| Used natural fertilizer | | | 7 | 33.3 |
| Used chemical fertilizer | | | 10 | 47.6 |
| Other chemical/metal exposure (n=11, 52.4%) | Carbon monoxide | | 3 | 27.3 |
| | Metals | Mercury | 2 | 18.2 |
| | | Zinc | 1 | 9.1 |
| | | Aluminum | 2 | 18.2 |
| | Paint/lacquer | | 3 | 27.3 |
| | Gasoline | | 4 | 36.4 |
| | Solvents | Acetone | 1 | 9.1 |
| | | Ethyl acetate | 1 | 9.1 |
| | | Freon | 1 | 9.1 |
| | | MEK | 1 | 9.1 |
| | | Drycleaning | 1 | 9.1 |
| Smoked cigarettes | | | 7 | 33.3 |
| Used other tobacco | | | 1 | 4.8 |
| Had anesthesia before PD diagnosis | | | 17 | 81. |
| Had amalgam fillings before PD dx | | | 13 | 61.9 |
| Had parent with PD (mother) | | | 2 | 9.5 |
| Had sibling with PD | | | 0 | 0 |
| Had other relative with PD | Paternal grandmother | | 4 | 19.1 |
| | Paternal aunt | | 1 | 4.8 |
| | Maternal aunt | | 1 | 4.8 |
| | Paternal cousin | | 1 | 4.8 |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 3-3. Genotype analysis and phenotype prediction in study population.**

| Number and percent of samples | PCR result interpretation | Predicted genotype | Predicted phenotype |
|---|---|---|---|
| 6 (31.6%) | wt/wt | Homozygous normal | EM |
| 13 (68.4%) | wt/*4 | Heterozygous | EM |

**Table 3-4. Crude odds ratios of PD risk by selected exposures using historical controls for comparison.**

| Exposure | Cases (n=21) | All controls (n=245) | OR (95% CI) | Control Group 1[1] (n=121) | OR (95% CI) | Control Group 2[2] (n=124) | OR (95% CI) |
|---|---|---|---|---|---|---|---|
| Pesticides | 6 | 51 | 1.52 (.50, 4.46) | 27 | 1.39 (.43, 4.33) | 24 | 1.67 (.51, 5.24) |
| Farming | 14 | 193 | 0.54 (.19, 1.56) | 91 | 0.66 (.22, 2.01) | 102 | .43 (.14, 1.35) |
| Gardening | 17 | 145 | 2.93 (.89,10.64) | 51 | 5.83 (1.70,21.93) | 94 | 1.36 (.39, 5.19) |
| Well water | 15 | 181 | 0.88 (.30, 2.68) | 88 | 0.94 (.31, 2.98) | 93 | 0.83 (.27, 2.65) |

[1] Control Group 1: 124 individuals between the ages of 45 and 80 years who were randomly selected from electoral rolls.

[2] Control Group 2: 121 patients with chronic cardiac disease recruited from physician offices.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER 4

# DETERMINATION OF THE CYTOCHROME P450 2D6*4 AND *5 POOR METABOLIZER GENOTYPES BY POLYMERASE CHAIN REACTION USING EPITHELIAL DNA FROM PARKINSON'S DISEASE PATIENTS

## 4.1 Introduction

The liver enzyme cytochrome P450 CYP2D6 (debrisoquine/sparteine hydroxylase) mediates the oxidative metabolism of many important drugs and environmental chemicals.[52,175] CYP2D6 is characterized by high genetic variability with more than 60 different CYP2D6 alleles described to date, leading to a broad spectrum of enzyme activity, from complete inactivity to ultra-high activity.[140] Approximately 7-10% of Caucasians in Europe and the United States completely lack CYP2D6 enzyme activity due to mutations in both alleles of the gene and are thus classified as poor metabolizers (PM's).[51,142] At least twelve distinct PM alleles have been identified that result in absent enzyme activity.[142] However, the PM phenotype is attributable to only five of the twelve mutations in more than 99% of Caucasians.[142] Moreover, of the five most common PM alleles, a single mutation (CYP2D6*4) accounts for more than 95% of PM alleles in this population.[132]

Variability in metabolism of drugs and other exogenous substances is an important predictor of individual response to potentially harmful toxins.[53,175,177] Gene mutations leading to deficient excretion of exogenous chemicals can increase susceptibility to environmentally related diseases.[53,175,177] Thus, CYP2D6 has been studied extensively in relation to its role in the etiology of diseases whose pathogenesis is also associated with exposure to environmental chemicals.[51] In particular, much



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

emphasis has been placed on the possible relationship between CYP2D6 PM's and Parkinson's disease (PD). [51,55,108,127-129,132-137,139,143,179] Results of early investigations relying on phenotyping methods to assess the possible association between CYP2D6 PM's and PD were largely inconclusive, likely due to methodologic problems characteristic of phenotypic studies such as inability to control for intraindividual variability and other confounding factors.[177] More recently, genotyping methods allowing direct determination of the genetic diversity in CYP2D6, and thus, individual metabolic capacity, have emerged as an alternative to traditional phenotyping methods. [182] Subsequently, the association between CYP2D6 polymorphisms and susceptibility to specific diseases such as PD has been examined using genotypic studies as a more objective measure of genotype, eliminating intraindividual variability while providing more rapid results. [182] Although numerous laboratory approaches have been used to determine CYP2D6 polymorphisms, most research effort has been placed on development and optimization of the simplest method: polymerase chain reaction (PCR)-based assays. [182] Moreover, independent laboratories have reported developing multiplex PCR reactions to simultaneously detect the five most common PM-associated CYP2D6 alleles representing greater than 99% of Caucasians with the PM phenotype. [142,150,205-207] However, the multiplex PCR assays may have certain limitations that must be considered. First, multiplex assays developed to date require the use of whole blood as a source for genomic DNA, the collection of which can be difficult in epidemiologic studies. [142,150,205-207] Buccal cells have recently emerged as the preferred source for DNA collection in genetic epidemiologic studies because they minimize expense, exposure to blood borne pathogens, and patient distress, thus potentially facilitating participation. [150,180] Yet the



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

utility of multiplex reactions has not been evaluated using this or other alternate DNA sources. Second, attempts to duplicate results of the reported techniques by other researchers have been unsuccessful,[150] perhaps suggesting that intrinsic flaws in particular methodologies developed in one molecular biology laboratory prevent their adaptability in other laboratories.

The research objectives of this study were to assess the performance of the multiplex PCR assays previously developed to detect the most common PM mutations using DNA derived from buccal cells and to develop a PCR assay to detect the CYP2D6*4 allele representing the biggest proportion of poor metabolizers. Additionally, we evaluated the performance of our PCR protocol by first determining the CYP2D6 *4 allele genotype in healthy control subjects using DNA derived from whole blood, then sequencing the relevant segment of the product obtained from the PCR for verification.

The specific aims of the study were as follows:

- Evaluate quantity and quality of DNA extracted from buccal swabs for amplification of large (>4 kb) gene segments
- Determine utility of polymerase chain reaction (PCR) in genotyping CYP2D6, a gene downstream from two highly homologous pseudogenes using buccal cells as source for genomic DNA
- Design PCR-based assay to detect the CYP2D6 *4 and *5 mutations using buccal cells as source for genomic DNA
- Verify CYP2D6 PCR results by sequencing gene product



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## 4.2 Subjects, materials, and methods

### 4.2.1 Subjects

A total of twenty-seven individuals participated in the study. They included twenty-one patients (14 males and 7 females) with PD enrolled in a gene-environment interaction study and six healthy volunteers. Subjects were unrelated Caucasian residents of Northern Colorado. All volunteers gave written informed consent to genotyping, and the procedure was approved in advance of the study by the University Human Research Committee. Among PD patients, mean age was 77 years in males (range 66 to 91 years) and 75 years in females (range 66 to 82 years). Healthy volunteers serving as control subjects included 3 males and 2 females whose ages ranged from 35 to 45 years.

### 4.2.2 DNA Source

Dry buccal brush samples were provided to PD patients for self-collection of epithelial DNA (MasterAmp Buccal Swabs, Epicentre Technologies, Madison, Wisconsin). Patients were instructed to thoroughly rinse their mouths with water prior to cell collection. In addition, patients were recommended to abstain from drinking coffee the morning of tissue collection. DNA was collected by rolling the buccal brush firmly on the inside of the cheek, approximately 20 times on each side. Brushes were allowed to air dry for 10-15 minutes at room temperature, after which they were transported to the laboratory and stored in the original packaging a maximum of 24 hours before extraction.

Buccal cells for control samples were collected in the same manner from healthy volunteers participating in the study. In addition, whole peripheral blood was also collected from healthy volunteers by a trained phlebotomist and extracted according to protocol described in the Epicentre MasterPure DNA Purification Kit.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 4.2.3 Buccal DNA Extraction

DNA was prepared using the Epicentre MasterAmp Buccal Swab DNA Extraction Solution. Swabs containing epithelial samples were placed directly into a tube containing the QuickExtract DNA Extraction Solution. Samples were then hand vortexed for 15 seconds followed by heat treatment for 30 minutes at 65°C to lyse epithelial cells and to degrade compounds inhibitory to amplification. Samples were hand vortexed again for 15 seconds and heated for 8 minutes at 98°C. The last step was repeated and centrifugation was used to pellet and remove cellular debris following heat treatment. The supernatant was then stored at –70°C until use.

#### 4.2.4 Buccal DNA concentration

Fluorometric evaluation of DNA yield indicated a wide range of concentrations from 96-7550 ng/ml from buccal cells. As expected, DNA concentrations from blood samples were significantly higher ranging from 3150-202160 ng/ml.

#### 4.2.5 CYP2D6 Genotype determination

##### 4.2.5.1 Long distance and multiplex polymerase chain reaction

Simultaneous detection of the five most common CYP2D6 null alleles has been described in several reports.[142,150,205-207] We examined the utility of the multiplex PCR method as described by Stuven et al using buccal cell-derived genomic DNA as the template.[142] Primers used for the assay were the same as those developed for the Stuven assay and are presented in Table 4-1.[142] The PCR method is described below.

###### 4.2.5.1.1 Preamplification

The CYP2D6 gene is downstream of two highly homologous pseudogenes (CYP2D7 and CYP2D8) in both of which several CYP2D6 mutations have also been



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

identified. Therefore, in order to reduce the likelihood of false positive or false negative detection of mutations by co-amplification of the pseudogene sequences, a 4666 bp fragment of CYP2D6 containing all nine exons is preamplified using primers complementary to unique CYP2D6 intron sequences.[142]

The preamplification method was performed exactly as described elsewhere.[142] Additionally, the ability of the preamplification assay to function in a range of PCR conditions was evaluated. Specifically, emphasis was placed on optimization of genomic DNA, $MgCl_2$, and primer concentrations as well as use of appropriate annealing temperatures. The modified protocols are described below. Total reaction volumes were 25 ul. Each reaction contained 25 or 50 mM MgC12, 0.1, 0.2, or 0.3 uM each of primers 1-5 and n, 100, 200, or 300 ng of genomic DNA, 200 uM of each dNTP, 2.5 U of Platinum Taq Polymerase (Invitrogen) or 2.5 U of Taq polymerase (Invitrogen), and 10X Reaction Buffer (Invitrogen). Reaction mixtures were overlaid with 10 ul of mineral oil to minimize evaportation and loss of sample during PCR cycling. Amplifications were performed using a Perkin Elmer Gene Amp 9600 Thermal Cycler. An initial denaturation step at 94°C for 3 minutes was followed by a range of 15-35 cycles of 94° C for 0.5 or 1 minute, 55 °-75 ° C for 0.5 minutes (annealing), and 72° C for 4.5 minutes (DNA synthesis). A terminal extension at 72° C for an additional 7 minutes completed the thermal cycling process. PCR products were verified by electrophoresis on 1% agarose gels containing 0.05% ethidium bromide and visualized on an Alpha Innotech AlphaImager 2200.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 4.2.5.1.2 Multiplex allele-specific PCR

According to the protocol outlined by Stuven et al., the preamplification product is used as the template in two separate PCR reactions with either the 5 wild-type or 5 mutation-specific primers to detect CYP2D6*3, *4, *7, *6, and *8 mutations. [142] Presence of CYP2D6 *5, the second most frequent mutation leading to complete gene deletion, can additionally be determined if there is absence of any product.

The multiplex allele-specific method developed by Stuven el al. was optimized with minor parametric modifications. Using the 4.7 kb CYP2D6-specific preamplification product as template, two separate reactions were performed for each DNA sample using wild-type and mutation-specific primers. Reactions contained 25 mM $MgCl_2$, 0.5 ul of preamplification product, 0.1 uM of each primer, 200 uM of each dNTP, 2.5 U of Platinum Taq Polymerase (Invitrogen), and 10X Reaction Buffer (Invitrogen) in a final volume of 25ul. Reaction mixtures were overlaid with mineral oil. Amplifications were performed on a Perkin Elmer Gene Amp 9600 Thermal Cycler. Thermal cycling was conducted with an initial denaturation step at 94°C for 3 minutes was followed by 25 cycles of 94° C for 1 minute, 55° C for 1 minute (annealing), and 72° C for 3 minutes (DNA synthesis). PCR products were checked as described above.

#### 4.2.5.1.3 Results

We were unable to generate the long CYP2D6 product either by following the original protocol or after repeated attempts at optimization of the original protocol. Likewise, attempting to use the long PCR product as a template in the nested mutliplex reactions yielded no results. Using information from a recently published report which found that buccal cells are not suitable for amplification of long PCR products,[150] we



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

designed a new PCR protocol to detect CYP2D6 *4, the most common CYP2D6 null allele. By default, the procedure would also detect CYP2D6 *5, the second most common null allele resulting in complete gene deletion, therefore, not producing any detectable product on electrophoresis. The new assay, involving a two-step PCR in order to prevent cross-reaction with pseudogenes, is presented in detail below.

#### 4.2.5.2 Long PCR of CYP2D6 fragment containing *4 mutation

We designed primers to preamplify a 1018 bp fragment of the CYP2D6 gene containing the *4 mutation (Table 4-2). Forward and reverse primers were designed to specifically complement sequences on the CYP2D6 gene only. The optimum length of the primers was determined to be 22 bases with shorter primers resulting in non-specific products. All primers were checked for possible homology with CYP2D7 and CYP2D8 pseudogenes using the Vector NTI computer program and were shown to be specific to CYP2D6 alone.

Total reaction volumes were 25 ul, containing 50 mM MgCl2, 10 mM of each dNTP, 0.3 uM of each primer (LF and LR), ~ 50 ng of genomic DNA, 2.5 U of Taq Polymerase (Invitrogen), and 10X Reaction Buffer (Invitrogen). Reaction mixtures were overlaid with mineral oil. Thermal cycling was carried out on a Perkin Elmer Gene Amp 9600 Thermal Cycler with an initial denaturation step at 94°C for 3 minutes followed by 30 cycles of 94° C for 1 minute, 58° C for 1.5 minutes (annealing), 72° C for 2 minutes (DNA synthesis), and a final extension at 72° C for 7 minutes. Amplification of the long PCR product was verified by electrophoresis as previously described.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 4.2.5.3 CYP2D6 *4 PCR assay

The long PCR product was used in two separate assays containing either the wild-type or mutation-specific primer to detect the CYP2D6 *4 mutation (Table 4-2). The forward primer was 22 bases, while the reverse primers were designed to end either exactly at the mutation site on the gene or at a base pair corresponding to the mutation site in the normal sequence. The position of primers was selected such as to obtain a product of less than 500 bp for optimization of PCR. Reaction volumes, concentrations, and thermal cycling parameters were the same as for the long PCR described above. CYP2D6 *4 product was resolved by electrophoresis on a 2% agarose-TBE gel stained with ethidium bromide. Electrophoresis was done in a 1X TBE medium at 100 V. Gels were run until the dye marker neared the end of the gel. Positive and negative controls were used in each assay. A 100 bp ladder (Invitrogen) was used as a molecular weight marker for reference.

#### 4.2.5.4 Results

We analyzed a total of 33 samples, including 21 patient DNA samples extracted from buccal cells, and 12 DNA samples derived from both buccal and peripheral blood cells from 6 healthy volunteers. Both the wild-type and *4 mutation alleles were detected using the assay specific for CYP2D6 *4. Results for a representative selection of patient DNA samples are presented in Figures 4-1 and 4-2 and summarized in Table 4-3. PCR results from buccal samples for selected normal subjects are presented in Figure 4-3 and summarized in Table 4-3. Results were interpreted as follows: samples containing a wild-type band in the absence of a corresponding band for the mutation were considered homozygous for the wild-type mutations, representing the normal extensive metabolizer



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(EM) phenotype. On the other hand, samples containing both a wild-type and a mutated allele indicate a heterozygous genotype, also leading to the normal or EM phenotype since the mutation is a recessive trait. Samples lacking either the wild-type or the mutated allele were considered to represent the CYP2D6 *5 phenotypes resulting in the complete deletion of the gene.

Patient samples were either heterozygous or homozygous wild-type indicating all patients tested were CYP2D6 extensive metabolizers (EMs). Likewise, healthy volunteer samples were all shown to be heterozygous, thus, all volunteers were also extensive metabolizers. There were no samples with only a mutant band representing the PM phenotype, nor did any samples contain the CYP2D6 *5 gene deletion allele.

Although clear results were obtained for both the wild-type and the mutated *4 alleles, despite all attempts to optimize the assay for specificity of the CYP2D6 *4 allele, some faint non-specific bands persisted, none of which were within the size range specified by the particular primers utilized. Therefore, the presence of these spurious bands did not compromise the prediction of the *4 phenotype. Moreover, there was no difference observed in results obtained from blood samples versus buccal samples in healthy volunteers. Thus, interpretations were made using results from buccal cells for consistency across subjects.

#### 4.2.5.4.1 Validation and DNA Sequencing

To test the accuracy of results of the PCR assay, fragments of the CYP2D6 gene containing the *4 mutation were amplified using DNA samples from buccal samples obtained from 4 normal subjects (with same primers used for the samples) and submitted for automated DNA sequencing as follows.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A PCR assay was used to amplify a fragment of the CYP2D6 gene containing the *4 mutation region using the forward primer designed for the CYP2D6 *4 assay (Figure 4-2). A new reverse primer was designed to amplify a fragment of the CYP2D6 gene containing the CYP2D6 *4 allele (Figure 4-2). Both primers were 22 bases long with a GC content of ~60%. PCR conditions were same as described for the CYP2D6 *4 assay. After resolution of the long PCR product on a 1% agarose gel stained with 0.05% ethidium bromide, the DNA template was excised from the gel using a clean razor blade and purified using the QIAquick Gel Extraction Kit (QIAGEN. The DNA sample was then resuspended in Tris and submitted to Sequinet Laboratories, Fort Collins, Colorado for sequencing at a recommended concentration of 2 ng/uL per 100 bases of PCR product length. PCR product concentrations were estimated by running a small volume of the purified PCR product on an agarose gel adjacent to a DNA Mass Ladder (Gibco BRL). One forward (IF) and two reverse primers (B*4N and B*4M) were provided for utilization in the sequencing reaction at a concentration of 3μM (Figure 4-2).

**4.2.5.4.2 Sequencing Results**

As mentioned previously, results of the PCR assay indicated a heterozygous genotype containing a wild-type and a mutated allele for all volunteers, including the 4 whose samples were submitted for sequence analysis. DNA sequencing results were compared to the original CYP2D6 gene sequence using the NCBI Basic Local Alignment Search Tool (BLAST) program (http://www.ncbi.nlm.nih.gov). All 4 DNA sequences matched with the corresponding CYP2D6 normal sequence, confirming the presence of the wild-type allele in all 4 samples. However, none of the samples sequenced matched for the CYP2D6 *4 mutation sequence, most likely due to co-amplification of



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

pseudogenes regions homologous with the mutation-specific primer. These results suggest that the PCR assay developed was accurate in determining the specific CYP2D6 *4 allele of interest as the sequence was unique to CYP2D6 at sites specified by our primers.

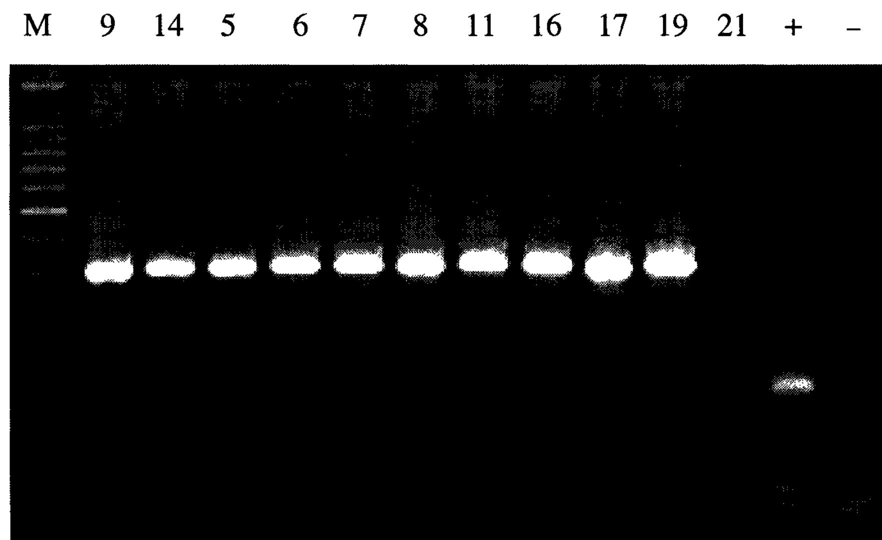



**Figure 4-1. Allele-specific PCR analysis of CYP2D6 mutation *4 (B) using the wild-type primer. Lane designations are as follows: "M" indicates a 100 bp molecular weight marker (Gibco BRL); numerals indicate sample ID as presented in Table 4-3; "+/-" indicate positive and negative control, respectively.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

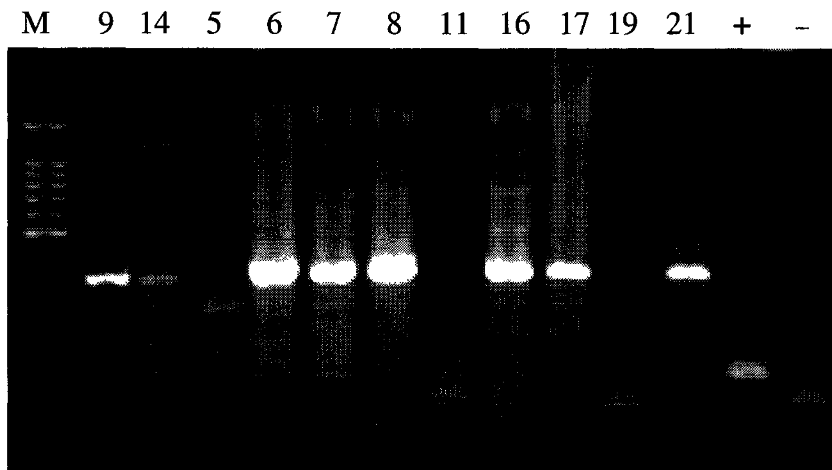



**Figure 4-2. Allele-specific PCR analysis of CYP2D6 mutation *4 (B) using the mutant primer. Lane designations are as follows: "M" indicates a 100 bp molecular weight marker (Gibco BRL); numerals indicate sample ID as presented in Table 4-3; "+/-" indicate positive and negative control, respectively.**

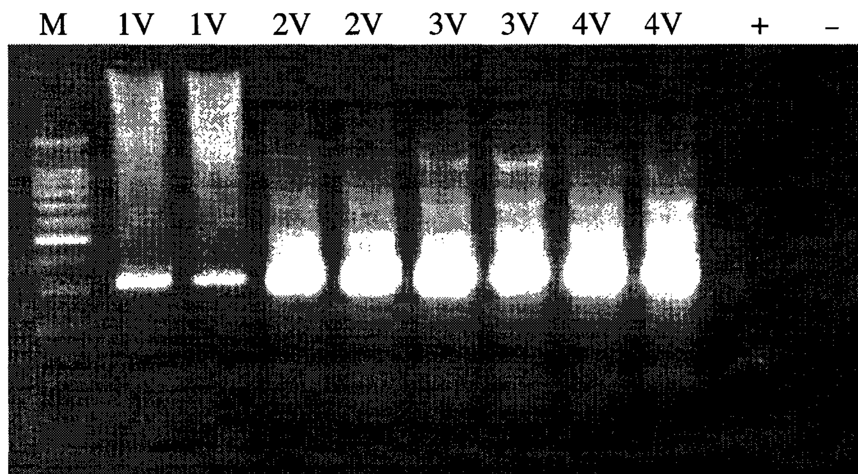



**Figure 4-3. Allele specific PCR analysis of CYP2D6 mutation*4 (B) using wild-type (left lanes) and mutant (right lanes) primers for each sample. A 100 bp molecular weight marker (Gibco BRL) is presented in Lane 1. Lanes 10 and 11 show a positive and a negative control, respectively.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

### 4.3 Discussion

Polymorphisms in the CYP2D6 gene resulting in the PM phenotype are suspected to increase susceptibility to Parkinson's disease in the presence of certain environmental chemicals.[175] Therefore, accurate detection of the CYP2D6 PM phenotype is an important component of epidemiologic studies of gene-environment interactions in the etiology of PD. Moreover, collection of genomic DNA using buccal swabs minimizes patient distress, expense, and exposure to blood-borne pathogens, and is thus an ideal DNA collection method in genetic epidemiologic studies. Thus, developing a rapid and reliable assay to detect the most common CYP2D6 PM phenotype using buccal swabs as a source of DNA would significantly facilitate epidemiologic investigations requiring CYP2D6 genotyping and provide a unique opportunity to investigate the contribution of this gene in PD. However, although several PCR assays to simultaneously detect 5 or more of the most common CYP2D6 PM alleles have been reportedly developed, all assays were designed to work with DNA obtained from peripheral blood of patients.

We examined the performance of one such method developed by Stuven et al. [142]using genomic DNA derived from buccal swabs but were unable to reproduce their results despite repeated attempts and optimizations. We then developed a PCR assay that allows the use of buccal cell-derived DNA to detect the two most common PM alleles accounting for over 95% of all PMs in Caucasians. Our goal in the design of the PCR assay was to create a simple, robust, rapid, and economical test, and one that would eliminate the need for other more extensive laboratory techniques such as RFLP analysis. Using DNA from buccal cells to preamplify a smaller segment of the CYP2D6 gene containing the single most common CYP2D6 PM mutation in Caucasians, we designed a



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

new PCR assay to detect CYP2D6 *4 and *5 alleles, the two most common CYP2D6 PM alleles in Caucasians. We used the preamplified gene product in two additional *4 allele-specific PCR assay to detect either the wild-type or mutation allele and succeeded in developing an assay for the CYP2D6 gene. Testing the accuracy of products generated from both PCR assays through sequencing, we were able to verify our results for the wild-type sequence. In contrast, due to cross-reactivity of the mutation-specific primer with pseudogenes, the mutation sequence obtained from the assay did not match the CYP2D6 *4 mutation sequence. However, since all our tested subjects possessed at least one wild-type allele, and because the wild-type allele is dominant, we could conclude that none of our subjects exhibited the PM phenotype. Thus, it was not necessary to redesign and verify an assay to accurately predict the *4 mutation allele. Moreover, complete elimination of spurious bands resulting from co-amplification with pseudogenes could not be achieved despite carefully designing all primers to be specific to our gene of interest. Surprisingly, the performance of the assay did not improve by using peripheral blood-derived DNA as the template, possibly suggesting persisting non-specificity of selected primers to pseudogenes which are highly homologous to this area of the CYP2D6 gene.

Two factors proved to be major limitations in the reliable assessment of the CYP2D6 genotype: quality of patient DNA and high level of homology of the CYP2D6 gene with the two neighboring pseudogenes, CYP2D7 and CYP2D8. Given this high degree of homology, use of the entire gene in allele-specific PCR assays results in a multitude of spurious products that render accurate assessment of CYP2D6 polymorphisms impossible. Therefore, any PCR-based assay to detect CYP2D6



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

polymorphisms must be followed by other laborious and costly techniques to try and distinguish the important products from all others. Thus, preamplification of a large segment of the gene containing the polymorphisms of interest using primers unique to CYP2D6 is a critical first step in any CYP2D6 genotyping assay. The preamplified gene product can then be used as a more specific template in subsequent allele-specific PCR assays, thereby eliminating undesired pseudogene products. However, the process required to extract DNA from buccal samples involves vigorous and repeated vortexing of the sample, which most likely results in excessive shearing of genomic DNA into fragments that are smaller than the large gene segments needed to be amplified, thus resulting in no PCR product after attempted amplification.

Based on our findings, we conclude that buccal cells are an inadequate source of DNA for multiplex PCR methods for simultaneous detection of CYP2D6 PM alleles as described in the literature. Moreover, for the first time, we demonstrated the use of buccal cells as a source of DNA for the accurate and rapid detection of the CYP2D6 *4 wild-type allele using a PCR-based assay. We were also able to detect the CYP2D6 *4 mutation allele using the same PCR-based assay. Thus, our assay can currently be used to rule out the presence of the PM phenotype in study subjects.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 4-2. Oligonucleotide primer sequences with position numbers according to Kimura et al. (1989). All primers are located in CYP2D6 with (-) and (+) indicating positions in the 5' and 3' flanking regions, respectively**

| Name | Used for | Sequence (5' to 3') | Position | Reference |
|---|---|---|---|---|
| P 1-5 | CYP2D6 (5') preamplification | GGTAAGGGCCTGGAGCAGGAA | -169/-146 | Stuven et al. (1989) |
| N | CYP2D6 (3') preamplification | GCCTCAACGTACCCCTGTCTC | +119/+99 | Stuven et al. (1989) |
| M | Multiplex forward | GTGGGGCTAATGCCTT | 1644/1659 | New design |
| A*3N | 2637 A | GGGTCCCAGGTCATCCT | 2653/2637 | Stuven et al. (1989) |
| A*3M | 2637 C (*3 mutation) | GGGTCCCAGGTCATCCG | 2653/2637 | Stuven et al. (1989) |
| B*4N | 1934 G | CGAAAGGGGCGTCC | 1947/1934 | Stuven et al. (1989) |
| B*4M | 1934 A (*4 mutation) | CGAAAGGGGCGTCT | 1947/1934 | Stuven et al. (1989) |
| E*7N | 3023 A | GCTGCACATCCGGAT | 3037/3023 | Stuven et al. (1989) |
| E*7M | 3023 C (*7 mutation) | GCTGCACATCCGGAG | 3037/3023 | Stuven et al. (1989) |
| T*6N | 1795 T | CTCCTCGGTCACCCA | 1809/1795 | Stuven et al. (1989) |
| T*6M | 1795 G (*6 mutation) | CTCCTCGGTCACCCC | 1809/1795 | Stuven et al. (1989) |
| G*8N | 1846 G | TCTGCCCATCACCCACC | 1862/1846 | Stuven et al. (1989) |
| G*8M | 1846 T (*8 mutation) | TCTGCCCATCACCCACA | 1862/1846 | Stuven et al. (1989) |

**Table 4-3. Newly designed oligonucleotide primer sequences with position numbers according to Kimura et al. (1989) within the CYP2D6 gene.**

| Name | Used for | Sequence (5' to 3') | Position |
|---|---|---|---|
| LF | 2D6 fragment (5')preamplification | TGATGACCGTAGTCCGAGCTGG | 1213/1192 |
| LR | 2D6 fragment (3')preamplification | GGGTCCCAGGTCATCCTGTGCT | 2653/2632 |
| IF | Internal 5' forward | TGGTGGATGGTGGGGCTAATGC | 1656/1635 |
| B*4N | 1934 G | CGTTGGGGCGAAAGGGGCGTCC | 1955/1934 |
| B*4M | 1934 A (*4 mutation) | CGTTGGGGCGAAAGGGGCGTCT | 1955/1934 |
| IR | Reverse primer for DNA sequencing | ACGACCATGTCTGAGATGTCC | 2269/2249 |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 4-3. Genotype analysis and phenotype prediction for DNA samples presented in Figures 4-1 to 4-3.**

| Sample ID | Lane number | PCR result interpretation | Predicted genotype | Predicted phenotype |
|---|---|---|---|---|
| 5 | 3 (Figures 3-1 and 3-2) | wt/wt | Homozygous normal | EM |
| 6 | 4 (Figures 3-1 and 3-2) | wt/*4 | Heterozygous | EM |
| 7 | 5 (Figures 3-1 and 3-2) | wt/*4 | Heterozygous | EM |
| 8 | 6 (Figures 3-1 and 3-2) | wt/*4 | Heterozygous | EM |
| 9 | 1 (Figures 3-1 and 3-2) | wt/*4 | Heterozygous | EM |
| 11 | 7 (Figures 3-1 and 3-2) | wt/*4 | Homozygous normal | EM |
| 14 | 2 (Figures 3-1 and 3-2) | wt/*4 | Heterozygous | EM |
| 16 | 8 (Figures 3-1 and 3-2) | wt/*4 | Heterozygous | EM |
| 17 | 9 (Figures 3-1 and 3-2) | wt/*4 | Heterozygous | EM |
| 19 | 10 (Figures 3-1 and -2) | wt/wt | Homozygous normal | EM |
| 21 | 11 (Figures 3-1 and -2) | wt/*4 | Heterozygous | EM |
| 1V | 1 (Figure 4-3) | wt/*4 | Heterozygous | EM |
| 2V | 2 (Figure 4-3) | wt/*4 | Heterozygous | EM |
| 3V | 3 (Figure 4-3) | wt/*4 | Heterozygous | EM |
| 4V | 4 (Figure 4-3) | wt/*4 | Heterozygous | EM |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER 5

# SIMULATION STUDY TO ASSESS DEVIATION OF RISK ESTIMATES FROM TRUE POPULATION RISK IN CASE-CONTROL AND CASE-ONLY STUDIES OF GENE-ENVIRONMENT INTERACTION

## 5.1 Introduction

The traditional case-control epidemiologic study design is well suited for the evaluation of gene-environment interactions but often requires much larger sample sizes than are necessary for the evaluation of main effects.[9,59,65,171] Moreover, ascertainment of an appropriate control group, an inherent challenge leading to potential confounding and other biases in all case-control studies, is made even more difficult when genetic traits are a concern in such a study.[9,57] For instance, confounding can result if a genotype occurs with variable frequency within subgroups of a given population due to population admixture, leading to "population stratification" among cases and controls, and inaccurate study results.[57,194]

In order to address the important limitations of the case-control study arising from unavoidable errors in control selection and often prohibitively large sample size requirements, the case-only study design has recently been proposed as a more efficient method to evaluate gene-environment interaction effects in disease causation.[194] The case-only model follows the same epidemiologic principles of case selection as the case-control study with the important exception that the subjects consist of cases only.[9,58,194] The design is valid only if the susceptibility genotype and the environmental exposure are independent in the population.[58] If this assumption is met, theoretical models have shown that case-only estimates of gene-environment interaction yield higher precision



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

than the same estimates derived from a case-control study.[58] The increase in precision results from a decrease in standard errors due to the elimination of control group variability, without compromising power to detect interaction effects.[58] Therefore, the case-only method has been promoted as a more appropriate design when the primary research objective is assessment of interaction effects.[194]

The purpose of this study was to use statistical simulations to compare the relative accuracy of the case-control and case-only study designs to detect gene-environment interactions as measured by increases in precision (leading to a reduction in random error) and validity (resulting in a reduction in systematic error). Precision was determined through evaluation of the magnitude and frequency with which estimated main and interaction effects obtained from random samples from a given population deviated relative to the "true" risk in that population. Since the scope of this dissertation was limited to the special case of Type 2 gene-environment interaction (GEI) in which risk of disease among non-exposed susceptibles is assumed to be unity,[6,7] our simulations were likewise designed to evaluate the Type 2 model of GEI exclusively. Additionally, our simulations were based on the assumption that genotype and environment are independent among the population, a requirement for the validity of case-only estimates.[58] We determined the proportion (frequency) of various estimated odds ratios from both case-control and case-only studies included in increasingly larger magnitude of ranges (±10% to ±50%) around the "true" odds ratio for each of several simulated populations. We then evaluated the possible effects of violation of the two assumptions on our simulation results. Accuracy of risk estimates from both the case-control and case-only methods was determined for different samples with sample sizes ranging from



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

100 to 1000 (using increments of 100) for thirty-six populations representing different combinations of prevalences of disease, exposure, and genotype, and magnitude of effects.

In this chapter we demonstrate that if the independence and Type 2 assumptions are satisfied, risk estimates obtained from the case-only design approximate "true" (population) risks with higher accuracy than estimates from case-control studies, even if both assumptions are violated in the sample, a common occurrence. Interestingly, we show that case-only estimates continue to out-perform case-control estimates when sample sizes are below 400, even if the independence assumption is violated to a moderate degree (OR for association between genotype and exposure = 1.5) in the population. In fact, we demonstrate that accuracy of results from the case-only study strongly depend on other design factors such as prevalences of disease, exposure, and genotype, satisfaction of the Type 2 assumption, magnitude of interaction, and various interactions of design factors rather than on presence of the independence assumption alone. Although our results indicate that violation of the independence assumption by an odds ratio of 1.5 has a slight effect on estimates obtained from both study designs, the case-only results become less accurate while case-control results shows a marginal improvement as compared to results obtained from a population in which the independence assumption is met. Results from both case-only and case-control study designs continue to either decrease or increase in accuracy commensurate with an increase in the magnitude by which the independence assumption is violated in the population. Additionally, we show that, given low genotype and/or exposure prevalences in the population, both the case-control study and the case-only study with small sample



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sizes (100-200 subjects) are at substantially increased risk of selecting diseased subjects who lack either the genotype or exposure, rendering risk estimates from either study unattainable and/or invalid. Moreover, our results demonstrate that a large proportion of both case-only and case-control estimates fail to approximate the "true" risk within a ±30% range, even when the number of subjects is increased to 1000.

Based on our findings, we suggest the use of case-only designs when the study objective is to determine Type 2 GEI effects. However, we caution against use of the case-only design when the independence assumption is likely to be violated in the population. When both main and GEI effects are of equal interest, requiring a case-control study design, we recommend increasing the sample size of the study beyond that required to detect the main effects alone in order to allow adequate power to detect GEI effects, especially when genotype and/or exposure is rare.

**5.2 Methods**

Since the risk of disease in any given population is impossible to ascertain, we constructed hypothetical populations to represent frequencies of the disease and risk factors of interest as reported in the literature. We randomly sampled cases and controls from such populations constructed to represent various realistic and extreme (best-case or worst-case) scenarios while incrementally increasing sample size. Appropriate risk estimates were calculated for each unique scenario. We chose Parkinson's disease (PD), a neurodegenerative disorder with unknown etiology but for which the interaction of certain genes and environmental exposures are highly suspected as our disease of interest. Pesticide exposure was selected as the environmental risk factor due to its strong association with PD. CYP2D6, a gene encoding for a biotransformation enzyme known



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to detoxify many types of pesticides and other chemicals associated with PD was chosen as the susceptibility gene. The interaction of pesticides with the CYP2D6 poor metabolizer polymorphism represents a Type 2 biological model of interaction whereby presence of the genotype alone is neither necessary nor sufficient to cause disease.[6,7] Hence, risk of disease for those with the genotype in the absence of exposure was assumed to be unity.

The simulation consisted of two stages:

1) Creating hypothetical populations representing a wide range of prevalences of PD, CYP2D6 PM genotype, and pesticide exposure, and for whom "true" risk of disease was known. The following risks were pre-specified in the construction of each population:
    a. Relative risk of disease for exposed susceptibles (i.e., those with CYP2D6 PM genotype who report pesticide exposure) compared to unexposed, genetically non-susceptible individuals
    b. Relative risk of disease for unexposed susceptibles
    c. The factor by which disease risk for exposed susceptibles differs from the multiplied effects of exposure and genetic susceptibility on disease risk, described as a synergy index on a multiplicative scale (SIM)
    d. Interaction effects between pesticide exposure and susceptibility genotype among non-diseased (controls) individuals
2) Sampling from each population (while increasing the number of subjects) to compute relevant risk estimates in order to compare to the "true" risk.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

### 5.2.1 Construction of populations

A "typical" population of size 100,000 was created to represent reported prevalences of PD in individuals aged 65 or older, CYP2D6 PM genotype among Caucasians, and pesticide exposure, as well as the risk of pesticide exposure on development of PD. Based on published reports, prevalences of disease, genotype, and exposure were set to be 5, 7, and 10%, respectively to represent a realistic situation.{55/id;215/id;218/id} Prevalence of disease was slightly overestimated to account for under- and mis-diagnosis of PD. The marginal odds ratio for disease among the exposed versus the unexposed was estimated at 2 based on numerous reports and a meta-analysis.{55/id} Two risk estimates, odds ratio of disease from genotype in the absence of exposure and the odds ratio of association between genotype and exposure among controls were prespecified to be unity in order to accommodate the Type 2 model of gene-environment interaction and the independence assumption for case-only studies, respectively. We found no published reports for magnitude of interaction for exposed susceptibles. However, most Type 2 gene-environment interaction effects are considered to have a much larger effect that the product of the main effects, {87/id} and the dose-response relationship for exposure to pesticides that in turn bioaccumulate due to insufficient metabolic capacity can be considered to be non-linear; therefore, we assumed the interaction effects to be greater than additive.

Additional populations of 100,000 were created representing various combinations of half and twice the value of each input parameter (with the typical parameters as the center point) in order to examine other scenarios ranging from "worst-case" with the lowest prevalences and odds ratios (half the typical values) to "best-case"



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with highest prevalences and odds ratios (twice the typical values). Three additional populations were constructed to evaluate the accuracy of estimates obtained from each study design relative to small interaction effects, and when the independence assumption is violated to various degrees in the population. Table 5-1 presents a summary of the parameters used in each population. Population frequencies were computed using the Maple V software package, a general-purpose computer algebra system from Waterloo Maple Software (WMS).

### 5.2.2 Sampling from each population

Populations were first stratified by disease status, from which random samples of increasingly larger sizes were taken using a 1:1 case to control ratio. Total numbers of subjects ranged from 100 (50 cases and 50 controls) to 1000. Sample replicates were generated using the SAS software procedure SURVEYSELECT (SAS version 8.01, SAS Institute Inc., Cary, NC, USA).

Use of 1000 replicates in the simulation study was evaluated by computing the relative standard errors of the percent of estimates of interaction effects within 10% (conservative), 30% (acceptable) and 50% (unreliable) of the "true" risk based on the Typical population representing expected risks in a real scenario in which the "true" magnitude of interaction is expected to be 3.0. Relative standard errors (RSE) were calculated as the ratio of the standard errors of the estimate accurate within given ranges of "true" interaction effect of 3 (for Typical population) and percent of SIM or COR estimates. Table 5-2 is a summary of the analysis. One thousand replicates produced relative errors of less than five percent for case-only estimates in the acceptable range (±30% of "true" risk) even in samples with the smallest number of subjects (n=100). In



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

order to maintain the same precision for the corresponding case-control estimates using the same number of subjects the number of replicates would have had to be increased four-fold for the smallest sample size (n=100). However, relative error for the case-control estimates remained below ten percent with 1000 replicates and was thus considered acceptable. For sample sizes of 500 or more, much fewer than 1000 replicates would have been required for precision of estimates at 5% relative error. Thus, for the sake of consistency, we chose to sample 1000 replicates for each of the 10 sample sizes from each of the 36 simulated populations

**5.2.3 Determination of relative contribution of population parameters to accuracy of GEI estimates**

Multi-way analysis of variance was performed to determine the relative contribution of all population parameters and assumptions to the accuracy of estimates using the SAS software procedure GLM (SAS version 8.01, SAS Institute Inc., Cary, NC, USA). The percentage of odds ratio estimates for interaction effects that were accurate within ±30% of the "true" risk was the outcome of interest. Population characteristics examined included all input parameters with more than one level used in different combinations to construct all populations: prevalences of disease ($P_D$), exposure ($P_E$), and genotype ($P_G$), odds ratio for exposure alone ($OR_{Eg}$), odds ratio for exposure and genotype ($OR_{EG}$), magnitude of interaction effect beyond the multiplicative risk (SIM or COR), the two study designs of interest (case-control and case-only), and number of subjects (100-1000). Two additional variables, the independence and Type 2 assumptions, were also examined. Recall that the Type 2 assumption is satisfied when the odds ratio of disease among the non-exposed susceptibles ($OR_{Ge}$) is equal to one, and



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the independence assumption requires the odds ratio of association between genotype and exposure among controls to be equal to one. The assumptions were determined to be violated if estimates were less than 0.5 or greater than 1.5 (outside the ±50% range of 1).

### 5.2.4 Analysis

For simplicity, both genotype and environmental exposure were assumed to be dichotomous. Disease risks were estimated using logistic regression models. All models were constructed using SAS software procedure GENMOD (SAS version 8.01, SAS Institute Inc., Cary, NC, USA) to obtain estimates and are described in the following sections.

Six odds ratios (five of which can be obtained from a case-control study and one that can be calculated using a case-only design) with 95% confidence intervals were calculated for each entire population and are explained in detail in the next three sections. Odds ratios for the whole population were considered to represent the "true" risks for that population. Respective odds ratios and 95% confidence intervals were computed for each sample drawn from the population and were considered risk estimates. Mean odds ratios were computed from all 1000 replicates for each sample size and for each population by averaging the odds ratios and the average of the coefficients was calculated and exponentiated.

#### 5.2.4.1 Case-control estimates

We calculated the following three odds ratios and the SIM for each replicate using a case-control design:

$\hat{OR}_{GE}$ = odds ratio of PD for those with both the CYP2D6 PM susceptibility genotype and pesticide exposure relative to those without either risk factor,



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$\hat{O}R_{Ge}$ = among those not exposed to pesticides, odds ratio of PD for those with the CYP2D6 PM susceptibility genotype relative to those without the genotype (this ratio is assumed to be one in a Type 2 model of gene-environment interaction),

$\hat{O}R_{Eg}$ = among those with the CYP2D6 PM genotype, odds ratio of PD for those exposed to pesticides relative to those not exposed, and

$S\hat{I}M$ (synergy index on multiplicative scale) = deviation from multiplicative model of interaction,

where

E = presence of environmental exposure,

e = absence of environmental exposure,

G = presence of susceptibility genotype, and

g = absence of susceptibility genotype.

#### 5.2.4.3 Case-only estimate

The case-only odds ratio (COR) relates exposure and genotype among cases by comparing the odds of exposure among those with the susceptibility gene to the odds of exposure among those without the susceptibility gene. The COR is equivalent to the SIM and is used to assess deviations from a multiplicative model of interaction in studies of gene-environment interaction in cases where the gene and the exposure are independent among the general population. If the independence assumption is met, then a COR (or SIM) value of greater than one would indicate more than multiplicative effects between exposure and genotype, whereas a COR (or SIM) of less than one would refer to a less than multiplicative effect.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 5.2.4.3 Control-only estimate

As mentioned in the previous section, results from a case-only study are valid only if the odds ratio among the controls (i.e., general population) is unity. Verification of this assumption, of course, is not possible when conducting a case-only study. However, in this simulation, we were able to easily obtain the odds ratio among the controls using the same logistic regression model as for the case-only design except where D = 0 (i.e., no disease).

## 5.3 Results

Accuracy of estimates improved commensurate with increasing sample sizes. As expected, for both study designs, accuracy also improved with increasing population parameters (odds ratios and frequencies of disease, genotype, and exposure) since the sample size affects power which is a function of said parameters. A consistent pattern was observed across all populations. Population 1 represented the worst-case scenario (i.e., lowest prevalences and odds ratios), Typical Population represented expected prevalences and odds ratios according to published reports, and Population 32 represented the best-case scenario (i.e., highest prevalences and odds ratios). Results from the remaining populations representing combinations of parameters between worst- and best-case scenarios (Populations 2-32) will not be discussed in detail as they did not provide any additional information. Of importance in the interpretation of results is that although all simulations were based on 1000 replicates, valid estimates were unobtainable for some replicates, thereby reducing the total number of samples used for analysis. Table 5-3 presents frequencies and percentages for all replicates in which either exposure or genotype was absent in all diseased subjects. As can be seen in the table, only 54% of



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

replicates with 100 subjects and 73% of those with 200 subjects from Population 1 through 8 (with small values for prevalences of gene and exposure) were useable for calculation of the case-only estimates. The same pattern was found for replicates of size 100 and 200 in Population 9 which combines lowest prevalence of gene with lowest interaction effects. Frequencies of useable replicates of size 100 and 200 increased with increasing prevalence of genotype until reaching 100% at Population 16 with higher prevalences of genotype, exposure, and odds ratios for interaction (Table 5-1).

Results of analyses for the Typical Population and Populations 1 and 32 are presented in Tables 5-4 to 5-19. Each table summarizes the accuracy of estimates obtained based on valid replicates for each of ten different sample sizes from a given population with respect to five increasingly wider ranges (±10% to ±50%) around the "true" risk in that population. Results are provided as percentages of valid estimates that fall within each given range. Invalid replicates included samples in which the prevalences of exposure and genotype were low and which failed to include diseased subjects with either the genotype or exposure. Since genotype among cases is used as the outcome in the logistic regression model for the case-only study, the model cannot be applied to replicates with missing values for the outcome. All corresponding case-control models also failed to converge in the absence of diseased subjects with the risk factors of interest. Hence neither case-only nor case-control estimates could be obtained from such replicates.

Finally, we found very little difference between the mean of odds ratios and the exponentiated mean of coefficients of the logistic regression model (see Tables 5-4 and 5-5). Therefore, the means of odds ratios will be discussed exclusively.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

### 5.3.1 Evaluation and comparison of accuracy of case-only and case-control interaction estimates

Recall that percentages presented in Tables 5-4 through 5-19 are based on the number of replicates for which estimates could be computed; and, as such, are conditional on the percentage of estimates reported in the first column of each table. As can be seen, in the Typical population, over 50% of samples from a case-control study with 100 subjects were inadequate for analysis (based on above criteria) as compared to only 2.5% of samples from a case-only study using an equal number of subjects. Since the number of cases in a case-control study of 100 subjects with a 1:1 ratio of cases to controls provides only half the number of cases as in a case-only study of the same size, we also compared the two designs with equal numbers of cases in each. The accuracy of case-control estimates continued to be lower than case-only estimates even when the same number of cases was compared for each design. For example, analysis based on the logistic regression model was possible for only 72.5% of replicates from a case-control study with 100 cases and 100 controls (compared to 97.5% from a case-only study of 100 subjects). Further, 1.1% of case-control models failed to converge even when the number of subjects was increased to 900 (Table 5-4). The number of replicates suitable for analysis from a case-control design increased dramatically with increasing number of subjects. After removing all replicates for which odds ratios could not be estimated, the case-only study still out-performed the case-control design irrespective of the number of subjects in the sample.

Evaluating the remaining samples for which GEI estimates could be computed, less than 10% of those generated from a case-control study of size 200 were accurate



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

within ±10% of the "true" risk. Increasing the sample size did improve accuracy but only slightly, with less than 20% accuracy within a ±10% range of "true" risk for case-control estimates from the largest samples with 1000 subjects (Table 5-4). Comparatively, while the case-only study also performed rather poorly with the smallest number of subjects in the study (11.9% accuracy at 10%), increasing the number of subjects increased the accuracy of case-only estimates more than twice that of the case-control estimates such that 24.9% of case-only versus 12.3% of case-control estimates were accurate within 10% of the "true" risk for samples with 500 subjects (Table 5-4). Effect of sample size on accuracy of estimates from the Typical population is illustrated in Figure 5-1. Case-only estimates continued to be more accurate even when compared to case-control estimates from samples with equal number of cases (24.9%, n = 500 versus 17.2%, n = 1000). As expected, the percent within ranges improved as we increased the range of intervals around the "true" values, such that for a sample with 500 subjects, almost 90% of case-only estimates fell within the widest range (50%) of the "true" risk while only 60.5% of case-control estimates fell within the 50% range. Again, case-only estimates were still more accurate than case-control estimates given equal number of 500 cases (88.6%, n = 500 versus 74.8%, n = 1000). Results of averages from exponentiated coefficients corresponding to odds ratios discussed above (Table 5-4) are presented in Table 5-5. As can be seen, results obtained by the second method (averages from exponentiated coefficients) were very similar to results of odds ratios obtained from transformed coefficients and will not be discussed.

Results of worst- and best-case scenarios (Populations 1 and 32, respectively) followed a similar trend as in the Typical population (Tables 5-6 and 5-7). Specifically,



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

case-only estimates were more accurate than case-control estimates in all cases with accuracy improving with increasing sample size and population parameters. As expected, percentage of estimates within all ranges were higher in the best-case scenario than in the Typical population, while fewer estimates fell within the five ranges in the worst-case scenario as compared to the Typical population. It is important to interpret the accuracy within each range as a percentage of the total number of valid replicates (not necessarily 1000 replicates) depending on which the percentages may appear too optimistic.

#### 5.3.1.1 Evaluation and comparison of case-only and case-control study designs given weak interaction effect

Given that the magnitude of interaction effects may be less pronounced than what we assumed for our typical scenario (SIM and COR=3), we created an additional population (Population 33) with characteristics identical to the Typical Population, but in which the magnitude of interaction (SIM and COR estimates) was closer to unity (SIM and COR=1.2) in order to examine the respective performances of the two study designs given small magnitude of interaction effects. Results of the analysis are presented in Tables 5-20 through 5-26. Predictably, accuracy to detect departures from multiplicative interaction effects was reduced in both study designs as compared to the Typical population with larger interaction effects (Tables 5-20 and 5-21). Accuracy to detect all other estimates (i.e., odds ratio of exposure alone, odds ratio of genotype alone, and odds ratio between exposure and genotype among controls) was also markedly reduced (Tables 5-22 to 5-26). For example, approximately 10% fewer estimates of interaction effects were accurate within ±40% range of the "true" risk in Population 33 as compared



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to the Typical population for both case-only and case-control designs (45.3% versus 55.8% CORs and 26.6% versus 34.4% for SIMs). Similarly, accuracy to detect effects of exposure alone within ±40% of "true" risk was reduced from 93.4% in the Typical population to 56.2% in Population 33 (Tables 5-14 and 5-26), while accuracy to determine odds ratio of association between genotype and exposure (independence assumption) fell from 69.9% in the Typical population to 43.8% in Population 33 (Tables 5-10 and 5-24).

At odds ratios approaching unity, we observed a difference between the results obtained from the average of exponentiated coefficients from the logistic regression model and the odds ratio of the transformed averages. This discrepancy is likely due to narrow ranges (around the coefficient) obtained from exponentiation of very small coefficient values. In this case, ranges obtained from exponentiating values of ranges around the coefficient (0.18) are much more narrow than ranges computed from odds ratios of transformed coefficients as illustrated in Tables 5-20 and 5-21. The observed variation in the two methods diminished as the interaction effect became larger, as explained above.

#### 5.3.1.2 Relative contribution of population parameters to accuracy of GEI estimates

Relative contribution of population parameters to accuracy of GEI estimates was evaluated using multi-way analysis of variance. Results of main effects and first-order interactions for all population parameters are presented in Table 5-27. As can be seen, all main effect variables with the exception of the independence assumption made a statistically significant contribution to the variability in the accuracy of estimates at the 1% level of significance. As indicated by the Type III sum of squares, magnitude of



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interaction (SIM or COR) had the greatest effect (by two orders of magnitude) followed by the choice of study design (by one order of magnitude) in comparison to other variables. Violation of the Type 2 assumption and all population prevalences also made a highly statistically significant contribution to accuracy of estimates. However, all main effect variables including the independence assumption (which was not significant individually) were also found to be significant in two-way (first-order) interactions, thus, meaningful interpretation of main effect results was not possible. For example, although the independence assumption alone did not show a significant contribution to accuracy of estimates, the contribution of the variable was statistically significant in interactions with SIM or COR, study design, and $OR_{GE}$, suggesting that the effect of violation of the independence assumption on the variability of the estimate depends, in order of contribution, on these variables. First-order interactions of study design with both Type 2 and independence assumptions made the most significant contribution to the model (highest Type III sum of squares), followed by the interaction of the independence assumption with magnitude of the interaction (SIM or COR) and number of subjects in the sample. Prevalences of gene and exposure also contributed interactively to the variability of estimates with each other, as well as with prevalence of disease and $OR_{GE}$, but to a lesser degree. Thus, the results of this analysis suggest that magnitude of interaction, study design, and the two assumptions most statistically significantly contribute to accuracy of estimates in a highly inter-dependent manner.

Given the significant interaction effects of study design on accuracy of estimates, we stratified the data by study design in order to evaluate the effects of population parameters on estimated obtained from each design separately. Tables 5-28 and 5-29



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

summarize the effects of all variables and first order interactions on accuracy of case-only and case-control estimates, respectively. As in the overall analysis, all main effects significantly interacted with each other, leading to difficulties in interpretation of main effects.

Evaluating first-order interactions, sample size, prevalence of gene, magnitude of interaction (COR), and the independence assumption made the most significant contributions to accuracy of case-only estimates, although all other variables also contributed (to a lesser extent) in various combinations of interactions. The independence assumption was found to depend mostly on the magnitude of interaction (COR) followed by the prevalence of gene, $OR_{GE}$, prevalence of exposure, and prevalence of disease. Type 2 assumption depended mostly on size of the sample, followed by prevalence of exposure, $OR_{GE}$, and magnitude of interaction (Table 5-28). In contrast, case-control estimates depended most statistically significantly on the interactions of the independence assumption with magnitude of effects (SIM) and with $OR_{GE}$, followed by prevalence of gene interacting with sample size, and prevalence of exposure. The Type 2 assumption depended significantly on magnitude of interaction (SIM) only (Table 5-29).

Because sample size most strongly contributed to accuracy of both estimates by interacting with other variables, we further reduced the data by sample size to examine the contribution of parameters to the variability in estimates for each study type at three different sample sizes. Results are presented in Tables 5-30 and 5-31. For case-only studies with 100 subjects, all main effect variables interacted with each other with the exception of prevalence of disease, which was found to be slightly significant only as a



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

main effect. The strongest interactions were seen among prevalence of gene, prevalence of exposure, $OR_{GE}$, SIM or COR, and the Type 2 assumption. Additionally, the independence assumption interacted significantly with $OR_{GE}$ to explain the variability in case-only estimates, especially when study size was small. Accuracy of case-control estimates given the same number of subjects (n=100) depended mostly on the interactions between the independence assumption and SIM or COR, as well as with prevalences of gene and exposure while the Type 2 assumption was significant when interacting with the prevalence of gene.

Contribution of the independence variable increased (interacted significantly with more variables) when the number of subjects in the case-only study increased to 500. However, the most significant interaction variables were prevalence of exposure and $OR_{GE}$. The Type 2 assumption was only significant in relation to prevalence of exposure, while the independence assumption depended on prevalences of both exposure and gene as well as $OR_{GE}$. All main effects were significant in interactions except prevalence of disease, which contributed neither individually nor jointly. By contrast, interaction of the independence assumption with SIM made the most important difference in accuracy of the case-control estimate given the same number of subjects. Independence assumption was also highly significant with $OR_{GE}$ while the Type 2 assumption and prevalence of disease had no effect on SIM.

Finally, the COR for samples of size 1000 was mostly affected by interactions between prevalence of gene, exposure and disease with magnitude of interaction (COR) and $OR_{GE}$. Neither assumption seemed to have any effect on COR at this sample size. By comparison, the case-control estimate (SIM) continued to be highly affected by



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interactions of the independence assumption and SIM, $OR_{GE}$, and prevalence of gene, and the Type 2 assumption, while prevalence of disease did not have any effect on SIM at n=1000. These results suggest that the independence assumption is the most important contributor to the accuracy of the SIM irrespective of sample size, while the COR is affected by both assumptions at smaller sample sizes only, and depends more on prevalences of gene and exposure and the presence and magnitude of interaction (Tables 5-30 and 5-31).

Together, these results indicate that the independence and Type 2 assumptions strongly impact accuracy of both case-only and case-control estimates. Moreover, the results suggest that the degree to which each assumption affects variability in accuracy of estimates depends on other study parameters, particularly the sample size. Thus, it was important to examine the individual and joint contributions of both assumptions on accuracy of case-only and case-control estimates given different sample sizes from the same population. We selected three populations representing typical, worst- and best-case scenarios to perform the analysis which is discussed in detail in the following section.

#### 5.3.1.3 Effect of violation of the independence and Type 2 assumptions in the sample on accuracy of case-only and case-control estimates

Satisfaction of the assumption of independence between genotype and exposure among the population, while not necessary for the case-control study, has theoretically been shown crucial to the validity of estimates from case-only studies. Recall that this assumption is only satisfied when the odds ratio for association between exposure and genotype among controls is equal to one. Unfortunately, despite its importance, this



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

assumption cannot be verified in a case-only study, which, by design, eliminates the need for control selection. However, our simulated samples consisted of both cases and controls in equal numbers since both case-only and case-control estimates in our simulations were obtained from the same samples. Hence, we had the opportunity to evaluate the accuracy of the independence assumption (as measured by the odds ratio for association between gene and exposure among controls) in each sample, alone (Table 5-8), as well as relative to both study designs. Not surprisingly, for studies with 200 subjects, we found less than 10% of the odds ratios calculated to fall within ±10% of the "true" risk of one. It is worthy of note that valid estimates were obtained for100% of all replicates for all sample sizes, suggesting results of main effects are estimated with higher precision than those of interaction effects in the case-control model. Nevertheless, accuracy of estimates within the narrowest range improved only slightly to less than 20% for samples with 1000 subjects, suggesting that precision remained inadequate at conservative levels of accuracy. Accuracy did improve with increasing sample size, however, most notably in the widest range around the "true" risk (±50%). For example, over 50 and 75 percent of estimates were accurate within ±50% of "true" risk for samples with 500 and 1000 subjects, respectively. However, our results suggest that the independence assumption is often not satisfied in samples of reasonable size, even when there is independence between genotype and exposure in the general population from which the samples are drawn. The frequent violation of the independence assumption in our samples seemed to warrant a more thorough examination of the impact of the assumption on accuracy of estimates. However, as seen in the analysis of variance



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

results, it was important to examine the independence assumption in relation to the second a priori assumption which we next explored in detail.

The Type 2 biological model of gene-environment interaction in which genotype alone is neither necessary nor sufficient to cause disease is considered to be satisfied if, among the unexposed, the odds ratio of disease for those with the susceptibility genotype relative to those without the genotype ($OR_{Ge}$) is unity. Interestingly, as presented in Table 5-11, although the trend in accuracy of the Type 2 assumption was very similar to that of the independence assumption (i.e., increased in proportion to sample size and wider ranges around one), proportions of Type 2 estimates within each range were higher under all conditions as compared to the independence assumption estimates. This apparent discrepancy can likely be explained in relation to the sample size for the respective estimates. Specifically, recall that violation of the independence assumption is evaluated as the odds ratio for association between gene and exposure among controls only, whereas the Type 2 assumption is determined by estimating the odds ratio of disease among all (cases and controls) susceptibles versus non-susceptibles among the unexposed. As such, the latter assumption is evaluated from a sample including twice as many subjects as the former assumption given the 1:1 ratio of cases to controls, leading to higher precision as seen in our results. However, although the discrepancy was considerably less when results of each assumption were compared between samples with equal numbers of subjects, the proportions of estimates that met the independence assumption remained smaller.

We evaluated the individual as well as joint effects of both the independence and Type 2 assumptions on the accuracy of case-only and case-control estimates of



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

interaction effects (COR and SIM, respectively) for three populations: Typical, Population 1 and Population 32, representing expected, worst-case, and best-case scenarios with respect to all parameters (prevalences of gene, exposure, and disease; relative risk of exposure alone; and presence and magnitude of interaction). Effects of violation of the assumptions on accuracy of estimates were evaluated in two ways. First, relative proportions of estimates that met either or both the independence or Type 2 assumptions as compared to estimates that failed to meet either or both assumptions were evaluated in relation to accuracy within ±30% range around "true" risk. (Tables 5-32 through 5-37, "Column Percents"). Additionally, accuracy of estimates within ±30% range around "true" risk was examined relative to proportion of samples that met either or both assumptions as compared to samples that violated either or both assumptions. (Tables 5-32 through 5-37, "Row Percents").

Violation of both the independence and Type 2 assumptions affected accuracy of case-only and case-control estimates obtained from each of the three populations examined. However, the overall effect of violation of assumptions on study results depended on sample size. This was expected due to the large effect of sample size on accuracy. Very low precision resulted in failure of almost all estimates to meet accuracy criteria. For example, For example, approximately 75% of all case-only estimates obtained from samples of size 100 from the Typical population failed to meet accuracy criteria ±30% of the "true" risk (Table 5-32). Thus, any effects of violation of assumptions on accuracy were obscured by lack of precision. In contrast, precision was extremely high for very large sample sizes such that most estimates met the accuracy criteria irrespective of presence of systematic error. For example, 75% of case-only



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

estimates for sample sizes exceeding 400 fell within the range (Table 5-32). However, presence of bias most adversely affected case-only estimates when study precision was moderate (i.e., at neither extreme), such that estimates fell within or outside accuracy criteria of ±30% in approximately equal proportions. In such cases (n=200-400), a higher proportion of estimates that met both assumptions fell within the accuracy range as compared to estimates that met neither assumption. This discrepancy was almost entirely due to the effect of the Type 2 assumption on accuracy with the independence assumption having only a negligible effect. For instance, 61.4% of estimates that met the Type 2 assumption were within the ±30% range versus 45.7% of those that did not meet the assumption but that met accuracy criteria whereas the difference between estimates that met or did not meet the independence assumption with respect to accuracy was smaller (e.g., 48.2% versus 45.7% for sample size 400). The Type 2 assumption was met with much higher frequency than the independence assumption in estimates from all sample sizes.

Although a similar trend was observed for case-control estimates from the Typical population, this study design estimated interaction effects with far less precision such that estimates within and outside the ±30% range did not approach equal proportions until sample sizes were increased to 700 or greater. Thus, precision had a larger effect (relative to bias) for sample sizes between 100 and 600 such that low accuracy rates were primarily due to lack of precision versus presence of bias. The effect of bias increased with increasing sample size. Effects of bias dominated precision most notably at sample size 700 but declined thereafter. Nevertheless, violation of both assumptions had a negative impact on accuracy such that only 28.6% of estimates that met neither



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

assumption fell within the given range as compared to 61.2% that met both. However, almost all estimates from sample sizes greater than 700 that fell within the accuracy range met both assumptions (e.g., 403/452 for sample size 700).

Similar results were observed when evaluating the effect of each assumption alone. However, contrary to the effect on case-only estimates, violation of the independenc alone had a stronger impact on accuracy of case-control estimates. Interestingly, this effect was in a positive direction such that accuracy of case-control estimates increased when the independence assumption was violated. For example, among estimates in which the Type 2 assumption was violated, a higher proportion of estimates that did not meet the independence assumption fell within range (28.6%) as compared to those that met the independence assumption (11.9%) for sample size 700 (Table 5-33, "Column Percent") . In contrast, violation of the Type 2 assumption alone led to a reduction in accuracy (28.6% versus 47.1%). Thus, violation of both assumptions has a strong effect on case-control estimates relative to the case-only study. Most notably, violation of the independence assumption had a much stronger effect on case-control study results by improving accuracy as compared to the case-only study in which the effects were minimal (Tables 5-32 and 5-33).

As expected, analysis of results from Population 1 representing the worst-case scenario revealed a considerable lack of precision even at the highest sample size of 1000. Thus, lack of precision heavily weighted the results toward inaccuracy, rendering the effects of bias irrelevant. Case-only and case-control study results were also unaffected by presence of bias due to sufficiently high precision, especially with respect to case-only estimates (Table 5-36, "Column Percent"). Case-control estimates by comparison were



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

shown to be less precise, thus more sensitive to the effects of violation of assumptions. For example, at sample size 200, 43.9% of case-control estimates that met neither assumption fell within the acceptable range while 81.6% of those that met both assumptions met accuracy criteria (Table 5-37, "Column Percent"). However, effects of precision obscured those of bias for sample sizes above 300.

Overall, these observations support results obtained from the analysis of variance discussed above. Most importantly, these findings confirm that accuracy of both case-only estimates (COR) and case-control estimates (SIM) strongly depends on sample size, and that the effects of both assumptions on accuracy of estimates depend on, and are thus more meaningful when viewed in relation to the size of the study. Figure 5-1 graphically presents the effect of sample size on GEI estimates obtained from a population in which both assumptions are satisfied (i.e., Typical population) but where the violation of either or both independence or Type 2 assumptions is violated in the sample.

Additionally, the effect of both assumptions on accuracy of estimates followed the same pattern among the three populations, while accuracy improved across the populations, indicating that other design parameters such as prevalence of gene and magnitude of effect also interact with the assumptions to influence accuracy of estimates, as was seen in the analysis of variance results. Thus, differences in accuracy resulting from sample size as seen between range categories diminished the effect of either assumption at the two extreme sample sizes.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

#### 5.3.1.4 Effect of violation of the independence assumption in the population on accuracy of case-only and case-control estimates

We next explored whether violation of either assumption in the population would significantly impact accuracy of either estimate (COR or SIM) and, if so, at what magnitude. To that end, we created two additional populations (34 and 35) with parameters identical to those of the Typical Population but in which the independence assumption was violated to different degrees (odds ratio of association between genotype and exposure among controls = 1.5 and 3, respectively). Results are presented in Tables 5-38 and 5-39. As compared to the Typical Population (Table 5-4), accuracy of case-only estimates to detect GEI interaction effects was reduced commensurate with increasingly larger odds ratio estimates for independence. As illustrated in Table 5-39, the reduction in accuracy was especially pronounced when departure from the independence assumption was large (odds ratio of association between genotype and exposure among controls=3). In contrast, the accuracy of the case-control estimates (SIM) improved with increasing violations of the independence assumption, a phenomenon likely due to a shift in group frequencies from which the estimate is derived, resulting in larger numbers among groups with low frequencies such as the exposed susceptible controls (Table 5-38). As before, sample size interacted significantly with the independence assumption to affect accuracy of estimates. This interactive effect is demonstrated in Figures 5-2 and 5-3. As can be seen, the magnitude and direction of effect of violation of the independence assumption on case-only and case-control estimates depends on study size. For example, in Populations 34 and 35 (Figures 5-2 and 5-3), violation of the independence assumption reduces accuracy of case-only estimates



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

at all sample sizes compared to the Typical population. However, the effects are more pronounced at higher sample sizes as evidenced by the ability of the case-control design to more accurately predict "true" risk at higher sample sizes while the case-only design still out-performs the case-control estimates at lower sample sizes.

Finally, we determined the proportion of samples drawn from each of the two populations in which the independence assumption was met as determined by the accuracy of odds ratios for association between gene and exposure among controls around the expected odds ratio of unity (Tables 5-40 and 5-41). We found that samples drawn from these two populations less frequently and less closely met the independence assumption than samples taken from populations in which the independence assumption was actually met. Moreover, the accuracy of estimates declined proportional to larger departures from the assumption. For example, 40.9% of the odds ratio for independence (odds ratio of association between genotype and exposure among controls) were within ±30% of the expected ratio of unity for a study of size 600 in the Typical Population (Table 5-10) as compared to 29.5% in Population 34 (Table 5-40) and only 4.4% in Population 35 (Table 5-41) for the same study size.

### 5.3.2 Odds ratio for exposure alone

Main effects were estimated for all replicates as compared to only 49% of replicates for which GEI could be computed. Risk estimates of main effects (exposure alone) were significantly more accurate than GEI estimates (Table 5-12); 18.2% of estimates from the smallest sample size (n=100) were contained within the smallest range (±10%) of the "true" value. The proportion of estimates in the smallest range category increased considerably by doubling the sample size up to sample size 400 (36.4%).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

While the percentage of estimates in the smallest range continued to increase with increasing sample sizes, differences were less dramatic at larger sample sizes. For example, increasing the sample size by 100 led to only a 2% increase in the proportion of estimates in the ±10% range for sample sizes 400, 600, and 900. However, doubling sample sizes continued to make a big difference as shown by the approximately 15% increase in estimates from sample size 500 (38.8%) to 1000 (53.2%). Increasing sample sizes had a much smaller effect when the range around the "true" risk was increased to ±50%.

**5.4 Discussion**

The case-only study design (a modification of the case-control study) has been proposed as a more precise alternative to the case-control design in estimation of gene-environment interaction effects due to its elimination of controls and use of less variables resulting in a reduction of systematic (reduced bias) and random (increased precision) errors. Comparing the relative efficiencies of the two designs using hypothetical populations, we were able to confirm that estimates of gene-environment interaction (GEI) from a case-only study more accurately approximate the true (population) risks as compared to the same estimates obtained from its parent design. In fact, we demonstrated that the case-only study leads to more precise estimates irrespective of sample size of the study (n=100 to 1000), magnitude of risk, and frequencies of disease, exposure, and susceptibility genotype in the population.

However, we also illustrated that results of GEI interaction from the case-only study are much less accurate than case-control estimates if the independence assumption is violated in the population, an assumption that, unfortunately, cannot be verified while



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

conducting a real case-only study due to lack of controls. We further demonstrated that, when samples are drawn from populations in which the independence assumption is met but the Type 2 assumption is not satisfied, the case-only design yields more precise estimates than the case-control design.

We assessed the ability of each design to accurately estimate the "true" risk in the population by measuring the difference between risk estimates obtained from each design and the known "true" risk in our hypothetical populations. More than half the time, neither the case-only nor the case-control design had the ability to predict the "true" risk of gene-environment interaction in the population within a ±10% range given realistic study parameters. Case-control risk estimates of main effects (exposure and genotype) were more reliable, but less than 40% of estimates from a study of size 600 (1:1 ratio cases and controls) predicted the "true" risk within ±10%. As expected, the percentage of estimates from both study designs falling into wider intervals around the "true" risk was much higher. The accuracy of estimates also improved as the sample size and/or frequencies of disease, exposure, and genotype increased.

### 5.4.1 Strengths of the simulation study

We constructed a hypothetical population representing frequencies expected in a real-world population with respect to prevalences of our disease and risk factors of interest (Parkinson's disease, CYP2D6 polymorphism, and pesticide exposure in a genetically homogeneous Caucasian population). Moreover, the odds ratios for exposure and gene-environment interaction on development of the disease were designed to represent realistic risks in the population. All parameters used to construct the Typical population



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

were selected after careful and thorough review of the most recent information available in published reports.

Using parameters from our Typical population, we constructed thirty five additional populations in which we systematically modified the parameters of interest to represent lower and higher values than expected in order to assess a broad range of realistic and possible scenarios related to genotype, exposure, disease, magnitude of risks, and effect of underlying assumptions on risk estimates. In particular, evaluation of the presence and impact of violation of the independence assumption on the accuracy of case-only study results was a unique contribution given that it is impossible to ascertain this information in a real case-only investigation.

Another important advantage of our simulations was to allow comparisons of the relative accuracies of case-control and case-only estimates given identical characteristics of samples (i.e., both estimates could be obtained from the same samples). Furthermore, we were able to demonstrate the impact of 10 different sample sizes ranging from very small (n=100) to often prohibitively large (n=1000) on both designs under a wide range of scenarios. Lastly, results of our simulations are generalizable beyond genetic and environmental risk factors for Parkinson's disease to other Type 2 models of gene-environment interactions with parameters falling into our defined parameter space.

With respect to variability in our estimates, we used sufficiently large enough numbers of randomly generated replicates to achieve relative random errors of estimates below 10% at 95% confidence levels for reasonable accuracy of estimates (±30% range of "true" risk). Programming error was minimized by checking the customized application specific to our simulations for programming and rounding errors by hand



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

calculations of all population risks and a random sample of estimates. Consistency was also ascertained by running programs in duplicate to ensure that we obtain identical results each time. Finally, we referenced all commercial programming algorithms used in simulations (e.g., algorithms to generate random numbers, population frequencies, and estimates).

### 5.4.2 Limitations of the simulation study

In keeping with the purpose of other projects in this dissertation, results of our simulations were limited to only Type 2 GEI interactions and cannot be extrapolated to other equally important biologic models of interactions. Additionally, we concentrated on the typical scenario to specifically represent a realistic scenario for the interaction of CYP2D6 poor metabolizer phenotype and pesticide exposure in the pathogenesis of Parkinson's disease among Caucasians, thus further limiting the generalizability of our simulations. However, we addressed this limitation by constructing additional populations to represent other situations such as lower interaction effects and frequencies more commonly observed in environmental and occupational epidemiologic investigations.

Reported estimates were also subject to some unavoidable rounding error due to use of non-integer numbers in the simulations. However, differences in results were negligible as verified by hand calculations of the exact numbers. Another limitation was that the number of replicates used in our simulations (1000 for each sample size) resulted in relative standard errors greater than 10% for our most conservative estimates (±10% range of "true" risk) and for the smallest sample sizes used, although the error rate was reduced to below 10% for sample sizes greater than 200. Moreover, we obtained



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

negligible standard errors for estimates accurate within the more realistic range of ±30% around the "true" risk.

### 5.5 Conclusions

Our findings suggest that results obtained from retrospective studies using a case-control design may not accurately reflect the actual risk in the population. In particular, we show that GEI interaction effects as estimated in a case-control study are far less accurate than estimates of main effects from the same study, especially when the genotype and exposure are uncommon in the population. Hence, case-control studies in which the objective is to estimate effects main as well as multiplicative interaction effects, will require considerably larger sample sizes in order to achieve adequate power to detect the interaction. We further demonstrate that when determination of Type 2 GEI effects is the only objective in a study (i.e., main effects are not of interest), the case-only design yields more accurate estimates than those obtained from a case-control study. Importantly, our results indicate that violation of the independence assumption (as indicated by presence of an association between gene and exposure among controls) in the sample has little effect on case-only estimates if the assumption is satisfied in the underlying population and that the case-only design may lead to more accurate estimates than the case-control study even in such cases. However, we also demonstrate the failure of the case-only study to produce more accurate estimates when the independence assumption is severely violated in the population, therefore, case-only investigations should only be considered when there is reasonable evidence of independence between genotype and exposure in the general population. Our simulation results provide important information regarding the necessity and impact of the independence



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

assumption on the accuracy of case-only study results. Additionally, the results impart valuable comparisons of the accuracy of case-only versus case-control estimates for detection of a Type 2 gene-environment interaction under a broad range of possible conditions.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-1. Summary of input parameters for each hypothetical population.**

| Populations N= 100000 | Population parameters | | | | | |
|---|---|---|---|---|---|---|
| | Prevalence of Genotype | Prevalence of Exposure | Prevalence of Disease | $OR_{GE}$[1] | COR[2] SIM[3] | COOR[4] |
| Typical | 7% | 10% | 5% | 6 | 3.0 | 1 |
| 1 | 3.5% | 5% | 2.5% | 3 | 1.5 | 1 |
| 2 | 3.5% | 5% | 2.5% | 3 | 6.0 | 1 |
| 3 | 3.5% | 5% | 2.5% | 12 | 1.5 | 1 |
| 4 | 3.5% | 5% | 2.5% | 12 | 6.0 | 1 |
| 5 | 3.5% | 5% | 10% | 3 | 1.5 | 1 |
| 6 | 3.5% | 5% | 10% | 3 | 6.0 | 1 |
| 7 | 3.5% | 5% | 10% | 12 | 1.5 | 1 |
| 8 | 3.5% | 5% | 10% | 12 | 6.0 | 1 |
| 9 | 3.5% | 20% | 2.5% | 3 | 1.5 | 1 |
| 10 | 3.5% | 20% | 2.5% | 3 | 6.0 | 1 |
| 11 | 3.5% | 20% | 2.5% | 12 | 1.5 | 1 |
| 12 | 3.5% | 20% | 2.5% | 12 | 6.0 | 1 |
| 13 | 3.5% | 20% | 10% | 3 | 1.5 | 1 |
| 14 | 3.5% | 20% | 10% | 3 | 6.0 | 1 |
| 15 | 3.5% | 20% | 10% | 12 | 1.5 | 1 |
| 16 | 3.5% | 20% | 10% | 12 | 6.0 | 1 |
| 17 | 14% | 5% | 2.5% | 3 | 1.5 | 1 |
| 18 | 14% | 5% | 2.5% | 3 | 6.0 | 1 |
| 19 | 14% | 5% | 2.5% | 12 | 1.5 | 1 |
| 20 | 14% | 5% | 2.5% | 12 | 6.0 | 1 |
| 21 | 14% | 5% | 10% | 3 | 1.5 | 1 |
| 22 | 14% | 5% | 10% | 3 | 6.0 | 1 |
| 23 | 14% | 5% | 10% | 12 | 1.5 | 1 |
| 24 | 14% | 5% | 10% | 12 | 6.0 | 1 |
| 25 | 14% | 20% | 2.5% | 3 | 1.5 | 1 |
| 26 | 14% | 20% | 2.5% | 3 | 6.0 | 1 |
| 27 | 14% | 20% | 2.5% | 12 | 1.5 | 1 |
| 28 | 14% | 20% | 2.5% | 12 | 6.0 | 1 |
| 29 | 14% | 20% | 10% | 3 | 1.5 | 1 |
| 30 | 14% | 20% | 10% | 3 | 6.0 | 1 |
| 31 | 14% | 20% | 10% | 12 | 1.5 | 1 |
| 32 | 14% | 20% | 10% | 12 | 6.0 | 1 |
| 33 | 7% | 10% | 5% | 1.5 | 1.2 | 1 |
| 34 | 7% | 10% | 5% | 6 | 3.0 | 1.5 |
| 35 | 7% | 10% | 5% | 6 | 3.0 | 3 |

[1] $OR_{GE}$ = Odds ratio of disease for those with both the susceptibility genotype and environmental exposure relative to those without either risk factor.
[2] COR = Departure from multiplicative model of interaction in a case-only study.
[3] SIM = Deviation from multiplicative model of interaction in a case-control study.
[4] COOR= odds ratio for association between genotype and exposure among controls (This ratio is equal to one when the independence assumption holds).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-2. Relative standard errors of estimates for percentage of estimates within 10, 30, and 50% ranges[3-5] of the "true" odds ratios for the Typical population based on 1000 sample replicates in the simulation study (SIM=3 or COR=3).**

| Number of subjects | 10% Range[3] | SE[1] | RSE[2] | 30% Range[4] | SE | RSE | 50% Range[5] | SE | RSE |
|---|---|---|---|---|---|---|---|---|---|
| Case-control study design | | | | | | | | | |
| 100 | 11.9 | 1.0 | 8.6 | 9.9 | 0.9 | 9.1 | 18.4 | 1.2 | 0.07 |
| 200 | 3.4 | 0.6 | 16.9 | 22.7 | 1.3 | 5.7 | 39.1 | 1.5 | 0.04 |
| 300 | 13.0 | 1.1 | 8.2 | 28.6 | 1.4 | 4.9 | 47.8 | 1.6 | 0.03 |
| 400 | 7.3 | 0.8 | 11.3 | 34.4 | 1.5 | 4.4 | 55.6 | 1.6 | 0.03 |
| 500 | 18.6 | 1.2 | 6.6 | 36.2 | 1.5 | 4.1 | 60.5 | 1.5 | 0.02 |
| 600 | 10.5 | 1.0 | 9.2 | 37.5 | 1.5 | 4 | 59.4 | 1.6 | 0.03 |
| 700 | 19.2 | 1.2 | 6.5 | 42.8 | 1.6 | 3.7 | 65.7 | 1.5 | 0.02 |
| 800 | 12.2 | 1.0 | 8.5 | 48.7 | 1.6 | 3.3 | 68.7 | 1.5 | 0.02 |
| 900 | 24.9 | 1.4 | 5.5 | 50.4 | 1.6 | 3.2 | 73.9 | 1.4 | 0.02 |
| 1000 | 12.3 | 1.0 | 8.4 | 52.4 | 1.6 | 3.1 | 74.8 | 1.4 | 0.02 |
| Case-only study design | | | | | | | | | |
| 100 | 26.1 | 1.4 | 5.3 | 32.2 | 1.5 | 4.7 | 51.8 | 1.6 | 0.03 |
| 200 | 13.1 | 1.1 | 8.1 | 43.3 | 1.6 | 3.7 | 69.2 | 1.5 | 0.02 |
| 300 | 30.7 | 1.5 | 4.8 | 53.6 | 1.6 | 3 | 77.4 | 1.3 | 0.02 |
| 400 | 13.9 | 1.1 | 7.9 | 55.8 | 1.6 | 2.9 | 83 | 1.2 | 0.01 |
| 500 | 33.6 | 1.5 | 4.4 | 66.4 | 1.5 | 2.3 | 88.6 | 1 | 0.01 |
| 600 | 16.9 | 1.2 | 7.0 | 73.1 | 1.4 | 1.9 | 92.9 | 0.8 | 0.01 |
| 700 | 37.0 | 1.5 | 4.1 | 76.6 | 1.3 | 1.7 | 94.1 | 0.7 | 0.01 |
| 800 | 15.4 | 1.1 | 7.4 | 80.5 | 1.3 | 1.6 | 96.6 | 0.6 | 0.01 |
| 900 | 40.7 | 1.6 | 3.8 | 81.1 | 1.2 | 1.5 | 96.5 | 0.6 | 0.01 |
| 1000 | 17.2 | 1.2 | 6.9 | 85.7 | 1.1 | 1.3 | 97.5 | 0.5 | 0.01 |

[1] SE = Standard error of the estimate.

[2] RSE = Relative standard error of the estimate.

[3] Percentage of all sample SIM or COR estimates within ±10% of "true" SIM or COR.

[4] Percentage of all sample SIM or COR estimates within ±30% of "true" SIM or COR.

[5] Percentage of all sample SIM or COR estimates within ±50% of "true" SIM or COR.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-3. Replicates with missing frequencies.**

| Population | PrevG[1] | PrevE[2] | PrevD[3] | $OR_{GE}$[4] | COR[5] | Number of subjects | Frequency[6] | Percent replicates valid in case-only model[7] |
|---|---|---|---|---|---|---|---|---|
| 1 | 3.5% | 5% | 2.5% | 3 | 1.5 | 100 | 78 | 7.8 |
| 1 | 3.5% | 5% | 2.5% | 3 | 1.5 | 200 | 453 | 45.3 |
| 2 | 3.5% | 5% | 2.5% | 3 | 6.0 | 100 | 78 | 7.8 |
| 2 | 3.5% | 5% | 2.5% | 3 | 6.0 | 200 | 454 | 45.4 |
| 3 | 3.5% | 5% | 2.5% | 12 | 1.5 | 100 | 72 | 7.2 |
| 3 | 3.5% | 5% | 2.5% | 12 | 1.5 | 200 | 455 | 45.5 |
| 4 | 3.5% | 5% | 2.5% | 12 | 6.0 | 100 | 78 | 7.8 |
| 5 | 3.5% | 5% | 10% | 3 | 1.5 | 100 | 78 | 7.8 |
| 5 | 3.5% | 5% | 10% | 3 | 1.5 | 200 | 453 | 45.3 |
| 6 | 3.5% | 5% | 10% | 3 | 6.0 | 100 | 77 | 7.7 |
| 6 | 3.5% | 5% | 10% | 3 | 6.0 | 200 | 454 | 45.4 |
| 7 | 3.5% | 5% | 10% | 12 | 1.5 | 100 | 65 | 6.5 |
| 7 | 3.5% | 5% | 10% | 12 | 1.5 | 200 | 455 | 45.5 |
| 8 | 3.5% | 5% | 10% | 12 | 6.0 | 100 | 77 | 7.7 |
| 9 | 3.5% | 20% | 2.5% | 3 | 1.5 | 100 | 78 | 7.8 |
| 9 | 3.5% | 20% | 2.5% | 3 | 1.5 | 200 | 455 | 45.5 |
| 10 | 3.5% | 20% | 2.5% | 3 | 6.0 | 100 | 63 | 6.3 |
| 11 | 3.5% | 20% | 2.5% | 12 | 1.5 | 100 | 78 | 7.8 |
| 13 | 3.5% | 20% | 10% | 3 | 1.5 | 100 | 78 | 7.8 |
| 13 | 3.5% | 20% | 10% | 3 | 1.5 | 200 | 455 | 45.5 |
| 14 | 3.5% | 20% | 10% | 3 | 6.0 | 100 | 78 | 7.8 |
| 15 | 3.5% | 20% | 10% | 12 | 1.5 | 100 | 78 | 7.8 |
| 15 | 3.5% | 20% | 10% | 12 | 1.5 | 200 | 455 | 45.5 |

[1] PrevG= prevalence of genotype.
[2] PrevE= prevalence of exposure.
[3] PrevD= prevalence of disease.
[4] $OR_{GE}$ = Odds ratio of disease for those with both the susceptibility genotype and environmental exposure relative to those without either risk factor.
[5] COR = Departure from multiplicative model of interaction in a case-only study.
[6] Maximum number of frequencies = 1000.
[7] Percentage of replicates that contained at least one diseased subject with both genoytpe and exposure.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-4. Summary of the simulation study results[1] for gene-environment interaction with respect to case-only and case-control designs in the Typical Population[2]: Percentage of estimated SIM or COR[3] values that fall within specified ranges of "true" SIM or COR value (SIM =3 or COR = 3).**

| | | | Percentage of all sample SIM or COR estimates within specified range of true SIM or COR | | | | |
|---|---|---|---|---|---|---|---|
| | | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Number of Subjects | Study Design | Percentage of valid replicates[4] | SIM or COR= 2.7 to 3.3 | SIM or COR= 2.4 to 3.6 | SIM or COR= 2.1 to 3.9 | SIM or COR= 1.8 to 4.2 | SIM or COR= 1.5 to 4.5 |
| 100 | Case-only | 97.5 | 11.9 | 21.6 | 32.2 | 43.2 | 51.8 |
| | Case-control | 48.6 | 3.4 | 6.5 | 9.9 | 13.7 | 18.4 |
| 200 | Case-only | 99.8 | 13.0 | 29.8 | 43.3 | 57.4 | 69.2 |
| | Case-control | 72.5 | 7.3 | 14.1 | 22.7 | 30.9 | 39.1 |
| 300 | Case-only | 100.0 | 18.6 | 36.5 | 53.6 | 66.9 | 77.4 |
| | Case-control | 86.1 | 10.5 | 19.0 | 28.6 | 38.9 | 47.8 |
| 400 | Case-only | 100.0 | 19.2 | 36.5 | 55.8 | 71.9 | 83.0 |
| | Case-control | 92.2 | 12.2 | 22.9 | 34.4 | 45.8 | 55.6 |
| 500 | Case-only | 100.0 | 24.9 | 47.8 | 66.4 | 80.0 | 88.6 |
| | Case-control | 95.3 | 12.3 | 22.8 | 36.2 | 48.3 | 60.5 |
| 600 | Case-only | 100.0 | 26.1 | 50.2 | 73.1 | 86.2 | 92.9 |
| | Case-control | 96.2 | 13.1 | 25.7 | 37.5 | 50.7 | 59.4 |
| 700 | Case-only | 100.0 | 30.7 | 56.0 | 76.6 | 87.6 | 94.1 |
| | Case-control | 97.3 | 13.9 | 28.4 | 42.8 | 54.4 | 65.7 |
| 800 | Case-only | 100.0 | 33.6 | 59.8 | 80.5 | 91.3 | 96.6 |
| | Case-control | 98.0 | 16.9 | 32.7 | 48.7 | 60.4 | 68.7 |
| 900 | Case-only | 100.0 | 37.0 | 65.4 | 81.1 | 92.3 | 96.5 |
| | Case-control | 98.9 | 15.4 | 33.0 | 50.4 | 64.3 | 73.9 |
| 1000 | Case-only | 100.0 | 40.7 | 68.4 | 85.7 | 93.8 | 97.5 |
| | Case-control | 99.5 | 17.2 | 36.0 | 52.4 | 64.9 | 74.8 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR = Departure from multiplicative model of interaction in a case-only study;
SIM = Departure from multiplicative model of interaction in a case-control study.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-5. Summary of the simulation study results[1] for gene-environment interaction with respect to case-only and case-control designs in the Typical Population[2]: Percentage of estimated averages from exponentiated coefficients[3] [ln(SIM) or ln(COR)] that fall within specified ranges of "true" exponentiated coefficient of SIM or COR (beta=1.10).**

| | | | Percentage of all estimated coefficients within specified range of true coefficient | | | | |
|---|---|---|---|---|---|---|---|
| | | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Study Design | Percentage of valid Replicates[4] | SIM or COR = 2.7-3.4 | SIM or COR= 2.4-3.7 | SIM or COR= 2.2-4.2 | SIM or COR= 1.9-4.6 | SIM or COR= 1.7-5.1 |
| 100 | Case-only | 96.9 | 13.9 | 24.1 | 34.3 | 43.1 | 52.4 |
| | Case-control | 46.4 | 3.7 | 6.4 | 9.8 | 12.8 | 16.4 |
| 200 | Case-only | 99.7 | 14.6 | 32.7 | 46.3 | 60.6 | 70.7 |
| | Case-control | 72.2 | 8.0 | 15.5 | 23.6 | 31.0 | 38.0 |
| 300 | Case-only | 100.0 | 20.9 | 39.1 | 56.1 | 69.1 | 79.0 |
| | Case-control | 86.1 | 11.9 | 20.4 | 29.7 | 37.4 | 47.8 |
| 400 | Case-only | 100.0 | 21.4 | 41.7 | 59.4 | 75.1 | 85.4 |
| | Case-control | 92.3 | 13.9 | 24.5 | 35.5 | 46.8 | 54.2 |
| 500 | Case-only | 100.0 | 28.3 | 51.9 | 69.0 | 82.1 | 90.2 |
| | Case-control | 95.9 | 14.6 | 25.1 | 38.2 | 48.4 | 58.9 |
| 600 | Case-only | 100.0 | 29.5 | 55.1 | 76.2 | 87.6 | 94.3 |
| | Case-control | 97.5 | 14.7 | 27.8 | 39.4 | 50.2 | 61.9 |
| 700 | Case-only | 100.0 | 35.4 | 60.4 | 79.2 | 90.5 | 96.5 |
| | Case-control | 98.8 | 15.1 | 30.6 | 44.8 | 55.8 | 65.0 |
| 800 | Case-only | 100.0 | 37.7 | 65.4 | 84.4 | 93.2 | 98.2 |
| | Case-control | 99.3 | 19.5 | 35.7 | 50.8 | 62.0 | 70.7 |
| 900 | Case-only | 100.0 | 42.1 | 69.0 | 84.5 | 94.1 | 98.2 |
| | Case-control | 99.7 | 17.7 | 36.2 | 52.2 | 64.5 | 75.4 |
| 1000 | Case-only | 100.0 | 45.8 | 71.9 | 87.2 | 95.8 | 98.9 |
| | Case-control | 100.0 | 19.8 | 37.8 | 54.2 | 66.6 | 75.5 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR = Departure from multiplicative model of interaction in a case-only study;
SIM = Departure from multiplicative model of interaction in a case-control study.
[4]Percentage of replicates included (fitted logistic regression model converged or -20 < beta < 20).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-6. Summary of the simulation study results[1] for gene-environment interaction with respect to case-only and case-control designs in Population 1[2]: Percentage of estimated SIM or COR[3] values that fall within specified ranges of "true" SIM or COR value (SIM =1.5 or COR = 1.5).**

| | | | Percentage of all sample SIM or COR estimates within specified range of true SIM or COR | | | | |
|---|---|---|---|---|---|---|---|
| | | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Study Design | Percentage of valid replicates[4] | SIM or COR= 1.35 to 1.65 | SIM or COR= 1.20 to 1.80 | SIM or COR= 1.05 to 1.95 | SIM or COR= 0.9 to 2.10 | SIM or COR= 0.75 to 2.25 |
| 100 | Case-only | 0.3 | 3.9 | 5.1 | 5.1 | 5.1 | 5.1 |
| | Case-control | 0.1 | 0.2 | 0.8 | 2.0 | 3.7 | 5.9 |
| 200 | Case-only | 28.5 | 4.2 | 7.0 | 9.5 | 12.9 | 14.9 |
| | Case-control | 19.2 | 0.8 | 1.6 | 3.6 | 4.1 | 5.9 |
| 300 | Case-only | 100.0 | 10.4 | 19.3 | 28.7 | 38.1 | 47.0 |
| | Case-control | 42.6 | 1.6 | 3.7 | 6.4 | 9.1 | 13.4 |
| 400 | Case-only | 100.0 | 10.9 | 20.3 | 29.6 | 39.0 | 49.1 |
| | Case-control | 51.7 | 4.9 | 8.8 | 13.0 | 17.8 | 21.6 |
| 500 | Case-only | 100.0 | 12.7 | 24.6 | 35.8 | 46.6 | 54.8 |
| | Case-control | 58.1 | 4.6 | 10.7 | 16.1 | 21.2 | 27.2 |
| 600 | Case-only | 100.0 | 12.7 | 24.6 | 35.8 | 46.6 | 54.8 |
| | Case-control | 59.8 | 4.7 | 11.0 | 16.6 | 21.8 | 28.0 |
| 700 | Case-only | 100.0 | 14.2 | 28.7 | 42.0 | 57.3 | 67.3 |
| | Case-control | 71.1 | 8.4 | 16.0 | 24.1 | 31.0 | 38.8 |
| 800 | Case-only | 100.0 | 16.5 | 32.3 | 47.4 | 60.2 | 71.4 |
| | Case-control | 76.7 | 7.5 | 15.9 | 25.2 | 35.7 | 43.1 |
| 900 | Case-only | 100.0 | 17.7 | 35.4 | 52.4 | 65.1 | 76.1 |
| | Case-control | 79.5 | 8.2 | 17.8 | 24.9 | 35.7 | 44.2 |
| 1000 | Case-only | 100.0 | 21.0 | 38.0 | 54.2 | 67.4 | 78.0 |
| | Case-control | 83.9 | 8.1 | 17.9 | 28.3 | 37.8 | 47.1 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR = Departure from multiplicative model of interaction in a case-only study;
SIM = Departure from multiplicative model of interaction in a case-control study.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-7. Summary of the simulation study results[1] for gene-environment interaction with respect to case-only and case-control designs in Population 32[2]: Percentage of estimated SIM or COR[3] values that fall within specified ranges of "true" SIM or COR value (SIM =6 or COR = 6).**

| | | | Percentage of all sample SIM or COR estimates within specified range of true SIM or COR | | | | |
|---|---|---|---|---|---|---|---|
| | | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Study Design | Percentage of valid replicates[4] | SIM or COR= 5.4 to 6.6 | SIM or COR= 4.8 to 7.2 | SIM or COR= 4.2 to 7.8 | SIM or COR= 3.6 to 8.4 | SIM or COR= 3.0 to 9.0 |
| 100 | Case-only | 94.1 | 16.2 | 33.0 | 45.8 | 59.1 | 69.9 |
| | Case-control | 80.7 | 8.3 | 16.9 | 26.7 | 34.5 | 44.1 |
| 200 | Case-only | 98.5 | 21.2 | 41.3 | 60.1 | 74.2 | 82.7 |
| | Case-control | 88.5 | 14.1 | 28.7 | 43.2 | 53.1 | 63.9 |
| 300 | Case-only | 99.8 | 28.7 | 52.8 | 70.8 | 83.0 | 89.0 |
| | Case-control | 94.2 | 18.2 | 34.0 | 49.8 | 63.7 | 74.3 |
| 400 | Case-only | 99.9 | 27.0 | 56.7 | 75.3 | 86.7 | 91.5 |
| | Case-control | 95.7 | 20.2 | 38.6 | 53.9 | 68.9 | 78.9 |
| 500 | Case-only | 100.0 | 35.8 | 63.0 | 82.1 | 91.9 | 96.0 |
| | Case-control | 97.9 | 22.0 | 43.2 | 61.9 | 75.8 | 83.4 |
| 600 | Case-only | 100.0 | 38.9 | 67.1 | 85.6 | 94.5 | 97.3 |
| | Case-control | 98.9 | 24.0 | 44.6 | 63.9 | 78.0 | 86.3 |
| 700 | Case-only | 100.0 | 40.8 | 72.3 | 91.4 | 97.1 | 98.8 |
| | Case-control | 99.2 | 26.0 | 50.8 | 70.3 | 82.9 | 89.9 |
| 800 | Case-only | 100.0 | 44.5 | 77.1 | 92.8 | 97.2 | 98.8 |
| | Case-control | 99.3 | 26.6 | 52.8 | 73.1 | 84.6 | 90.5 |
| 900 | Case-only | 100.0 | 46.5 | 80.0 | 94.6 | 98.3 | 99.5 |
| | Case-control | 99.8 | 31.8 | 58.6 | 77.3 | 87.8 | 92.3 |
| 1000 | Case-only | 100.0 | 47.5 | 79.4 | 94.1 | 98.4 | 99.6 |
| | Case-control | 100.0 | 32.8 | 58.4 | 79.8 | 88.9 | 93.0 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR = Departure from multiplicative model of interaction in a case-only study;
SIM = Departure from multiplicative model of interaction in a case-control study.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-8. Summary of the simulation study results[1] for testing the independence assumption (COOR)[2] in Typical Population[3]: Percentage of estimated OR values that fall within specified ranges of "true" COOR (COOR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 100.0 | 3.1 | 5.2 | 7.5 | 11.5 | 13.4 |
| 200 | 100.0 | 7.1 | 16.8 | 27.6 | 34.1 | 41.4 |
| 300 | 100.0 | 10.0 | 18.4 | 26.7 | 35.3 | 47.5 |
| 400 | 100.0 | 12.3 | 23.5 | 33.7 | 44.1 | 52.6 |
| 500 | 100.0 | 13.2 | 27.0 | 37.7 | 50.2 | 60.5 |
| 600 | 100.0 | 11.6 | 26.3 | 40.9 | 51.9 | 63.5 |
| 700 | 100.0 | 13.7 | 28.0 | 41.9 | 56.3 | 67.5 |
| 800 | 100.0 | 16.6 | 31.9 | 47.4 | 60.9 | 71.6 |
| 900 | 100.0 | 16.4 | 34.5 | 52.2 | 66.8 | 76.1 |
| 1000 | 100.0 | 16.9 | 37.1 | 53.3 | 65.2 | 75.0 |

[1] Based on 1000 replicates.
[2] Odds ratio of association between genotype and exposure among controls.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).

**Table 5-9. Summary of the simulation study results[1] for testing the independence assumption (COOR)[2] in Population 1[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 200 | 2.1 | 0.0 | 0.1 | 0.1 | 0.1 | 0.2 |
| 300 | 99.7 | 1.0 | 2.1 | 3.4 | 4.3 | 5.4 |
| 400 | 100.0 | 4.1 | 7.9 | 12.0 | 14.7 | 17.4 |
| 500 | 100.0 | 8.4 | 14.8 | 21.2 | 25.4 | 29.1 |
| 600 | 100.0 | 10.0 | 19.6 | 28.0 | 32.2 | 36.3 |
| 700 | 100.0 | 12.5 | 22.5 | 30.4 | 37.7 | 43.5 |
| 800 | 100.0 | 7.6 | 16.2 | 27.3 | 39.8 | 48.7 |
| 900 | 100.0 | 7.9 | 16.7 | 27.4 | 37.6 | 49.7 |
| 1000 | 100.0 | 8.6 | 16.7 | 26.4 | 36.5 | 50.7 |

[1] Based on 1000 replicates.
[2] Odds ratio of association between genotype and exposure among controls.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-10. Summary of the simulation study results[1] for testing the independence assumption (COOR)[2] in Population 32[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 100.0 | 9.2 | 18.3 | 27.2 | 36.3 | 43.7 |
| 200 | 100.0 | 14.5 | 27.5 | 40.3 | 52.5 | 64.5 |
| 300 | 100.0 | 18.2 | 35.2 | 48.6 | 61.2 | 72.0 |
| 400 | 100.0 | 18.8 | 39.5 | 55.7 | 69.9 | 78.9 |
| 500 | 100.0 | 23.4 | 44.5 | 63.1 | 76.5 | 84.6 |
| 600 | 100.0 | 23.2 | 45.3 | 63.2 | 78.7 | 87.7 |
| 700 | 100.0 | 26.4 | 51.4 | 67.6 | 81.0 | 90.4 |
| 800 | 100.0 | 27.5 | 52.3 | 71.6 | 82.2 | 90.5 |
| 900 | 100.0 | 31.6 | 55.7 | 74.7 | 86.8 | 93.6 |
| 1000 | 100.0 | 30.3 | 58.4 | 76.5 | 88.9 | 94.4 |

[1] Based on 1000 replicates.
[2] Odds ratio of association between genotype and exposure among controls.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).

**Table 5-11. Summary of the simulation study results[1] for testing Type 2 gene-environment interaction assumption ($OR_{Ge}$)[2] in Typical Population[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 100.0 | 11.8 | 23.1 | 35.1 | 45.0 | 57.6 |
| 200 | 100.0 | 19.5 | 37.9 | 52.4 | 65.8 | 77.0 |
| 300 | 100.0 | 24.5 | 44.1 | 62.2 | 76.2 | 85.3 |
| 400 | 100.0 | 28.7 | 53.6 | 69.1 | 83.3 | 89.8 |
| 500 | 100.0 | 28.8 | 57.9 | 76.5 | 87.3 | 94.2 |
| 600 | 100.0 | 35.2 | 61.7 | 79.2 | 91.3 | 95.2 |
| 700 | 100.0 | 37.9 | 67.7 | 84.4 | 92.7 | 96.9 |
| 800 | 100.0 | 38.5 | 69.0 | 87.0 | 94.9 | 97.3 |
| 900 | 100.0 | 38.7 | 70.8 | 89.1 | 95.6 | 98.2 |
| 1000 | 100.0 | 42.1 | 74.5 | 91.2 | 97.3 | 98.8 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease, among unexposed, for susceptibles versus non-susceptibles.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-12. Summary of the simulation study results[1] for testing Type 2 gene-environment interaction assumption ($OR_{Ge}$)[2] in Population 1[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 95.4 | 14.7 | 20.3 | 27.4 | 35.6 | 46.6 |
| 200 | 99.7 | 16.1 | 28.7 | 41.8 | 52.6 | 63.3 |
| 300 | 100.0 | 18.2 | 35.4 | 50.3 | 62.7 | 73.9 |
| 400 | 100.0 | 20.6 | 40.3 | 56.5 | 69.8 | 80.1 |
| 500 | 100.0 | 22.3 | 44.7 | 64.4 | 77.6 | 85.2 |
| 600 | 100.0 | 24.2 | 48.6 | 68.2 | 81.1 | 88.0 |
| 700 | 100.0 | 31.9 | 53.9 | 73.0 | 85.6 | 91.4 |
| 800 | 100.0 | 31.0 | 55.1 | 73.4 | 86.4 | 92.8 |
| 900 | 100.0 | 31.2 | 59.5 | 79.0 | 90.4 | 94.9 |
| 1000 | 100.0 | 34.8 | 61.4 | 82.4 | 92.4 | 96.3 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease, among unexposed, for susceptibles versus non-susceptibles.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).

**Table 5-13. Summary of the simulation study results[1] for testing Type 2 gene-environment interaction assumption ($OR_{Ge}$)[2] in Population 32[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 100.0 | 15.6 | 30.0 | 44.3 | 56.1 | 66.0 |
| 200 | 100.0 | 23.9 | 44.2 | 62.7 | 75.6 | 84.7 |
| 300 | 100.0 | 28.0 | 52.3 | 71.0 | 85.1 | 92.0 |
| 400 | 100.0 | 27.3 | 57.0 | 76.0 | 88.1 | 93.9 |
| 500 | 100.0 | 34.3 | 62.4 | 81.4 | 92.6 | 97.0 |
| 600 | 100.0 | 37.7 | 66.3 | 85.3 | 93.7 | 97.6 |
| 700 | 100.0 | 41.5 | 69.0 | 87.1 | 95.5 | 97.7 |
| 800 | 100.0 | 45.6 | 72.7 | 89.5 | 96.2 | 98.4 |
| 900 | 100.0 | 45.4 | 76.8 | 91.0 | 97.7 | 99.2 |
| 1000 | 100.0 | 48.9 | 76.5 | 93.5 | 98.7 | 99.6 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease, among unexposed, for susceptibles versus non-susceptibles.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-14. Summary of the simulation study results[1] for effects of exposure alone ($OR_{Eg}$)[2] in Typical Population[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=2).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Percentage of valid replicates[4] | OR= 1.8 to 2.2 | OR= 1.6 to 2.4 | OR= 1.4 to 2.6 | OR= 1.2 to 2.8 | OR= 1.0 to 3.0 |
| 100 | 100.0 | 18.3 | 35.6 | 52.9 | 65.7 | 75.6 |
| 200 | 100.0 | 23.8 | 47.4 | 68.4 | 81.1 | 88.9 |
| 300 | 100.0 | 30.1 | 58.0 | 77.4 | 89.2 | 93.3 |
| 400 | 100.0 | 36.4 | 65.8 | 83.3 | 93.4 | 96.7 |
| 500 | 100.0 | 39.0 | 69.5 | 87.6 | 95.0 | 98.0 |
| 600 | 100.0 | 43.7 | 75.9 | 90.5 | 96.9 | 98.7 |
| 700 | 100.0 | 43.1 | 75.8 | 91.9 | 97.4 | 99.2 |
| 800 | 100.0 | 49.4 | 80.8 | 93.8 | 98.5 | 99.7 |
| 900 | 100.0 | 51.4 | 83.4 | 95.6 | 98.8 | 99.7 |
| 1000 | 100.0 | 53.2 | 85.2 | 96.7 | 99.1 | 99.8 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease, among non-susceptibles, for exposed versus unexposed.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).

**Table 5-15. Summary of the simulation study results[1] for effects of exposure alone ($OR_{Eg}$)[2] in Population 1[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=2).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Percentage of valid replicates[4] | OR= 1.8 to 2.2 | OR= 1.6 to 2.4 | OR= 1.4 to 2.6 | OR= 1.2 to 2.8 | OR= 1.0 to 3.0 |
| 100 | 98.9 | 13.4 | 26.1 | 37.6 | 48.5 | 60.6 |
| 200 | 99.7 | 17.6 | 35.4 | 52.8 | 67.9 | 77.7 |
| 300 | 100.0 | 23.2 | 45.4 | 63.0 | 76.9 | 85.7 |
| 400 | 100.0 | 24.9 | 53.1 | 71.9 | 85.2 | 90.9 |
| 500 | 100.0 | 32.0 | 58.3 | 77.7 | 89.8 | 94.2 |
| 600 | 100.0 | 34.6 | 60.2 | 81.1 | 90.8 | 95.4 |
| 700 | 100.0 | 34.9 | 64.6 | 84.7 | 93.2 | 96.9 |
| 800 | 100.0 | 39.1 | 71.1 | 88.5 | 95.1 | 97.3 |
| 900 | 100.0 | 43.0 | 73.9 | 89.2 | 96.6 | 98.9 |
| 1000 | 100.0 | 43.6 | 76.3 | 91.2 | 96.5 | 98.3 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease, among non-susceptibles, for exposed versus unexposed.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-16. Summary of the simulation study results[1] for effects of exposure alone ($OR_{Eg}$)[2] in Population 32[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=2).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Percentage of valid replicates[4] | OR= 1.8 to 2.2 | OR= 1.6 to 2.4 | OR= 1.4 to 2.6 | OR= 1.2 to 2.8 | OR= 1.0 to 3.0 |
| 100 | 100.0 | 18.2 | 37.8 | 54.2 | 68.3 | 79.5 |
| 200 | 100.0 | 28.7 | 55.3 | 74.5 | 86.0 | 92.5 |
| 300 | 100.0 | 35.5 | 62.6 | 83.0 | 93.3 | 96.2 |
| 400 | 100.0 | 38.1 | 69.7 | 87.7 | 95.7 | 97.8 |
| 500 | 100.0 | 44.4 | 74.0 | 90.8 | 97.3 | 99.3 |
| 600 | 100.0 | 49.1 | 78.9 | 93.0 | 97.5 | 99.3 |
| 700 | 100.0 | 53.7 | 83.8 | 95.9 | 98.8 | 99.9 |
| 800 | 100.0 | 56.1 | 87.5 | 96.7 | 99.2 | 99.8 |
| 900 | 100.0 | 59.6 | 89.3 | 97.5 | 99.5 | 100.0 |
| 1000 | 100.0 | 61.6 | 92.1 | 98.8 | 99.8 | 100.0 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease, among non-susceptibles, for exposed versus unexposed.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).

**Table 5-17. Summary of the simulation study results[1] for effects of multiplicative interaction ($OR_{GE}$)[2] in Typical Population[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=6).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Percentage of valid replicates[4] | OR= 5.4 to 6.6 | OR= 4.8 to 7.2 | OR= 4.2 to 7.8 | OR= 3.6 to 8.4 | OR= 3.0 to 9.0 |
| 100 | 48.6 | 3.2 | 6.0 | 11.0 | 14.0 | 21.6 |
| 200 | 72.1 | 9.2 | 15.3 | 25.0 | 32.1 | 41.5 |
| 300 | 80.8 | 10.0 | 20.0 | 30.0 | 39.1 | 48.8 |
| 400 | 82.0 | 10.7 | 24.7 | 35.7 | 47.6 | 57.7 |
| 500 | 85.6 | 13.4 | 26.8 | 38.8 | 51.3 | 61.9 |
| 600 | 89.5 | 13.9 | 26.9 | 37.9 | 50.7 | 61.5 |
| 700 | 92.0 | 13.9 | 30.4 | 44.2 | 58.1 | 68.5 |
| 800 | 93.3 | 19.0 | 34.8 | 50.0 | 61.7 | 69.8 |
| 900 | 94.3 | 19.9 | 36.6 | 52.9 | 65.6 | 74.6 |
| 1000 | 94.8 | 19.1 | 37.7 | 54.7 | 66.9 | 77.4 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease for exposed susceptibles versus unexposed non-susceptibles.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-18. Summary of the simulation study results[1] for effects of multiplicative interaction ($OR_{GE}$)[2] in Population 1[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=3).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Percentage of valid replicates[4] | OR= 2.6 to 3.2 | OR= 2.3 to 3.5 | OR= 2.0 to 3.8 | OR= 1.7 to 4.1 | OR= 1.5 to 4.4 |
| 100 | 35.7 | 0.8 | 1.0 | 2.9 | 3.3 | 3.3 |
| 200 | 30.5 | 0.3 | 0.6 | 4.8 | 5.4 | 5.5 |
| 300 | 42.6 | 2.3 | 3.1 | 11.4 | 11.9 | 13.8 |
| 400 | 50.7 | 5.8 | 8.2 | 17.8 | 18.5 | 23.2 |
| 500 | 58.1 | 7.8 | 9.8 | 21.6 | 22.7 | 31.9 |
| 600 | 64.0 | 9.3 | 12.4 | 23.4 | 24.5 | 36.1 |
| 700 | 71.1 | 9.4 | 12.9 | 26.4 | 27.6 | 41.0 |
| 800 | 76.8 | 11.1 | 16.0 | 30.2 | 32.2 | 45.9 |
| 900 | 79.5 | 12.9 | 17.9 | 31.2 | 33.0 | 48.4 |
| 1000 | 83.9 | 11.9 | 17.8 | 32.3 | 34.9 | 48.7 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease for exposed susceptibles versus unexposed non-susceptibles.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).

**Table 5-19. Summary of the simulation study results[1] for effects of multiplicative interaction ($OR_{GE}$)[2] in Population 32[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=12).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Percentage of valid replicates[4] | OR= 10.8 to 13.2 | OR= 9.6 to 14.4 | OR= 8.4 to 15.6 | OR= 7.2 to 16.8 | OR= 6.0 to 18.0 |
| 100 | 57.2 | 10.0 | 21.5 | 32.0 | 41.0 | 49.5 |
| 200 | 62.5 | 15.1 | 30.0 | 44.0 | 57.8 | 67.6 |
| 300 | 66.2 | 18.0 | 35.7 | 51.3 | 66.0 | 74.7 |
| 400 | 69.2 | 20.5 | 38.0 | 57.6 | 70.6 | 79.8 |
| 500 | 73.0 | 24.3 | 48.4 | 65.5 | 77.6 | 86.8 |
| 600 | 73.5 | 24.5 | 47.9 | 68.3 | 81.6 | 88.0 |
| 700 | 76.9 | 30.2 | 53.7 | 73.7 | 83.3 | 89.3 |
| 800 | 78.5 | 30.1 | 57.1 | 74.4 | 85.9 | 91.1 |
| 900 | 80.6 | 32.6 | 60.2 | 78.5 | 89.3 | 92.4 |
| 1000 | 81.6 | 34.1 | 62.8 | 81.7 | 90.5 | 93.6 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease for exposed susceptibles versus unexposed non-susceptibles.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-20. Summary of the simulation study results[1] for gene-environment interaction with respect to case-only and case-control designs in the Population 33[2]: Percentage of estimated SIM or COR[3] values that fall within specified ranges of "true" SIM or COR value (SIM=1.2 or COR=1.2).**

| | | | Percentage of all sample SIM or COR estimates within specified range of true SIM or COR | | | | |
|---|---|---|---|---|---|---|---|
| | | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Study Design | Percentage of valid replicates[4] | SIM or COR= 1.1 to 1.3 | SIM or COR= 0.96 to 1.4 | SIM or COR= 0.84 to 1.6 | SIM or COR= 0.72 to 1.7 | SIM or COR= 0.60 to 1.8 |
| 100 | Case-only | 99.5 | 7.0 | 12.9 | 21.4 | 25.7 | 31.3 |
| | Case-control | 50.0 | 2.8 | 4.8 | 8.1 | 10.3 | 12.7 |
| 200 | Case-only | 100.0 | 6.4 | 16.7 | 28.8 | 37.8 | 47.2 |
| | Case-control | 73.7 | 4.3 | 9.1 | 17.3 | 21.8 | 27.1 |
| 300 | Case-only | 100.0 | 9.6 | 20.4 | 36.7 | 46.9 | 56.5 |
| | Case-control | 87.3 | 5.6 | 13.1 | 21.4 | 28.1 | 35.0 |
| 400 | Case-only | 100.0 | 12.7 | 29.0 | 45.3 | 56.2 | 66.0 |
| | Case-control | 93.2 | 7.1 | 16.3 | 26.6 | 33.4 | 42.5 |
| 500 | Case-only | 100.0 | 13.9 | 31.5 | 49.5 | 60.3 | 69.1 |
| | Case-control | 96.4 | 8.9 | 17.9 | 30.1 | 40.1 | 48.6 |
| 600 | Case-only | 100.0 | 14.0 | 32.0 | 52.0 | 64.8 | 75.7 |
| | Case-control | 97.8 | 7.7 | 18.8 | 33.1 | 42.6 | 52.6 |
| 700 | Case-only | 100.0 | 17.4 | 36.0 | 56.9 | 70.0 | 80.3 |
| | Case-control | 98.8 | 9.7 | 22.5 | 37.3 | 46.0 | 54.8 |
| 800 | Case-only | 100.0 | 18.0 | 38.0 | 61.0 | 73.5 | 83.3 |
| | Case-control | 99.2 | 10.7 | 24.6 | 39.8 | 50.9 | 62.7 |
| 900 | Case-only | 100.0 | 20.4 | 40.0 | 63.1 | 77.6 | 86.2 |
| | Case-control | 99.7 | 12.6 | 25.3 | 42.7 | 53.5 | 63.6 |
| 1000 | Case-only | 100.0 | 19.5 | 43.2 | 65.5 | 79.1 | 88.9 |
| | Case-control | 100.0 | 11.4 | 26.6 | 42.8 | 54.6 | 65.6 |

[1] Based on 1000 replicates.

[2] See Table 5-1 for population parameters.

[3] COR = Departure from multiplicative model of interaction in a case-only study;
SIM = Departure from multiplicative model of interaction in a case-control study.

[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-21. Summary of the simulation study results[1] for gene-environment interaction with respect to case-only and case-control designs in the Population 33[2]: Percentage of estimated averages from exponentiated coefficients[3] [ln(SIM) or ln(COR)] that fall within specified ranges of "true" exponentiated coefficient of SIM or COR (beta=.18).**

| | | | Percentage of all estimated coefficients within specified range of true coefficient | | | | |
|---|---|---|---|---|---|---|---|
| | | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Study Design | Percentage of valid Replicates[4] | SIM or COR= 1.17-1.21 | SIM or COR= 1.15-1.25 | SIM or COR= 1.14-1.26 | SIM or COR= 1.12-1.28 | SIM or COR= 0.60-1.80 |
| 100 | Case-only | 63.5 | 0.1 | 2.4 | 5.6 | 6.8 | 7.2 |
| | Case-control | 31.4 | 1.0 | 1.5 | 1.6 | 2.2 | 3.0 |
| 200 | Case-only | 85.6 | 1.3 | 3.2 | 4.0 | 5.2 | 7.5 |
| | Case-control | 62.2 | 1.0 | 2.0 | 2.7 | 3.7 | 4.3 |
| 300 | Case-only | 95.8 | 1.2 | 4.1 | 5.4 | 7.6 | 10.7 |
| | Case-control | 84.2 | 1.4 | 2.6 | 3.6 | 4.6 | 5.9 |
| 400 | Case-only | 98.5 | 2.8 | 6.0 | 7.5 | 10.6 | 14.1 |
| | Case-control | 91.7 | 1.5 | 3.8 | 4.4 | 6.0 | 8.0 |
| 500 | Case-only | 99.1 | 3.9 | 6.7 | 8.6 | 11.6 | 14.9 |
| | Case-control | 95.7 | 1.9 | 4.1 | 5.5 | 7.9 | 9.2 |
| 600 | Case-only | 100.0 | 3.7 | 7.2 | 9.3 | 12.2 | 14.9 |
| | Case-control | 97.9 | 1.7 | 3.8 | 4.5 | 6.1 | 8.8 |
| 700 | Case-only | 100.0 | 3.8 | 8.1 | 10.1 | 14.6 | 18.7 |
| | Case-control | 98.9 | 1.7 | 4.0 | 5.1 | 7.9 | 10.7 |
| 800 | Case-only | 99.9 | 4.4 | 8.8 | 11.0 | 16.0 | 19.4 |
| | Case-control | 99.2 | 3.1 | 6.2 | 7.2 | 9.4 | 11.2 |
| 900 | Case-only | 100.0 | 4.9 | 8.7 | 11.7 | 17.3 | 21.8 |
| | Case-control | 99.7 | 2.4 | 5.2 | 6.3 | 9.8 | 12.9 |
| 1000 | Case-only | 100.0 | 4.9 | 10.2 | 11.6 | 16.3 | 21.4 |
| | Case-control | 100.0 | 3.6 | 5.9 | 7.2 | 9.3 | 12.0 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR = Departure from multiplicative model of interaction in a case-only study;
SIM = Departure from multiplicative model of interaction in a case-control study.
[4] Percentage of replicates included (fitted logistic regression model converged or -20 < beta < 20).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-22. Summary of the simulation study results[1] for effects of multiplicative interaction ($OR_{GE}$)[2] in Population33[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1.5).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Percentage of valid replicates[4] | OR= 1.4 to 1.7 | OR= 1.2 to 1.8 | OR= 1.1 to 2.0 | OR= .90 to 2.1 | OR= .75 to 2.3 |
| 100 | 48.6 | 50.0 | 2.4 | 5.1 | 7.9 | 11.2 |
| 200 | 72.1 | 73.8 | 5.4 | 11.4 | 16.3 | 22.9 |
| 300 | 80.8 | 87.3 | 6.3 | 14.8 | 23.3 | 31.3 |
| 400 | 82.0 | 93.2 | 7.6 | 17.8 | 26.1 | 36.4 |
| 500 | 85.6 | 96.4 | 10.9 | 20.4 | 30.0 | 41.9 |
| 600 | 89.5 | 97.7 | 9.8 | 20.0 | 31.2 | 43.7 |
| 700 | 92.0 | 98.7 | 12.7 | 25.7 | 37.0 | 47.8 |
| 800 | 93.3 | 99.1 | 11.8 | 26.9 | 39.0 | 52.8 |
| 900 | 94.3 | 99.7 | 15.8 | 28.7 | 41.9 | 54.5 |
| 1000 | 94.8 | 100.0 | 13.1 | 30.7 | 42.0 | 58.4 |

[1] Based on 1000 replicates.
[2] Odds ratio of association between genotype and exposure among controls.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).

**Table 5-23. Summary of the simulation study results[1] for effects of multiplicative interaction ($OR_{GE}$) in Population 33[2]: Percentage of exponentiated coefficients [ln($OR_{GE}$)] that fall within specified ranges of "true" exponentiated coefficient (beta=0.41).**

| | | Percentage of all estimated coefficients within specified range of true coefficient | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Percentage of valid replicates[4] | OR= 1.45 to 1.52 | OR= 1.38 to 1.60 | OR= 1.32 to 1.70 | OR= 1.27 to 1.77 | OR= 1.22 to 1.84 |
| 100 | 48.6 | 31.2 | 1.1 | 2.4 | 3.5 | 4.4 |
| 200 | 72.1 | 62.2 | 1.9 | 5.1 | 7.0 | 9.2 |
| 300 | 80.8 | 84.2 | 1.4 | 5.4 | 8.4 | 11.7 |
| 400 | 82.0 | 91.7 | 2.3 | 6.9 | 10.2 | 14.3 |
| 500 | 85.6 | 95.6 | 2.5 | 9.9 | 13.5 | 17.4 |
| 600 | 89.5 | 97.9 | 3.0 | 8.7 | 13.0 | 16.9 |
| 700 | 92.0 | 98.9 | 3.4 | 10.3 | 15.2 | 19.5 |
| 800 | 93.3 | 99.2 | 2.8 | 9.9 | 16.3 | 21.4 |
| 900 | 94.3 | 99.7 | 3.6 | 13.6 | 19.6 | 23.9 |
| 1000 | 94.8 | 100.0 | 2.8 | 11.6 | 17.1 | 23.9 |

[1] Based on 1000 replicates.
[2] Odds ratio of association between genotype and exposure among controls.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic reg. model converged or –20 < beta < 20).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-24. Summary of the simulation study results[1] for testing the independence assumption (COOR)[2] in Population 33[3]: Percentage of estimated OR values that fall within specified ranges of "true" COOR (COOR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 99.8 | 3.6 | 6.4 | 9.2 | 12.5 | 15.1 |
| 200 | 100.0 | 7.2 | 16.3 | 26.8 | 34.0 | 41.2 |
| 300 | 100.0 | 9.2 | 18.3 | 26.2 | 36.1 | 48.2 |
| 400 | 100.0 | 13.2 | 23.9 | 34.1 | 43.8 | 53.1 |
| 500 | 100.0 | 13.2 | 26.6 | 39.6 | 49.7 | 60.9 |
| 600 | 100.0 | 12.2 | 27.0 | 40.0 | 53.2 | 64.5 |
| 700 | 100.0 | 13.8 | 28.2 | 42.7 | 56.2 | 67.7 |
| 800 | 100.0 | 18.2 | 34.9 | 48.1 | 61.0 | 72.9 |
| 900 | 100.0 | 16.2 | 33.7 | 52.7 | 66.6 | 76.8 |
| 1000 | 100.0 | 18.5 | 37.6 | 54.1 | 67.1 | 75.2 |

[1] Based on 1000 replicates.
[2] Odds ratio of association between genotype and exposure among controls.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).

**Table 5-25. Summary of the simulation study results[1] for testing Type 2 gene-environment interaction assumption ($OR_{Ge}$)[2] in Population 33[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 100.0 | 15.2 | 26.6 | 37.7 | 49.3 | 61.1 |
| 200 | 100.0 | 19.0 | 39.7 | 55.3 | 67.6 | 78.2 |
| 300 | 100.0 | 28.0 | 50.3 | 65.8 | 78.4 | 86.1 |
| 400 | 100.0 | 27.6 | 52.9 | 71.9 | 83.7 | 91.3 |
| 500 | 100.0 | 30.9 | 59.3 | 77.2 | 88.7 | 95.3 |
| 600 | 100.0 | 32.9 | 62.3 | 82.2 | 91.8 | 96.1 |
| 700 | 100.0 | 34.7 | 64.7 | 84.9 | 92.8 | 97.3 |
| 800 | 100.0 | 41.3 | 70.8 | 86.2 | 94.6 | 97.2 |
| 900 | 100.0 | 40.8 | 72.8 | 89.8 | 96.5 | 98.7 |
| 1000 | 100.0 | 43.5 | 74.8 | 92.3 | 97.7 | 98.9 |

[1] Based on 1000 replicates.
[2] Odds ratio of disease, among unexposed, for susceptibles versus non-susceptibles.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-26. Summary of the simulation study results[1] for effects of exposure alone $(OR_{Eg})^2$ in Population 33[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=2).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Percentage of valid replicates[4] | OR= 1.8 to 2.2 | OR= 1.6 to 2.4 | OR= 1.4 to 2.6 | OR= 1.2 to 2.8 | OR= 1.0 to 3.0 |
| 100 | 100.0 | 9.3 | 18.7 | 30.4 | 47.4 | 64.7 |
| 200 | 100.0 | 9.3 | 21.5 | 37.2 | 54.6 | 74.4 |
| 300 | 100.0 | 6.3 | 17.7 | 31.7 | 54.6 | 77.3 |
| 400 | 100.0 | 5.0 | 14.6 | 30.6 | 56.2 | 82.7 |
| 500 | 100.0 | 4.3 | 11.7 | 29.4 | 57.9 | 85.4 |
| 600 | 100.0 | 2.4 | 10.0 | 28.8 | 59.7 | 89.9 |
| 700 | 100.0 | 1.8 | 9.3 | 28.7 | 60.5 | 90.5 |
| 800 | 100.0 | 1.1 | 6.8 | 23.3 | 59.6 | 93.8 |
| 900 | 100.0 | 1.5 | 5.8 | 21.9 | 61.3 | 94.5 |
| 1000 | 100.0 | 0.6 | 5.0 | 21.9 | 62.5 | 94.3 |

[1] Based on 1000 replicates.

[2] Odds ratio of disease, among non-susceptibles, for exposed versus unexposed.

[3] See Table 5-1 for population parameters.

[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-27. Relative contribution of design parameters[1], number of subjects, study type, and independence[2] and Type2 assumptions[2] (and all 1st order interactions) to the accuracy[3] of Type 2 gene-environment interaction estimates.**

| Source of Error | Type III Sum of Squares | F Value | Pr > F |
|---|---|---|---|
| Prevalence of gene | 16987 | 25.9 | <.0001 |
| Prevalence of exposure | 17311 | 26.4 | <.0001 |
| Prevalence of disease | 17137 | 26.1 | <.0001 |
| $OR_{GE}$ | 9482 | 14.5 | 0.0001 |
| SIM or COR | 1460722 | 2226.2 | <.0001 |
| COOR | 101 | 0.2 | 0.6945 |
| $OR_{Ge}$ | 20620 | 31.4 | <.0001 |
| Number of Subjects | 15225 | 2.6 | 0.0059 |
| Study Type | 169975 | 259.0 | <.0001 |
| COOR x $OR_{Ge}$ | 2499 | 3.8 | 0.0511 |
| COOR x $OR_{GE}$ | 19113 | 29.1 | <.0001 |
| COOR x PrevD | 41 | 0.1 | 0.8028 |
| COOR x PrevE | 1280 | 2.0 | 0.1627 |
| COOR x PrevG | 15 | 0.0 | 0.8817 |
| COOR x SIM or COR | 106986 | 163.0 | <.0001 |
| COOR x Size | 990 | 0.2 | 0.9971 |
| COOR x Study Type | 151802 | 231.3 | <.0001 |
| $OR_{Ge}$ x $OR_{GE}$ | 7741 | 11.8 | 0.0006 |
| $OR_{Ge}$ x PrevD | 108 | 0.2 | 0.6845 |
| $OR_{Ge}$ x PrevE | 5002 | 7.6 | 0.0058 |
| $OR_{Ge}$ x PrevG | 2190 | 3.3 | 0.0679 |
| $OR_{Ge}$ x SIM or COR | 437 | 0.7 | 0.4148 |
| $OR_{Ge}$ x Size | 13035 | 2.2 | 0.0191 |
| $OR_{Ge}$ x Study Type | 192859 | 293.9 | <.0001 |
| $OR_{GE}$ x SIM or COR | 36618 | 55.8 | <.0001 |
| $OR_{GE}$ x Size | 3128 | 0.5 | 0.8539 |
| $OR_{GE}$ x Study Type | 34988 | 53.3 | <.0001 |
| PrevD x $OR_{GE}$ | 474 | 0.7 | 0.3956 |
| PrevD x SIM or COR | 11525 | 17.6 | <.0001 |
| PrevD x Size | 1914 | 0.3 | 0.9674 |
| PrevD x Study Type | 1029 | 1.6 | 0.2105 |
| PrevE x $OR_{GE}$ | 83191 | 126.8 | <.0001 |
| PrevE x PrevD | 280 | 0.4 | 0.5136 |
| PrevE x SIM or COR | 10013 | 15.3 | 0.0001 |
| PrevE x Size | 10860 | 1.8 | 0.0568 |
| PrevE x Study Type | 30660 | 46.7 | <.0001 |
| PrevG x $OR_{GE}$ | 4 | 0.0 | 0.9409 |
| PrevG x PrevD | 140 | 0.2 | 0.6440 |
| PrevG x PrevE | 33327 | 50.8 | <.0001 |
| PrevG x SIM or COR | 26585 | 40.5 | <.0001 |
| PrevG x Size | 63490 | 10.8 | <.0001 |
| PrevG x Study Type | 67298 | 102.6 | <.0001 |
| SIM or COR x Size | 26955 | 4.6 | <.0001 |
| SIM or COR x Study Type | 22684 | 35.1 | <.0001 |
| Size x Study Type | 47821 | 8.1 | <.0001 |

[1] Prevalence of gene, prevalence of exposure, prevalence of disease, $OR_{GE}$ (odds ratio for disease among exposed susceptibles versus unexposed non-susceptibles), SIM and COR (magnitude of multiplicative interaction), COOR (odds ratio of association between genotype and exposure among controls), $OR_{Ge}$ (odds ratio of disease among non-exposed susceptibles).

[2] COOR within ±50% of 1 (0.5-1.5); $OR_{Ge}$ within ±50% of 1 (0.5-1.5).

[3] Estimate within ±30% of "true" OR (OR=3).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-28. Relative contribution of design parameters[1], number of subjects, and the independence[2] and Type2 assumptions[2] (and 1st order interactions) to the accuracy[3] of case-only estimates (COR) of Type 2 gene-environment interaction estimates.**

| Source of Error | Type III Sum of Squares | F Value | Pr > F |
|---|---|---|---|
| PrevG | 25764 | 43.5 | <.0001 |
| PrevE | 11774 | 19.9 | <.0001 |
| PrevD | 14836 | 25.0 | <.0001 |
| $OR_{GE}$ | 13464 | 22.7 | <.0001 |
| COR | 1041373 | 1757.7 | <.0001 |
| COOR | 136 | 0.2 | 0.6322 |
| $OR_{Ge}$ | 10550 | 17.8 | <.0001 |
| Size | 22717 | 4.3 | <.0001 |
| PrevG x PrevE | 13373 | 22.6 | <.0001 |
| PrevG x PrevD | 19 | 0.0 | 0.8568 |
| PrevG x $OR_{GE}$ | 2357 | 4.0 | 0.0463 |
| PrevG x COR | 40170 | 67.8 | <.0001 |
| PrevG x Size | 60459 | 11.3 | <.0001 |
| PrevG x COOR | 12434 | 21.0 | <.0001 |
| PrevG x $OR_{Ge}$ | 823 | 1.4 | 0.2388 |
| PrevE x PrevD | 546 | 0.9 | 0.3371 |
| PrevE x $OR_{GE}$ | 76629 | 129.3 | <.0001 |
| PrevE x COR | 1623 | 2.7 | 0.0982 |
| PrevE x Size | 4580 | 0.9 | 0.5617 |
| PrevE x COOR | 2511 | 4.2 | 0.0397 |
| PrevE x $OR_{Ge}$ | 11794 | 19.9 | <.0001 |
| PrevD x $OR_{Ge}$ | 216 | 0.4 | 0.5462 |
| PrevD x COR | 5935 | 10.0 | 0.0016 |
| PrevD x Size | 1874 | 0.4 | 0.9573 |
| PrevD x COOR | 1840 | 3.1 | 0.0783 |
| PrevD x $OR_{Ge}$ | 1381 | 2.3 | 0.1271 |
| $OR_{GE}$ x COR | 8703 | 14.7 | 0.0001 |
| $OR_{GE}$ x Size | 4107 | 0.8 | 0.6441 |
| $OR_{GE}$ x COOR | 3374 | 5.7 | 0.0172 |
| $OR_{GE}$ x $OR_{Ge}$ | 6059 | 10.2 | 0.0014 |
| COR x Size | 27790 | 5.2 | <.0001 |
| COR x COOR | 24654 | 41.6 | <.0001 |
| COR x $OR_{Ge}$ | 1913 | 3.2 | 0.0726 |
| Size x COOR | 7014 | 1.3 | 0.224 |
| Size x $OR_{Ge}$ | 16967 | 3.2 | 0.0008 |
| COOR x $OR_{Ge}$ | 22 | 0.0 | 0.8486 |

[1] Prevalence of gene, prevalence of exposure, prevalence of disease, $OR_{GE}$ (odds ratio for disease among exposed susceptibles versus unexposed non-susceptibles), SIM and COR (magnitude of multiplicative interaction), COOR (odds ratio of association between genotype and exposure among controls), $OR_{Ge}$ (odds ratio of disease among non-exposed susceptibles).

[2] COOR within ±50% of 1 (0.5-1.5); $OR_{Ge}$ within ±50% of 1 (0.5-1.5).

[3] Estimate within ±30% of "true" OR (OR=3).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-29. Relative contribution of design parameters[1], number of subjects, and the independence[2] and Type2 assumptions[2] (and 1st order interactions) to the accuracy[3] of case-control estimates (SIM) of Type 2 gene-environment interaction estimates.**

| Source of Error | Type III Sum of Squares | F Value | Pr > F |
|---|---|---|---|
| PrevG | 317 | 1.3 | 0.259 |
| PrevE | 4905 | 19.7 | <.0001 |
| PrevD | 3542 | 14.2 | 0.0002 |
| $OR_{GE}$ | 657 | 2.6 | 0.1046 |
| SIM | 429511 | 1725.4 | <.0001 |
| COOR | 1988 | 8.0 | 0.0048 |
| $OR_{Ge}$ | 3372 | 13.6 | 0.0002 |
| Size | 1573 | 0.7 | 0.7073 |
| PrevG x PrevE | 16804 | 67.5 | <.0001 |
| PrevG x PrevD | 59 | 0.2 | 0.6265 |
| PrevG x $OR_{GE}$ | 2049 | 8.2 | 0.0042 |
| PrevG x SIM | 514 | 2.1 | 0.151 |
| PrevG x Size | 30989 | 13.8 | <.0001 |
| PrevG x COOR | 3219 | 12.9 | 0.0003 |
| PrevG x $OR_{Ge}$ | 232 | 0.9 | 0.3348 |
| PrevE x PrevD | 265 | 1.1 | 0.302 |
| PrevE x $OR_{GE}$ | 11380 | 45.7 | <.0001 |
| PrevE x SIM | 13224 | 53.1 | <.0001 |
| PrevE x Size | 6519 | 2.9 | 0.0021 |
| PrevE x COOR | 1254 | 5.0 | 0.025 |
| PrevE x $OR_{Ge}$ | 6 | 0.0 | 0.8801 |
| PrevD x $OR_{Ge}$ | 153 | 0.6 | 0.4339 |
| PrevD x SIM | 1142 | 4.6 | 0.0324 |
| PrevD x Size | 1677 | 0.8 | 0.6644 |
| PrevD x COOR | 1303 | 5.2 | 0.0223 |
| PrevD x $OR_{Ge}$ | 269 | 1.1 | 0.2992 |
| $OR_{GE}$ x SIM | 9861 | 39.6 | <.0001 |
| $OR_{GE}$ x Size | 7093 | 3.2 | 0.0009 |
| $OR_{GE}$ x COOR | 45693 | 183.6 | <.0001 |
| $OR_{GE}$ x $OR_{Ge}$ | 420 | 1.7 | 0.1941 |
| SIM x Size | 6839 | 3.1 | 0.0013 |
| SIM x COOR | 201453 | 809.3 | <.0001 |
| SIM x $OR_{Ge}$ | 2418 | 9.7 | 0.0019 |
| Size x COOR | 4145 | 1.9 | 0.0557 |
| Size x $OR_{Ge}$ | 2201 | 1.0 | 0.4525 |
| COOR x $OR_{Ge}$ | 15 | 0.1 | 0.806 |

[1] Prevalence of gene, prevalence of exposure, prevalence of disease, $OR_{GE}$ (odds ratio for disease among exposed susceptibles versus unexposed non-susceptibles), SIM and COR (magnitude of multiplicative interaction), COOR (odds ratio of association between genotype and exposure among controls), $OR_{Ge}$ (odds ratio of disease among non-exposed susceptibles).

[2] COOR within ±50% of 1 (0.5-1.5); $OR_{Ge}$ within ±50% of 1 (0.5-1.5).

[3] Estimate within ±30% of "true" OR (OR=3).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-30. Relative contribution of design parameters[1], number of subjects, and the independence[2] and Type2 assumptions[2] (and all 1st order interactions) to the accuracy[3] of case-control estimates (COR) of Type 2 gene-environment interaction estimates.**

| Source of Error | Type III Sum of Squares | F Value | Pr > F |
|---|---|---|---|
| | Number of Subjects = 100 | | |
| PrevG | 693 | 7.4 | 0.008 |
| PrevE | 1627 | 17.5 | <.0001 |
| PrevD | 603 | 6.5 | 0.0131 |
| ORGE | 931 | 10.0 | 0.0023 |
| COR | 13678 | 146.8 | <.0001 |
| COOR | 52 | 0.6 | 0.4564 |
| $OR_{Ge}$ | 50 | 0.5 | 0.466 |
| PrevG x PrevE | 5576 | 59.8 | <.0001 |
| PrevG x PrevD | 165 | 1.8 | 0.188 |
| PrevG x $OR_{GE}$ | 693 | 7.4 | 0.008 |
| PrevG x COR | 784 | 8.4 | 0.0049 |
| PrevG x COOR | 227 | 2.4 | 0.123 |
| PrevG x $OR_{Ge}$ | 1392 | 14.9 | 0.0002 |
| PrevE x PrevD | 465 | 5.0 | 0.0286 |
| PrevE x $OR_{GE}$ | 1362 | 14.6 | 0.0003 |
| PrevE x COR | 96 | 1.0 | 0.3141 |
| PrevE x COOR | 46 | 0.5 | 0.4838 |
| PrevE x $OR_{Ge}$ | 765 | 8.2 | 0.0054 |
| PrevD x $OR_{GE}$ | 25 | 0.3 | 0.6074 |
| PrevD x COR | 105 | 1.1 | 0.2911 |
| PrevD x COOR | 100 | 1.1 | 0.3032 |
| PrevD x $OR_{Ge}$ | 21 | 0.2 | 0.64 |
| $OR_{GE}$ x COR | 3050 | 32.7 | <.0001 |
| $OR_{GE}$ x COOR | 492 | 5.3 | 0.0244 |
| $OR_{GE}$ x ORGe | 2565 | 27.5 | <.0001 |
| COR x COOR | 5 | 0.1 | 0.822 |
| COR x $OR_{Ge}$ | 224 | 2.4 | 0.125 |
| COOR x $OR_{Ge}$ | 37 | 0.4 | 0.5305 |

[1] Prevalence of gene, prevalence of exposure, prevalence of disease, $OR_{GE}$ (odds ratio for disease among exposed susceptibles versus unexposed non-susceptibles), SIM and COR (magnitude of multiplicative interaction), COOR (odds ratio of association between genotype and exposure among controls), $OR_{Ge}$ (odds ratio of disease among non-exposed susceptibles).
[2] COOR within ±50% of 1 (0.5-1.5); $OR_{Ge}$ within ±50% of 1 (0.5-1.5).
[3] Estimate within ±30% of "true" OR (OR=3).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-30 (continued). Relative contribution of design parameters[1], number of subjects, and the independence[2] and Type2 assumptions[2] (and all 1st order interactions) to the accuracy[3] of case-control estimates (COR) of Type 2 gene-environment interaction estimates.**

| Source of Error | Type III Sum of Squares | F Value | Pr > F |
|---|---|---|---|
| | Number of Subjects = 500 | | |
| PrevG | 173 | 0.4 | 0.547 |
| PrevE | 471 | 1.0 | 0.3214 |
| PrevD | 584 | 1.2 | 0.2698 |
| ORGE | 890 | 1.9 | 0.1739 |
| COR | 85582 | 180.3 | <.0001 |
| COOR | 17 | 0.0 | 0.8511 |
| $OR_{Ge}$ | 1881 | 4.0 | 0.0491 |
| PrevG x PrevE | 1478 | 3.1 | 0.0806 |
| PrevG x PrevD | 0 | 0.0 | 0.9758 |
| PrevG x $OR_{GE}$ | 703 | 1.5 | 0.2263 |
| PrevG x COR | 2397 | 5.1 | 0.0267 |
| PrevG x COOR | 2938 | 6.2 | 0.0144 |
| PrevG x $OR_{Ge}$ | 3 | 0.0 | 0.9409 |
| PrevE x PrevD | 3 | 0.0 | 0.9333 |
| PrevE x $OR_{GE}$ | 11016 | 23.2 | <.0001 |
| PrevE x COR | 2279 | 4.8 | 0.0307 |
| PrevE x COOR | 1323 | 2.8 | 0.098 |
| PrevE x $OR_{Ge}$ | 1365 | 2.9 | 0.093 |
| PrevD x $OR_{GE}$ | 249 | 0.5 | 0.4706 |
| PrevD x COR | 456 | 1.0 | 0.329 |
| PrevD x COOR | 406 | 0.9 | 0.3574 |
| PrevD x $OR_{Ge}$ | 218 | 0.5 | 0.4998 |
| $OR_{GE}$ x COR | 2222 | 4.7 | 0.0328 |
| $OR_{GE}$ x COOR | 2069 | 4.4 | 0.0392 |
| $OR_{GE}$ x ORGe | 387 | 0.8 | 0.3687 |
| COR x COOR | 1157 | 2.4 | 0.1215 |
| COR x $OR_{Ge}$ | 39 | 0.1 | 0.7748 |
| COOR x $OR_{Ge}$ | 98 | 0.2 | 0.6505 |

[1] Prevalence of gene, prevalence of exposure, prevalence of disease, $OR_{GE}$ (odds ratio for disease among exposed susceptibles versus unexposed non-susceptibles), SIM and COR (magnitude of multiplicative interaction), COOR (odds ratio of association between genotype and exposure among controls), $OR_{Ge}$ (odds ratio of disease among non-exposed susceptibles).
[2] COOR within ±50% of 1 (0.5-1.5); $OR_{Ge}$ within ±50% of 1 (0.5-1.5).
[3] Estimate within ±30% of "true" OR (OR=3).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-30 (continued). Relative contribution of design parameters[1], number of subjects, and the independence[2] and Type2 assumptions[2] (and all 1st order interactions) to the accuracy[3] of case-control estimates (COR) of Type 2 gene-environment interaction estimates.**

| Source of Error | Type III Sum of Squares | F Value | Pr > F |
|---|---|---|---|
| | Number of Subjects = 1000 | | |
| PrevG | 8 | 0.0 | 0.9243 |
| PrevE | 115 | 0.1 | 0.7112 |
| PrevD | 187 | 0.2 | 0.6372 |
| ORGE | 121 | 0.1 | 0.7044 |
| COR | 16188 | 19.4 | <.0001 |
| COOR | 1001 | 1.2 | 0.2763 |
| $OR_{Ge}$ | 1082 | 1.3 | 0.2577 |
| PrevG x PrevE | 804 | 1.0 | 0.3289 |
| PrevG x PrevD | 682 | 0.8 | 0.3682 |
| PrevG x $OR_{GE}$ | 133 | 0.2 | 0.6904 |
| PrevG x COR | 22454 | 26.9 | <.0001 |
| PrevG x COOR | 500 | 0.6 | 0.4408 |
| PrevG x $OR_{Ge}$ | 6 | 0.0 | 0.9309 |
| PrevE x PrevD | 2 | 0.0 | 0.964 |
| PrevE x $OR_{GE}$ | 8917 | 10.7 | 0.0015 |
| PrevE x COR | 151 | 0.2 | 0.6719 |
| PrevE x COOR | 189 | 0.2 | 0.6352 |
| PrevE x $OR_{Ge}$ | 98 | 0.1 | 0.7328 |
| PrevD x $OR_{GE}$ | 159 | 0.2 | 0.6638 |
| PrevD x COR | 4861 | 5.8 | 0.0178 |
| PrevD x COOR | 121 | 0.2 | 0.7038 |
| PrevD x $OR_{Ge}$ | 13 | 0.0 | 0.9 |
| $OR_{GE}$ x COR | 13 | 0.0 | 0.9028 |
| $OR_{GE}$ x COOR | 142 | 0.2 | 0.6812 |
| $OR_{GE}$ x ORGe | 112 | 0.1 | 0.7154 |
| COR x COOR | 1669 | 2.0 | 0.1607 |
| COR x $OR_{Ge}$ | 80 | 0.1 | 0.7582 |
| COOR x $OR_{Ge}$ | 888 | 1.1 | 0.3048 |

[1] Prevalence of gene, prevalence of exposure, prevalence of disease, $OR_{GE}$ (odds ratio for disease among exposed susceptibles versus unexposed non-susceptibles), SIM and COR (magnitude of multiplicative interaction), COOR (odds ratio of association between genotype and exposure among controls), $OR_{Ge}$ (odds ratio of disease among non-exposed susceptibles).
[2] COOR within ±50% of 1 (0.5-1.5); $OR_{Ge}$ within ±50% of 1 (0.5-1.5).
[3] Estimate within ±30% of "true" OR (OR=3).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-31. Relative contribution of design parameters[1], number of subjects, and the independence[2] and Type2 assumptions[2] (and all 1st order interactions) to the accuracy[3] of case-control estimates (SIM) of Type 2 gene-environment interaction estimates.**

| Source of Error | Type III Sum of Squares | F Value | Pr > F |
|---|---|---|---|
| | Number of Subjects = 100 | | |
| PrevG | 5 | 0.1 | 0.7531 |
| PrevE | 673 | 13.3 | 0.0005 |
| PrevD | 59 | 1.2 | 0.2834 |
| ORGE | 40 | 0.8 | 0.3803 |
| SIM | 5255 | 103.5 | <.0001 |
| COOR | 28 | 0.6 | 0.4619 |
| $OR_{Ge}$ | 7 | 0.1 | 0.7203 |
| PrevG x PrevE | 356 | 7.0 | 0.0099 |
| PrevG x PrevD | 82 | 1.6 | 0.2085 |
| PrevG x $OR_{GE}$ | 1 | 0.0 | 0.9084 |
| PrevG x SIM | 0 | 0.0 | 0.933 |
| PrevG x COOR | 133 | 2.6 | 0.1097 |
| PrevG x $OR_{Ge}$ | 187 | 3.7 | 0.0587 |
| PrevE x PrevD | 13 | 0.3 | 0.6134 |
| PrevE x $OR_{GE}$ | 79 | 1.6 | 0.2148 |
| PrevE x SIMorCOR | 359 | 7.1 | 0.0096 |
| PrevE x COOR | 219 | 4.3 | 0.0412 |
| PrevE x $OR_{Ge}$ | 89 | 1.8 | 0.1898 |
| PrevD x $OR_{GE}$ | 94 | 1.9 | 0.1766 |
| PrevD x SIM | 1 | 0.0 | 0.8971 |
| PrevD x COOR | 56 | 1.1 | 0.2963 |
| PrevD x $OR_{Ge}$ | 7 | 0.1 | 0.7161 |
| $OR_{GE}$ x SIM | 104 | 2.1 | 0.1565 |
| $OR_{GE}$ x COOR | 692 | 13.6 | 0.0004 |
| $OR_{GE}$ x ORGe | 25 | 0.5 | 0.4825 |
| SIM x COOR | 4285 | 84.4 | <.0001 |
| SIM x $OR_{Ge}$ | 29 | 0.6 | 0.4533 |
| COOR x $OR_{Ge}$ | 179 | 3.5 | 0.0645 |

[1] Prevalence of gene, prevalence of exposure, prevalence of disease, $OR_{GE}$ (odds ratio for disease among exposed susceptibles versus unexposed non-susceptibles), SIM and COR (magnitude of multiplicative interaction), COOR (odds ratio of association between genotype and exposure among controls), $OR_{Ge}$ (odds ratio of disease among non-exposed susceptibles).
[2] COOR within ±50% of 1 (0.5-1.5); $OR_{Ge}$ within ±50% of 1 (0.5-1.5).
[3] Estimate within ±30% of "true" OR (OR=3).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-31 (continued). Relative contribution of design parameters[1], number of subjects, and the independence[2] and Type2 assumptions[2] (and all 1st order interactions) to the accuracy[3] of case-control estimates (SIM) of Type 2 gene-environment interaction estimates.**

| Source of Error | Type III Sum of Squares | F Value | Pr > F |
|---|---|---|---|
| | Number of Subjects = 500 | | |
| PrevG | 30 | 0.2 | 0.6964 |
| PrevE | 19 | 0.1 | 0.7567 |
| PrevD | 41 | 0.2 | 0.645 |
| ORGE | 265 | 1.4 | 0.2445 |
| SIM | 30585 | 158.4 | <.0001 |
| COOR | 1061 | 5.5 | 0.021 |
| $OR_{Ge}$ | 247 | 1.3 | 0.2603 |
| PrevG x PrevE | 4031 | 20.9 | <.0001 |
| PrevG x PrevD | 80 | 0.4 | 0.521 |
| PrevG x $OR_{GE}$ | 479 | 2.5 | 0.1182 |
| PrevG x SIM | 521 | 2.7 | 0.1035 |
| PrevG x COOR | 484 | 2.5 | 0.1164 |
| PrevG x $OR_{Ge}$ | 30 | 0.2 | 0.6951 |
| PrevE x PrevD | 16 | 0.1 | 0.7718 |
| PrevE x $OR_{GE}$ | 1534 | 8.0 | 0.0058 |
| PrevE x SIMorCOR | 3462 | 17.9 | <.0001 |
| PrevE x COOR | 362 | 1.9 | 0.1736 |
| PrevE x $OR_{Ge}$ | 3 | 0.0 | 0.8939 |
| PrevD x $OR_{GE}$ | 56 | 0.3 | 0.5897 |
| PrevD x SIM | 109 | 0.6 | 0.4546 |
| PrevD x COOR | 2 | 0.0 | 0.9169 |
| PrevD x $OR_{Ge}$ | 19 | 0.1 | 0.7557 |
| $OR_{GE}$ x SIM | 1441 | 7.5 | 0.0074 |
| $OR_{GE}$ x COOR | 8190 | 42.4 | <.0001 |
| $OR_{GE}$ x ORGe | 84 | 0.4 | 0.5098 |
| SIM x COOR | 28672 | 148.5 | <.0001 |
| SIM x $OR_{Ge}$ | 442 | 2.3 | 0.1335 |
| COOR x $OR_{Ge}$ | 438 | 2.3 | 0.135 |

[1] Prevalence of gene, prevalence of exposure, prevalence of disease, $OR_{GE}$ (odds ratio for disease among exposed susceptibles versus unexposed non-susceptibles), SIM and COR (magnitude of multiplicative interaction), COOR (odds ratio of association between genotype and exposure among controls), $OR_{Ge}$ (odds ratio of disease among non-exposed susceptibles).

[2] COOR within ±50% of 1 (0.5-1.5); $OR_{Ge}$ within ±50% of 1 (0.5-1.5).

[3] Estimate within ±30% of "true" OR (OR=3).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-31 (continued). Relative contribution of design parameters[1], number of subjects, and the independence[2] and Type2 assumptions[2] (and all 1st order interactions) to the accuracy[3] of case-control estimates (SIM) of Type 2 gene-environment interaction estimates.**

| Source of Error | Type III Sum of Squares | F Value | Pr > F |
|---|---|---|---|
| | Number of Subjects = 1000 | | |
| PrevG | 3 | 0.0 | 0.9377 |
| PrevE | 37 | 0.1 | 0.7703 |
| PrevD | 31 | 0.1 | 0.7888 |
| ORGE | 7 | 0.0 | 0.8961 |
| SIM | 8649 | 20.0 | <.0001 |
| COOR | 4613 | 10.7 | 0.0015 |
| $OR_{Ge}$ | 2126 | 4.9 | 0.0291 |
| PrevG x PrevE | 346 | 0.8 | 0.3734 |
| PrevG x PrevD | 106 | 0.2 | 0.6222 |
| PrevG x $OR_{GE}$ | 598 | 1.4 | 0.2427 |
| PrevG x SIM | 52 | 0.1 | 0.729 |
| PrevG x COOR | 4809 | 11.1 | 0.0012 |
| PrevG x $OR_{Ge}$ | 1 | 0.0 | 0.967 |
| PrevE x PrevD | 27 | 0.1 | 0.8039 |
| PrevE x $OR_{GE}$ | 1761 | 4.1 | 0.0465 |
| PrevE x SIMorCOR | 2 | 0.0 | 0.9424 |
| PrevE x COOR | 2158 | 5.0 | 0.0279 |
| PrevE x $OR_{Ge}$ | 15 | 0.0 | 0.8522 |
| PrevD x $OR_{GE}$ | 85 | 0.2 | 0.6591 |
| PrevD x SIM | 541 | 1.3 | 0.266 |
| PrevD x COOR | 85 | 0.2 | 0.6586 |
| PrevD x $OR_{Ge}$ | 0 | 0.0 | 0.9879 |
| $OR_{GE}$ x SIM | 683 | 1.6 | 0.2118 |
| $OR_{GE}$ x COOR | 3764 | 8.7 | 0.004 |
| $OR_{GE}$ x ORGe | 31 | 0.1 | 0.7904 |
| SIM x COOR | 21892 | 50.7 | <.0001 |
| SIM x $OR_{Ge}$ | 207 | 0.5 | 0.4904 |
| COOR x $OR_{Ge}$ | 3438 | 8.0 | 0.0059 |

[1] Prevalence of gene, prevalence of exposure, prevalence of disease, $OR_{GE}$ (odds ratio for disease among exposed susceptibles versus unexposed non-susceptibles), SIM and COR (magnitude of multiplicative interaction), COOR (odds ratio of association between genotype and exposure among controls), $OR_{Ge}$ (odds ratio of disease among non-exposed susceptibles).

[2] COOR within ±50% of 1 (0.5-1.5); $OR_{Ge}$ within ±50% of 1 (0.5-1.5).

[3] Estimate within ±30% of "true" OR (OR=3).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-32. Summary of the simulation study results[1] for the Typical population[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" COR[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 100 | | | |
| No | No | No | 256 | 39.1 | 66.7 |
| No | No | Yes | 310 | 47.4 | 64.3 |
| No | Yes | No | 28 | 4.3 | 70.0 |
| No | Yes | Yes | 60 | 9.2 | 63.8 |
| Yes | No | No | 128 | 37.0 | 33.3 |
| Yes | No | Yes | 172 | 49.7 | 35.7 |
| Yes | Yes | No | 12 | 3.5 | 30.0 |
| Yes | Yes | Yes | 34 | 9.8 | 36.2 |
| | | Number of Subjects = 200 | | | |
| No | No | No | 85 | 16.2 | 54.8 |
| No | No | Yes | 223 | 42.5 | 51.7 |
| No | Yes | No | 47 | 9.0 | 62.7 |
| No | Yes | Yes | 170 | 32.4 | 50.1 |
| Yes | No | No | 70 | 14.7 | 45.2 |
| Yes | No | Yes | 208 | 43.8 | 48.3 |
| Yes | Yes | No | 28 | 5.9 | 37.3 |
| Yes | Yes | Yes | 169 | 35.6 | 49.9 |
| | | Number of Subjects = 300 | | | |
| No | No | No | 46 | 10.6 | 52.9 |
| No | No | Yes | 183 | 42.4 | 41.8 |
| No | Yes | No | 33 | 7.6 | 55.0 |
| No | Yes | Yes | 170 | 39.4 | 41.0 |
| Yes | No | No | 41 | 7.2 | 47.1 |
| Yes | No | Yes | 255 | 44.9 | 58.2 |
| Yes | Yes | No | 27 | 4.8 | 45.0 |
| Yes | Yes | Yes | 245 | 43.1 | 59.0 |
| | | Number of Subjects = 400 | | | |
| No | No | No | 25 | 6.3 | 54.3 |
| No | No | Yes | 165 | 41.8 | 38.6 |
| No | Yes | No | 29 | 7.3 | 51.8 |
| No | Yes | Yes | 176 | 44.6 | 37.4 |
| Yes | No | No | 21 | 3.5 | 45.7 |
| Yes | No | Yes | 263 | 43.5 | 61.4 |
| Yes | Yes | No | 27 | 4.5 | 48.2 |
| Yes | Yes | Yes | 294 | 48.6 | 62.6 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-only study) = 3.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Among unexposed, OR for disease between susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-32 (continued). Summary of the simulation study results[1] for the Typical population[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" COR[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 500 | | | |
| No | No | No | 11 | 3.6 | 40.7 |
| No | No | Yes | 114 | 37.1 | 31.0 |
| No | Yes | No | 10 | 3.3 | 32.3 |
| No | Yes | Yes | 172 | 56.0 | 30.0 |
| Yes | No | No | 16 | 2.3 | 59.3 |
| Yes | No | Yes | 254 | 36.7 | 69.0 |
| Yes | Yes | No | 21 | 3.0 | 67.7 |
| Yes | Yes | Yes | 402 | 58.0 | 70.0 |
| | | Number of Subjects = 600 | | | |
| No | No | No | 9 | 3.8 | 39.1 |
| No | No | Yes | 74 | 31.6 | 21.6 |
| No | Yes | No | 9 | 3.8 | 36.0 |
| No | Yes | Yes | 142 | 60.7 | 23.3 |
| Yes | No | No | 14 | 1.8 | 60.9 |
| Yes | No | Yes | 268 | 35.0 | 78.4 |
| Yes | Yes | No | 16 | 2.1 | 64.0 |
| Yes | Yes | Yes | 468 | 61.1 | 76.7 |
| | | Number of Subjects = 700 | | | |
| No | No | No | 5 | 2.5 | 35.7 |
| No | No | Yes | 65 | 32.2 | 20.9 |
| No | Yes | No | 6 | 3.0 | 35.3 |
| No | Yes | Yes | 126 | 62.4 | 19.1 |
| Yes | No | No | 9 | 1.1 | 64.3 |
| Yes | No | Yes | 246 | 30.8 | 79.1 |
| Yes | Yes | No | 11 | 1.4 | 64.7 |
| Yes | Yes | Yes | 532 | 66.7 | 80.9 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-only study) = 3.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Among unexposed, OR for disease between susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-33. Summary of the simulation study results[1] for the Typical population[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" SIM[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 100 | | | |
| No | No | No | 349 | 38.8 | 90.9 |
| No | No | Yes | 457 | 50.8 | 94.8 |
| No | Yes | No | 26 | 2.9 | 65.0 |
| No | Yes | Yes | 68 | 7.6 | 72.3 |
| Yes | No | No | 35 | 35.0 | 9.1 |
| Yes | No | Yes | 25 | 25.0 | 5.2 |
| Yes | Yes | No | 14 | 14.0 | 35.0 |
| Yes | Yes | Yes | 26 | 26.0 | 27.7 |
| | | Number of Subjects = 200 | | | |
| No | No | No | 125 | 16.4 | 80.6 |
| No | No | Yes | 395 | 52.0 | 91.6 |
| No | Yes | No | 50 | 6.6 | 66.7 |
| No | Yes | Yes | 190 | 25.0 | 56.0 |
| Yes | No | No | 30 | 12.5 | 19.4 |
| Yes | No | Yes | 36 | 15.0 | 8.4 |
| Yes | Yes | No | 25 | 10.4 | 33.3 |
| Yes | Yes | Yes | 149 | 62.1 | 44.0 |
| | | Number of Subjects = 300 | | | |
| No | No | No | 70 | 10.0 | 80.5 |
| No | No | Yes | 386 | 55.1 | 88.1 |
| No | Yes | No | 37 | 5.3 | 61.7 |
| No | Yes | Yes | 208 | 29.7 | 50.1 |
| Yes | No | No | 17 | 5.7 | 19.5 |
| Yes | No | Yes | 52 | 17.4 | 11.9 |
| Yes | Yes | No | 23 | 7.7 | 38.3 |
| Yes | Yes | Yes | 207 | 69.2 | 49.9 |
| | | Number of Subjects = 400 | | | |
| No | No | No | 32 | 5.0 | 69.6 |
| No | No | Yes | 375 | 58.7 | 87.6 |
| No | Yes | No | 29 | 4.5 | 51.8 |
| No | Yes | Yes | 203 | 31.8 | 43.2 |
| Yes | No | No | 14 | 3.9 | 30.4 |
| Yes | No | Yes | 53 | 14.7 | 12.4 |
| Yes | Yes | No | 27 | 7.5 | 48.2 |
| Yes | Yes | Yes | 267 | 74.0 | 56.8 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-control study) = 3.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Among unexposed, OR for disease between susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-33 (continued). Summary of the simulation study results[1] for the Typical population[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" SIM[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 500 | | | |
| No | No | No | 22 | 3.6 | 81.5 |
| No | No | Yes | 327 | 53.4 | 88.9 |
| No | Yes | No | 18 | 2.9 | 58.1 |
| No | Yes | Yes | 245 | 40.0 | 42.7 |
| Yes | No | No | 5 | 1.3 | 18.5 |
| Yes | No | Yes | 41 | 10.6 | 11.1 |
| Yes | Yes | No | 13 | 3.4 | 41.9 |
| Yes | Yes | Yes | 329 | 84.8 | 57.3 |
| | | Number of Subjects = 600 | | | |
| No | No | No | 16 | 2.7 | 69.6 |
| No | No | Yes | 306 | 50.8 | 89.5 |
| No | Yes | No | 14 | 2.3 | 56.0 |
| No | Yes | Yes | 266 | 44.2 | 43.6 |
| Yes | No | No | 7 | 1.8 | 30.4 |
| Yes | No | Yes | 36 | 9.0 | 10.5 |
| Yes | Yes | No | 11 | 2.8 | 44.0 |
| Yes | Yes | Yes | 344 | 86.4 | 56.4 |
| | | Number of Subjects = 700 | | | |
| No | No | No | 10 | 1.8 | 71.4 |
| No | No | Yes | 274 | 50.0 | 88.1 |
| No | Yes | No | 9 | 1.6 | 52.9 |
| No | Yes | Yes | 255 | 46.5 | 38.8 |
| Yes | No | No | 4 | 0.9 | 28.6 |
| Yes | No | Yes | 37 | 8.2 | 11.9 |
| Yes | Yes | No | 8 | 1.8 | 47.1 |
| Yes | Yes | Yes | 403 | 89.2 | 61.2 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-control study) = 3.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Among unexposed, OR for disease between susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-34. Summary of the simulation study results[1] for Population 1[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" COR[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 400 | | | |
| No | No | No | 159 | 18.3 | 85.9 |
| No | No | Yes | 558 | 64.1 | 87.1 |
| No | Yes | No | 13 | 1.5 | 92.9 |
| No | Yes | Yes | 140 | 16.1 | 87.5 |
| Yes | No | No | 26 | 20.0 | 14.1 |
| Yes | No | Yes | 83 | 63.8 | 12.9 |
| Yes | Yes | No | 1 | 0.8 | 7.1 |
| Yes | Yes | Yes | 20 | 15.4 | 12.5 |
| | | Number of Subjects = 500 | | | |
| No | No | No | 104 | 12.4 | 80.6 |
| No | No | Yes | 497 | 59.2 | 85.7 |
| No | Yes | No | 17 | 2.0 | 89.5 |
| No | Yes | Yes | 222 | 26.4 | 81.6 |
| Yes | No | No | 25 | 15.6 | 19.4 |
| Yes | No | Yes | 83 | 51.9 | 14.3 |
| Yes | Yes | No | 2 | 1.3 | 10.5 |
| Yes | Yes | Yes | 50 | 31.3 | 18.4 |
| | | Number of Subjects = 600 | | | |
| No | No | No | 80 | 9.8 | 88.9 |
| No | No | Yes | 440 | 53.9 | 80.4 |
| No | Yes | No | 24 | 2.9 | 80.0 |
| No | Yes | Yes | 272 | 33.3 | 81.7 |
| Yes | No | No | 10 | 5.4 | 11.1 |
| Yes | No | Yes | 107 | 58.2 | 19.6 |
| Yes | Yes | No | 6 | 3.3 | 20.0 |
| Yes | Yes | Yes | 61 | 33.2 | 18.3 |
| | | Number of Subjects = 700 | | | |
| No | No | No | 51 | 6.1 | 82.3 |
| No | No | Yes | 416 | 50.1 | 82.7 |
| No | Yes | No | 21 | 2.5 | 87.5 |
| No | Yes | Yes | 342 | 41.2 | 83.2 |
| Yes | No | No | 11 | 6.5 | 17.7 |
| Yes | No | Yes | 87 | 51.2 | 17.3 |
| Yes | Yes | No | 3 | 1.8 | 12.5 |
| Yes | Yes | Yes | 69 | 40.6 | 16.8 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-only study) = 1.5.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Among unexposed, OR for disease between susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-34 (continued). Summary of the simulation study results[1] for Population 1[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" COR[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 800 | | | |
| No | No | No | 33 | 4.1 | 71.7 |
| No | No | Yes | 378 | 47.3 | 80.9 |
| No | Yes | No | 24 | 3.0 | 92.3 |
| No | Yes | Yes | 365 | 45.6 | 79.2 |
| Yes | No | No | 13 | 6.5 | 28.3 |
| Yes | No | Yes | 89 | 44.5 | 19.1 |
| Yes | Yes | No | 2 | 1.0 | 7.7 |
| Yes | Yes | Yes | 96 | 48.0 | 20.8 |
| | | Number of Subjects = 1000 | | | |
| No | No | No | 25 | 3.2 | 89.3 |
| No | No | Yes | 373 | 47.1 | 78.5 |
| No | Yes | No | 15 | 1.9 | 65.2 |
| No | Yes | Yes | 379 | 47.9 | 80.0 |
| Yes | No | No | 3 | 1.4 | 10.7 |
| Yes | No | Yes | 102 | 49.0 | 21.5 |
| Yes | Yes | No | 8 | 3.8 | 34.8 |
| Yes | Yes | Yes | 95 | 45.7 | 20.0 |
| | | Number of Subjects = 1000 | | | |
| No | No | No | 16 | 2.1 | 72.7 |
| No | No | Yes | 364 | 48.7 | 77.3 |
| No | Yes | No | 9 | 1.2 | 60.0 |
| No | Yes | Yes | 359 | 48.0 | 73.0 |
| Yes | No | No | 6 | 2.4 | 27.3 |
| Yes | No | Yes | 107 | 42.5 | 22.7 |
| Yes | Yes | No | 6 | 2.4 | 40.0 |
| Yes | Yes | Yes | 133 | 52.8 | 27.0 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-only study) = 1.5.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Among unexposed, OR for disease between susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-35. Summary of the simulation study results[1] for Population 1[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" SIM[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 400 | | | |
| No | No | No | 179 | 18.8 | 96.8 |
| No | No | Yes | 628 | 65.9 | 98.0 |
| No | Yes | No | 13 | 1.4 | 92.9 |
| No | Yes | Yes | 133 | 14.0 | 83.1 |
| Yes | No | No | 6 | 12.8 | 3.2 |
| Yes | No | Yes | 13 | 27.7 | 2.0 |
| Yes | Yes | No | 1 | 2.1 | 7.1 |
| Yes | Yes | Yes | 27 | 57.4 | 16.9 |
| | | Number of Subjects = 500 | | | |
| No | No | No | 126 | 13.4 | 97.7 |
| No | No | Yes | 565 | 60.0 | 97.4 |
| No | Yes | No | 15 | 1.6 | 78.9 |
| No | Yes | Yes | 236 | 25.1 | 86.8 |
| Yes | No | No | 3 | 5.2 | 2.3 |
| Yes | No | Yes | 15 | 25.9 | 2.6 |
| Yes | Yes | No | 4 | 6.9 | 21.1 |
| Yes | Yes | Yes | 36 | 62.1 | 13.2 |
| | | Number of Subjects = 600 | | | |
| No | No | No | 85 | 9.2 | 94.4 |
| No | No | Yes | 535 | 57.8 | 97.8 |
| No | Yes | No | 27 | 2.9 | 90.0 |
| No | Yes | Yes | 279 | 30.1 | 83.8 |
| Yes | No | No | 5 | 6.8 | 5.6 |
| Yes | No | Yes | 12 | 16.2 | 2.2 |
| Yes | Yes | No | 3 | 4.1 | 10.0 |
| Yes | Yes | Yes | 54 | 73.0 | 16.2 |
| | | Number of Subjects = 700 | | | |
| No | No | No | 59 | 6.6 | 95.2 |
| No | No | Yes | 482 | 53.7 | 95.8 |
| No | Yes | No | 19 | 2.1 | 79.2 |
| No | Yes | Yes | 338 | 37.6 | 82.2 |
| Yes | No | No | 3 | 2.9 | 4.8 |
| Yes | No | Yes | 21 | 20.6 | 4.2 |
| Yes | Yes | No | 5 | 4.9 | 20.8 |
| Yes | Yes | Yes | 73 | 71.6 | 17.8 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-control study) = 1.5.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Among unexposed, OR for disease between susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-35 (continued). Summary of the simulation study results[1] for Population 1[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" SIM[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 800 | | | |
| No | No | No | 43 | 4.8 | 93.5 |
| No | No | Yes | 444 | 49.4 | 95.1 |
| No | Yes | No | 23 | 2.6 | 88.5 |
| No | Yes | Yes | 389 | 43.3 | 84.4 |
| Yes | No | No | 3 | 3.0 | 6.5 |
| Yes | No | Yes | 23 | 22.8 | 4.9 |
| Yes | Yes | No | 3 | 3.0 | 11.5 |
| Yes | Yes | Yes | 72 | 71.3 | 15.6 |
| | | Number of Subjects = 900 | | | |
| No | No | No | 26 | 2.9 | 92.9 |
| No | No | Yes | 459 | 50.7 | 96.6 |
| No | Yes | No | 21 | 2.3 | 91.3 |
| No | Yes | Yes | 400 | 44.2 | 84.4 |
| Yes | No | No | 2 | 2.1 | 7.1 |
| Yes | No | Yes | 16 | 17.0 | 3.4 |
| Yes | Yes | No | 2 | 2.1 | 8.7 |
| Yes | Yes | Yes | 74 | 78.7 | 15.6 |
| | | Number of Subjects = 1000 | | | |
| No | No | No | 19 | 2.1 | 86.4 |
| No | No | Yes | 456 | 50.7 | 96.8 |
| No | Yes | No | 13 | 1.4 | 86.7 |
| No | Yes | Yes | 411 | 45.7 | 83.5 |
| Yes | No | No | 3 | 3.0 | 13.6 |
| Yes | No | Yes | 15 | 14.9 | 3.2 |
| Yes | Yes | No | 2 | 2.0 | 13.3 |
| Yes | Yes | Yes | 81 | 80.2 | 16.5 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-control study) = 1.5.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Among unexposed, OR for disease between susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-36. Summary of the simulation study results[1] for Population 32[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" COR[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 100 | | | |
| No | No | No | 72 | 24.6 | 35.6 |
| No | No | Yes | 83 | 28.3 | 23.0 |
| No | Yes | No | 61 | 20.8 | 43.6 |
| No | Yes | Yes | 77 | 26.3 | 25.9 |
| Yes | No | No | 130 | 18.4 | 64.4 |
| Yes | No | Yes | 278 | 39.3 | 77.0 |
| Yes | Yes | No | 79 | 11.2 | 56.4 |
| Yes | Yes | Yes | 220 | 31.1 | 74.1 |
| | | Number of Subjects = 200 | | | |
| No | No | No | 15 | 11.0 | 26.3 |
| No | No | Yes | 28 | 20.6 | 9.4 |
| No | Yes | No | 33 | 24.3 | 34.4 |
| No | Yes | Yes | 60 | 44.1 | 10.9 |
| Yes | No | No | 42 | 4.9 | 73.7 |
| Yes | No | Yes | 270 | 31.3 | 90.6 |
| Yes | Yes | No | 63 | 7.3 | 65.6 |
| Yes | Yes | Yes | 489 | 56.6 | 89.1 |
| | | Number of Subjects = 300 | | | |
| No | No | No | 6 | 8.6 | 17.1 |
| No | No | Yes | 12 | 17.1 | 4.9 |
| No | Yes | No | 11 | 15.7 | 24.4 |
| No | Yes | Yes | 41 | 58.6 | 6.1 |
| Yes | No | No | 29 | 3.1 | 82.9 |
| Yes | No | Yes | 233 | 25.1 | 95.1 |
| Yes | Yes | No | 34 | 3.7 | 75.6 |
| Yes | Yes | Yes | 634 | 68.2 | 93.9 |
| | | Number of Subjects = 400 | | | |
| No | No | No | 2 | 5.6 | 9.5 |
| No | No | Yes | 6 | 16.7 | 3.2 |
| No | Yes | No | 6 | 16.7 | 15.0 |
| No | Yes | Yes | 22 | 61.1 | 2.9 |
| Yes | No | No | 19 | 2.0 | 90.5 |
| Yes | No | Yes | 184 | 19.1 | 96.8 |
| Yes | Yes | No | 34 | 3.5 | 85.0 |
| Yes | Yes | Yes | 727 | 75.4 | 97.1 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-only study) = 6.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Odds ratios for disease among unexposed susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-36 (continued). Summary of the simulation study results[1] for Population 32[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" COR[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 500 | | | |
| No | No | No | 0 | 0.0 | 0.0 |
| No | No | Yes | 1 | 7.1 | 0.7 |
| No | Yes | No | 3 | 21.4 | 13.0 |
| No | Yes | Yes | 10 | 71.4 | 1.2 |
| Yes | No | No | 7 | 0.7 | 100.0 |
| Yes | No | Yes | 146 | 14.8 | 99.3 |
| Yes | Yes | No | 20 | 2.0 | 87.0 |
| Yes | Yes | Yes | 813 | 82.5 | 98.8 |
| | | Number of Subjects = 600 | | | |
| No | No | No | 0 | 0.0 | 0.0 |
| No | No | Yes | 2 | 22.2 | 1.8 |
| No | Yes | No | 1 | 11.1 | 7.1 |
| No | Yes | Yes | 6 | 66.7 | 0.7 |
| Yes | No | No | 10 | 1.0 | 100.0 |
| Yes | No | Yes | 111 | 11.2 | 98.2 |
| Yes | Yes | No | 13 | 1.3 | 92.9 |
| Yes | Yes | Yes | 857 | 86.5 | 99.3 |
| | | Number of Subjects = 700 | | | |
| No | No | No | 0 | 0.0 | 0.0 |
| No | No | Yes | 0 | 0.0 | 0.0 |
| No | Yes | No | 0 | 0.0 | 0.0 |
| No | Yes | Yes | 1 | 100.0 | 0.1 |
| Yes | No | No | 5 | 0.5 | 100.0 |
| Yes | No | Yes | 91 | 9.1 | 100.0 |
| Yes | Yes | No | 21 | 2.1 | 100.0 |
| Yes | Yes | Yes | 882 | 88.3 | 99.9 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR (departure from multiplicative model of interaction in a case-only study) = 6.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Odds ratios for disease among unexposed susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-37. Summary of the simulation study results[1] for Population 32[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" SIM[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 100 | | | |
| No | No | No | 136 | 24.6 | 67.3 |
| No | No | Yes | 278 | 50.3 | 77.0 |
| No | Yes | No | 51 | 9.2 | 36.4 |
| No | Yes | Yes | 88 | 15.9 | 29.6 |
| Yes | No | No | 66 | 14.8 | 32.7 |
| Yes | No | Yes | 83 | 18.6 | 23.0 |
| Yes | Yes | No | 89 | 19.9 | 63.6 |
| Yes | Yes | Yes | 209 | 46.8 | 70.4 |
| | | Number of Subjects = 200 | | | |
| No | No | No | 32 | 8.8 | 56.1 |
| No | No | Yes | 206 | 56.6 | 69.1 |
| No | Yes | No | 25 | 6.9 | 26.0 |
| No | Yes | Yes | 101 | 27.7 | 18.4 |
| Yes | No | No | 25 | 3.9 | 43.9 |
| Yes | No | Yes | 92 | 14.5 | 30.9 |
| Yes | Yes | No | 71 | 11.2 | 74.0 |
| Yes | Yes | Yes | 448 | 70.4 | 81.6 |
| | | Number of Subjects = 300 | | | |
| No | No | No | 11 | 4.3 | 31.4 |
| No | No | Yes | 153 | 59.8 | 62.4 |
| No | Yes | No | 6 | 2.3 | 13.3 |
| No | Yes | Yes | 86 | 33.6 | 12.7 |
| Yes | No | No | 24 | 3.2 | 68.6 |
| Yes | No | Yes | 92 | 12.4 | 37.6 |
| Yes | Yes | No | 39 | 5.2 | 86.7 |
| Yes | Yes | Yes | 589 | 79.2 | 87.3 |
| | | Number of Subjects = 400 | | | |
| No | No | No | 8 | 4.2 | 38.1 |
| No | No | Yes | 109 | 56.8 | 57.4 |
| No | Yes | No | 5 | 2.6 | 12.5 |
| No | Yes | Yes | 70 | 36.5 | 9.3 |
| Yes | No | No | 13 | 1.6 | 61.9 |
| Yes | No | Yes | 81 | 10.0 | 42.6 |
| Yes | Yes | No | 35 | 4.3 | 87.5 |
| Yes | Yes | Yes | 679 | 84.0 | 90.7 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] SIM (departure from multiplicative model of interaction in a case-control study) = 6.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Odds ratios for disease among unexposed susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-37 (continued). Summary of the simulation study results[1] for Population 32[2] as percentages of combinations of the following factors: number of estimates categorized by ±30% of "true" SIM[3] and number of estimates within categories of the independence[4] and Type 2 assumptions[5].**

| Estimates within ±30% of "true" risk | Independence assumption met | Type2 assumption met | Frequency of estimates per category | Row Percent[6] | Column Percent[7] |
|---|---|---|---|---|---|
| | | Number of Subjects = 500 | | | |
| No | No | No | 3 | 2.1 | 42.9 |
| No | No | Yes | 79 | 55.6 | 53.7 |
| No | Yes | No | 4 | 2.8 | 17.4 |
| No | Yes | Yes | 56 | 39.4 | 6.8 |
| Yes | No | No | 4 | 0.5 | 57.1 |
| Yes | No | Yes | 68 | 7.9 | 46.3 |
| Yes | Yes | No | 19 | 2.2 | 82.6 |
| Yes | Yes | Yes | 767 | 89.4 | 93.2 |
| | | Number of Subjects = 600 | | | |
| No | No | No | 2 | 1.8 | 20.0 |
| No | No | Yes | 54 | 49.1 | 47.8 |
| No | Yes | No | 0 | 0.0 | 0.0 |
| No | Yes | Yes | 54 | 49.1 | 6.3 |
| Yes | No | No | 8 | 0.9 | 80.0 |
| Yes | No | Yes | 59 | 6.6 | 52.2 |
| Yes | Yes | No | 14 | 1.6 | 100.0 |
| Yes | Yes | Yes | 809 | 90.9 | 93.7 |
| | | Number of Subjects = 700 | | | |
| No | No | No | 0 | 0.0 | 0.0 |
| No | No | Yes | 31 | 44.3 | 34.1 |
| No | Yes | No | 0 | 0.0 | 0.0 |
| No | Yes | Yes | 39 | 55.7 | 4.4 |
| Yes | No | No | 5 | 0.5 | 100.0 |
| Yes | No | Yes | 60 | 6.5 | 65.9 |
| Yes | Yes | No | 21 | 2.3 | 100.0 |
| Yes | Yes | Yes | 844 | 90.8 | 95.6 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] SIM (departure from multiplicative model of interaction in a case-control study) = 6.
[4] Odds ratios for association between genotype and exposure among controls is in 0.5-1.5 range.
[5] Odds ratios for disease among unexposed susceptibles vs. non-susceptibles is in 0.5-1.5 range.
[6] Percentage of estimates in "true" risk range categories that fall within assumption categories.
[7] Percent of estimates in assumption categories that fall within "true" risk range categories.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-38. Summary of the simulation study results[1] for gene-environment interaction with respect to case-only and case-control designs in the Population 34[2]: Percentage of estimated SIM or COR[3] values that fall within specified ranges of "true" SIM or COR value (SIM=3 or COR=3).**

| | | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|---|
| | | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Study Design | Percentage of valid replicates[4] | OR= 2.7 to 3.3 | OR= 2.4 to 3.6 | OR= 2.1 to 3.9 | OR= 1.8 to 4.2 | OR= 1.5 to 4.5 |
| 100 | Case-only | 95.0 | 8.6 | 17.0 | 27.9 | 35.5 | 43.3 |
| | Case-control | 59.2 | 4.1 | 8.1 | 12.3 | 17.5 | 23.4 |
| 200 | Case-only | 98.6 | 11.1 | 21.2 | 32.4 | 41.3 | 48.8 |
| | Case-control | 81.6 | 7.2 | 14.8 | 21.9 | 30.1 | 36.7 |
| 300 | Case-only | 99.8 | 10.7 | 19.3 | 30.7 | 39.2 | 48.6 |
| | Case-control | 92.1 | 9.4 | 17.6 | 28.0 | 35.1 | 45.0 |
| 400 | Case-only | 100.0 | 10.1 | 21.5 | 31.5 | 41.4 | 48.2 |
| | Case-control | 94.4 | 11.4 | 23.8 | 34.9 | 45.7 | 54.4 |
| 500 | Case-only | 100.0 | 12.4 | 22.5 | 32.4 | 42.1 | 50.4 |
| | Case-control | 95.9 | 14.1 | 26.9 | 37.4 | 48.5 | 59.4 |
| 600 | Case-only | 100.0 | 9.8 | 20.4 | 30.3 | 39.2 | 50.2 |
| | Case-control | 97.2 | 11.3 | 23.6 | 37.8 | 50.5 | 63.5 |
| 700 | Case-only | 100.0 | 8.1 | 18.4 | 29.3 | 41.5 | 52.0 |
| | Case-control | 98.1 | 14.8 | 27.4 | 43.8 | 57.2 | 67.5 |
| 800 | Case-only | 100.0 | 8.2 | 16.4 | 26.0 | 36.0 | 47.6 |
| | Case-control | 99.1 | 16.8 | 32.0 | 45.9 | 59.9 | 71.6 |
| 900 | Case-only | 100.0 | 5.5 | 13.5 | 24.5 | 36.9 | 48.6 |
| | Case-control | 99.5 | 17.6 | 34.5 | 49.3 | 63.6 | 75.4 |
| 1000 | Case-only | 100.0 | 5.5 | 12.4 | 23.0 | 36.6 | 50.9 |
| | Case-control | 99.5 | 16.8 | 35.0 | 50.9 | 65.6 | 75.9 |

[1] Based on 1000 replicates.
[2] See Table 5-1 for population parameters.
[3] COR = Departure from multiplicative model of interaction in a case-only study;
SIM = Departure from multiplicative model of interaction in a case-control study.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-39. Summary of the simulation study results[1] for gene-environment interaction with respect to case-only and case-control designs in the Population 35[2]: Percentage of estimated SIM or COR[3] values that fall within specified ranges of "true" SIM or COR value (SIM=3 or COR=3).**

| | | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|---|
| | | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample Size | Study Design | Percentage of valid replicates[4] | OR= 2.7 to 3.3 | OR= 2.4 to 3.6 | OR= 2.1 to 3.9 | OR= 1.8 to 4.2 | OR= 1.5 to 4.5 |
| 100 | Case-only | 73.5 | 2.6 | 4.9 | 7.7 | 10.3 | 12.8 |
| | Case-control | 71.6 | 6.4 | 13.1 | 17.9 | 25.8 | 31.8 |
| 200 | Case-only | 83.1 | 1.0 | 1.9 | 2.8 | 3.3 | 4.5 |
| | Case-control | 89.3 | 8.2 | 15.5 | 24.2 | 34.3 | 43.9 |
| 300 | Case-only | 88.6 | 0.1 | 0.7 | 1.2 | 1.6 | 2.4 |
| | Case-control | 94.8 | 10.8 | 22.1 | 32.8 | 42.8 | 52.2 |
| 400 | Case-only | 92.4 | 0.1 | 0.3 | 0.6 | 1.1 | 1.8 |
| | Case-control | 98.0 | 12.8 | 25.7 | 38.4 | 50.6 | 60.3 |
| 500 | Case-only | 96.6 | 0.0 | 0.0 | 0.0 | 0.1 | 0.4 |
| | Case-control | 98.5 | 13.3 | 28.4 | 45.4 | 58.7 | 68.3 |
| 600 | Case-only | 97.6 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 |
| | Case-control | 99.2 | 15.9 | 31.4 | 47.2 | 60.7 | 70.7 |
| 700 | Case-only | 98.2 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 |
| | Case-control | 99.8 | 17.5 | 34.3 | 50.8 | 63.3 | 75.0 |
| 800 | Case-only | 98.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | Case-control | 99.8 | 19.6 | 35.3 | 52.4 | 67.1 | 76.1 |
| 900 | Case-only | 98.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | Case-control | 100.0 | 19.7 | 36.0 | 54.4 | 68.1 | 80.0 |
| 1000 | Case-only | 99.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 |
| | Case-control | 100.0 | 19.6 | 40.8 | 58.0 | 71.0 | 81.0 |

[1] Based on 1000 replicates

[2] See Table 5-1 for population parameters

[3] COR = Departure from multiplicative model of interaction in a case-only study;
SIM = Departure from multiplicative model of interaction in a case-control study

[4] Percentage of replicates included (fitted logistic regression model converged or OR <15)



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Table 5-40. Summary of the simulation study results[1] for testing the independence assumption (COOR)[2] in Population 34[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 98.9 | 2.7 | 5.7 | 8.6 | 11.6 | 14.1 |
| 200 | 100.0 | 7.6 | 14.8 | 21.6 | 28.5 | 34.3 |
| 300 | 100.0 | 7.3 | 15.2 | 22.7 | 31.4 | 41.4 |
| 400 | 100.0 | 9.4 | 18.5 | 26.4 | 34.0 | 41.2 |
| 500 | 100.0 | 9.7 | 21.0 | 27.7 | 36.0 | 43.8 |
| 600 | 100.0 | 9.3 | 19.8 | 29.5 | 39.2 | 46.2 |
| 700 | 100.0 | 11.2 | 21.4 | 31.2 | 40.4 | 47.5 |
| 800 | 100.0 | 9.9 | 24.2 | 33.7 | 42.5 | 49.9 |
| 900 | 100.0 | 12.3 | 23.7 | 33.6 | 42.8 | 50.9 |
| 1000 | 100.0 | 10.4 | 21.2 | 31.2 | 42.2 | 49.9 |

[1] Based on 1000 replicates.
[2] Odds ratio of association between genotype and exposure among controls.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).

**Table 5-41. Summary of the simulation study results[1] for testing the independence assumption (COOR)[2] in Population 35[3]: Percentage of estimated OR values that fall within specified ranges of "true" OR (OR=1).**

| | | Percentage of all sample OR estimates within specified range of true OR | | | | |
|---|---|---|---|---|---|---|
| | | ±10% | ±20% | ±30% | ±40% | ±50% |
| Sample size | Percentage of valid replicates[4] | OR= 0.9 to 1.1 | OR= 0.8 to 1.2 | OR= 0.7 to 1.3 | OR= 0.6 to 1.4 | OR= 0.5 to 1.5 |
| 100 | 94.8 | 1.1 | 2.8 | 3.9 | 6.2 | 7.4 |
| 200 | 99.9 | 3.8 | 8.2 | 10.4 | 13.9 | 15.7 |
| 300 | 99.9 | 2.7 | 5.4 | 8.2 | 11.9 | 15.4 |
| 400 | 100.0 | 2.7 | 4.7 | 7.2 | 9.4 | 11.7 |
| 500 | 100.0 | 2.4 | 4.1 | 6.1 | 7.6 | 8.8 |
| 600 | 100.0 | 1.5 | 2.7 | 4.4 | 6.0 | 7.6 |
| 700 | 100.0 | 0.9 | 1.7 | 2.6 | 4.4 | 5.7 |
| 800 | 100.0 | 0.7 | 1.3 | 2.2 | 3.0 | 4.2 |
| 900 | 100.0 | 0.2 | 0.9 | 1.5 | 2.3 | 3.4 |
| 1000 | 100.0 | 0.4 | 0.7 | 1.6 | 2.8 | 3.7 |

[1] Based on 1000 replicates.
[2] Odds ratio of association between genotype and exposure among controls.
[3] See Table 5-1 for population parameters.
[4] Percentage of replicates included (fitted logistic regression model converged or OR <15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

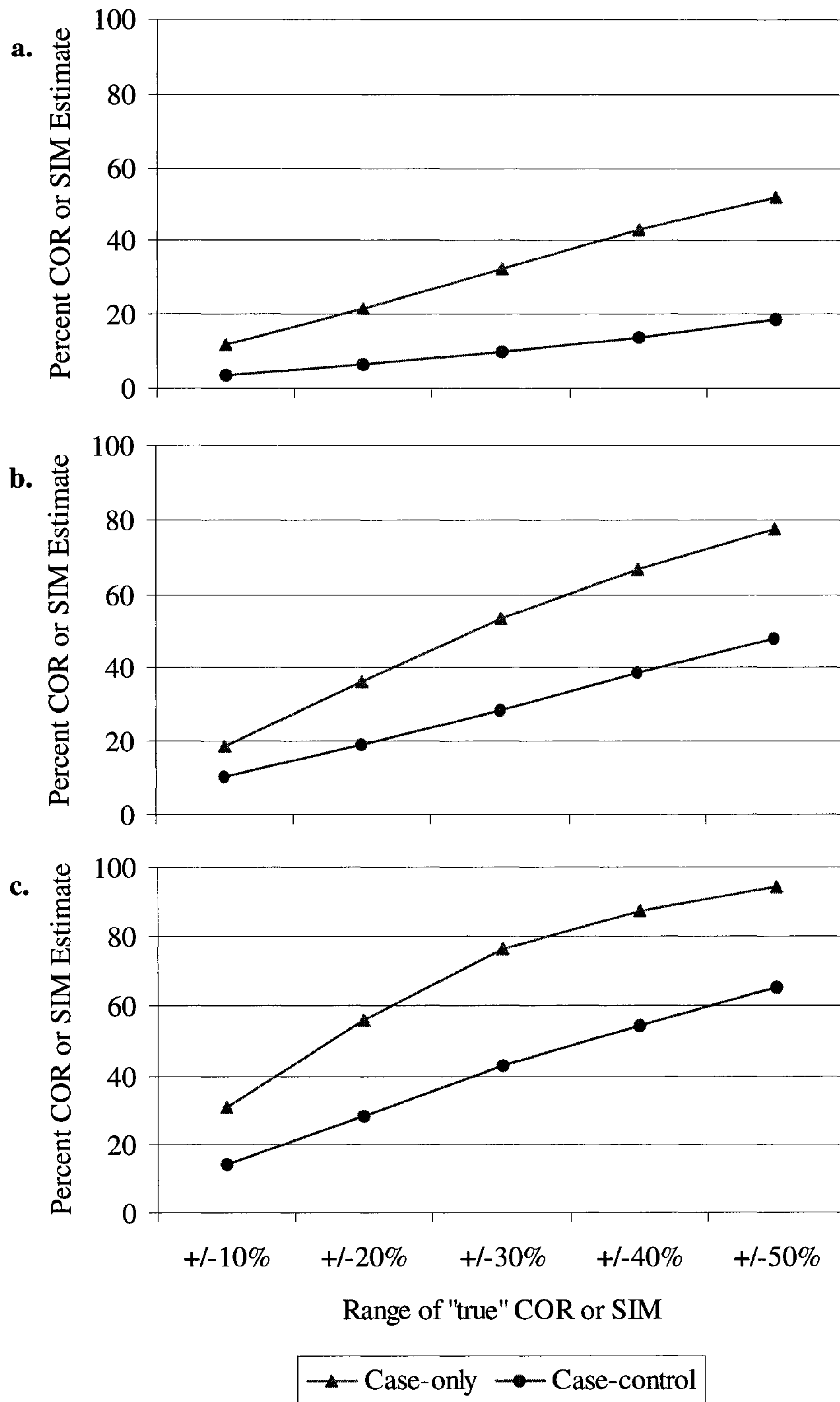



**Figure 5-1. Effect of sample size on accuracy of case-only and case-control estimates from the Typical population ( a: n=100; b: n=300; c: n=700).** Numerals with lines indicate percent of 1000 replicates included (fitted logistic model converged or OR < 15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

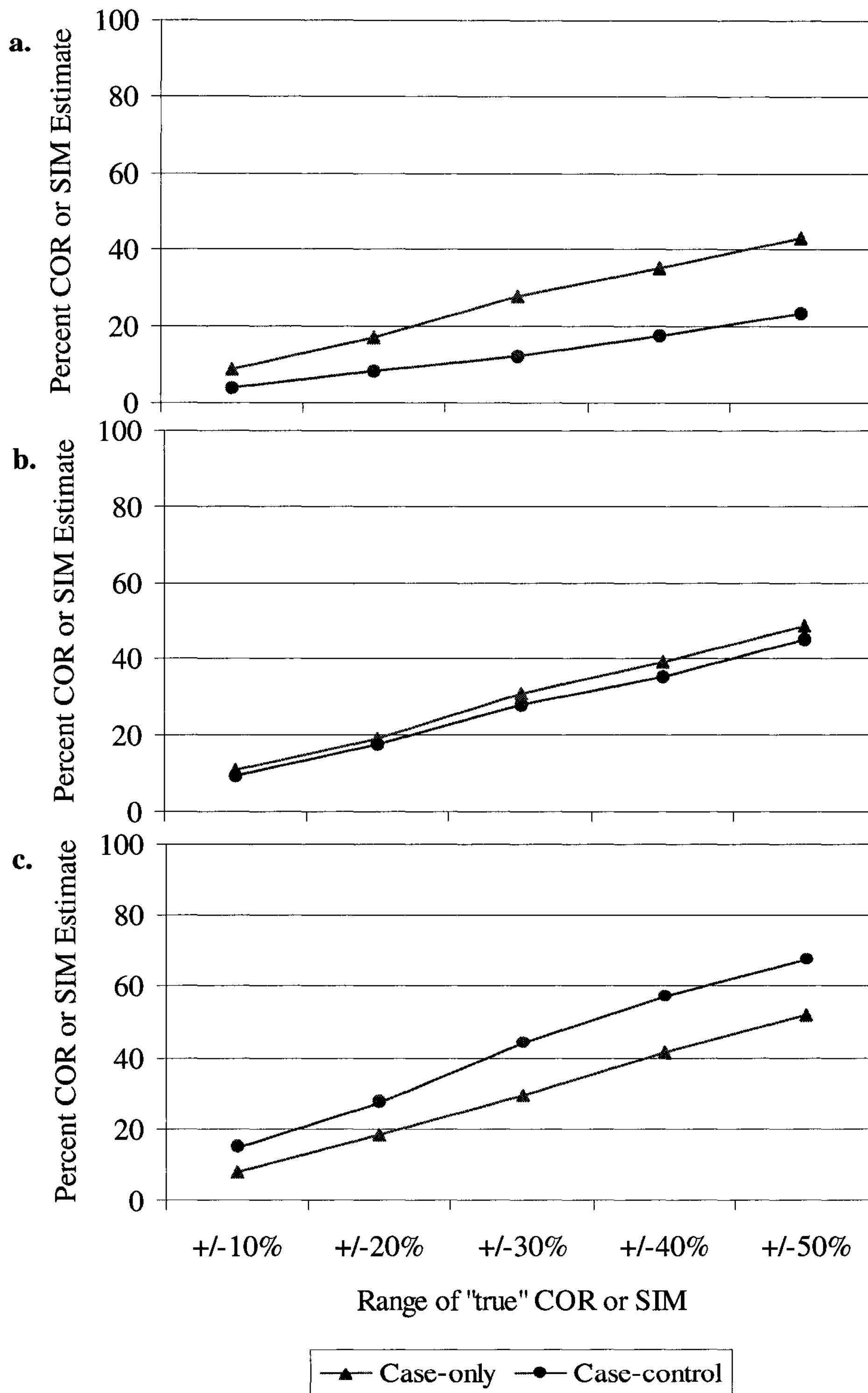



**Figure 5-2. Effect of sample size on accuracy of case-only and case-control estimates from the Typical population when the independence assumption is violated by a magnitude of 1.5 ( a: n=100; b: n=300; c: n=700).** Numerals with lines indicate percent of 1000 replicates included (fitted logistic model converged or OR < 15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

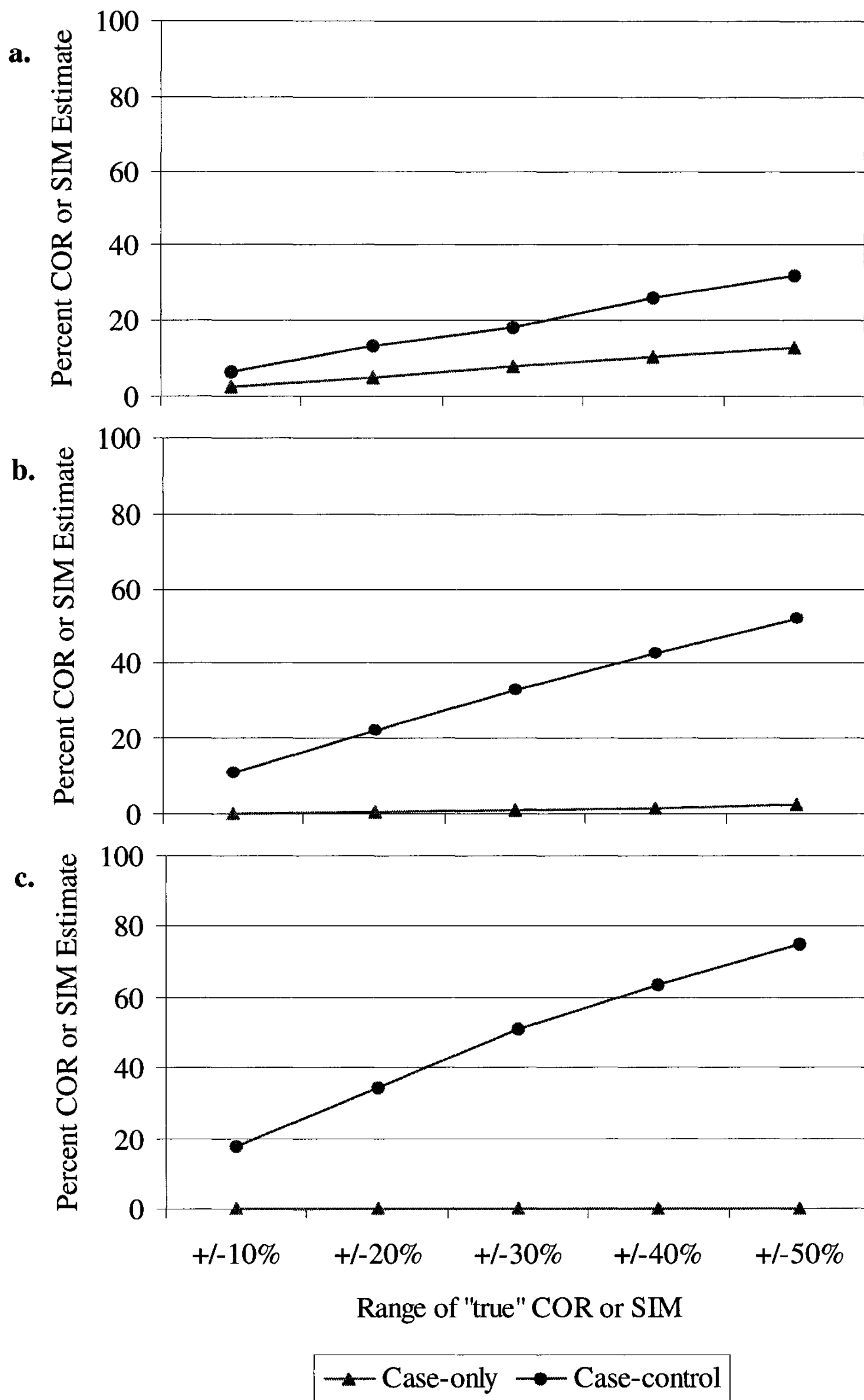



**Figure 5-3. Effect of sample size on accuracy of case-only and case-control estimates from the Typical population when the independence assumption is violated by a magnitude of 3 ( a: n=100; b: n=300; c: n=700).** Numerals with lines indicate percent of 1000 replicates included (fitted logistic model converged or OR < 15).



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER 6

# CONCLUSIONS

Advances in medicine and biotechnology have led to health improvements in ways that were, until recently, inconceivable. Although impressive, scientific progress thus far has been mostly limited to improvements in diagnostic techniques thought to result in disease prevention, and drug therapies used to treat symptoms or delay the onset of disease. Meanwhile, the etiology of many diseases continues to elude our understanding. As a result, we are increasingly challenged with diseases of senescence such as cancer and neurodegenerative diseases which will continue to grow in importance commensurate with increasing life expectancies. This changing face of public health in developed countries has necessitated development of new epidemiologic paradigms that more strongly emphasize the multifactorial origin of chronic disease. Fortunately, knowledge of genetics and availability of novel molecular techniques to detect genetic susceptibility to disease have advanced concurrent with increasing epidemiologic interest in the combined contribution of genes and environment in the pathogenesis of diseases. Thus far, however, few epidemiologic studies have explored the role of various gene-environment interactions in the etiology of neurodegenerative processes. This dissertation was undertaken to develop, evaluate, and integrate new concepts in epidemiologic investigations toward a fuller understanding of such interaction effects using Parkinson's disease as a model.

There is strong evidence of an association between pesticide exposure and Parkinson's disease (PD), a common disorder of the brain whose etiology remains



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

unknown despite a century of research. Genetic susceptibility is also thought to play an important role in the development of this devastating disease. In our study, we hypothesized that individuals who have had prolonged exposure to pesticides but fail to efficiently eliminate them due to polymorphisms that reduce the effectiveness of a detoxification enzyme encoded by the CYP2D6 gene, have a multiplicatively increased risk of PD as compared to individuals with the normal form of the gene who are also exposed to pesticides. We used the case-only design in which only subjects with disease are studied. This type of study has been proposed as a more efficient yet precise alternative to the traditional case-control study when the goal of an investigation is to evaluate gene-environment interaction effects, as in this case.

Thirty patients were needed in the study in order to detect an interaction effect with a magnitude of ten with 80% power based on prevalences of genotype, exposure, disease, and risk of pesticide exposure obtained from published reports. However, given time, resources, and financial constraints, we were able to recruit only 21 PD patients. We determined amount and duration of pesticide exposure for each subject in a detailed personal interview. To assess genotype, we asked patients to provide self-collected samples of epithelial tissue as a source for genomic DNA. This method is preferred by epidemiologists as a safer and more convenient protocol for both investigators and patients, thus maximizing study participation. Although the majority of patients reported exposure to pesticides, there were no CYP2D6 poor metabolizers among our subjects, preventing assessment of interaction effects. Using simulation studies, we later demonstrated the high probability of failure to enroll genetically susceptible cases into a study with a small sample size, especially when the genotype of interest is rare (7%



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

prevalence in our population). Moreover, due to limitations inherent in our selected study design, we were unable to determine the risk of exposure alone. However, using data from a published case-control study in which controls shared similar characteristics to our cases, we were able to determine a statistically significant association between gardening and PD, and an increased but not statistically significant risk for ever versus never pesticide exposure in our subjects as compared to the historical controls. However, these results are subject to more bias and random error, and thus, are likely to be less accurate than what we may have obtained with an internal control group.

The CYP2D6 gene encoding for a Phase I detoxification enzyme is characterized as highly polymorphic, with many different forms leading to poor metabolism of substrates specific to the enzyme, including pesticides. Although many alleles for the poor metabolizer phenotype (PM) have been described, the phenotype is attributable to five distinct PM alleles in over 99% of Caucasians, and protocols have been developed for simultaneous detection of all five PM alleles. Published protocols to date have been based on blood tissue as the source for genomic DNA. The second part of this dissertation involved the evaluation of a published polymerase chain reaction (PCR) protocol for simultaneous detection of the five alleles for use with DNA derived from buccal cells.

Results of the published assay proved to be irreproducible, despite numerous attempts at modification and optimization of the protocol. This failure was most likely due to excessive fragmentation of buccal DNA during the extraction protocol which involves repeated vigorous mixing of the product making it unsuitable for use in the assay. We were able to design a new PCR assay to detect the two most common



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CYP2D6 PM alleles in Caucasians using DNA from epithelial cells. We subsequently validated our results by sequencing the PCR product. Although our assay provides a robust, rapid, and economical alternative to published protocols requiring blood samples, we were unable to completely eliminate the possibility of spurious results from other non-functional but highly homologous pseudogenes that make interpretation of results difficult.

Based on our findings, we conclude that buccal cells are an inadequate source of DNA for accurate detection of the five most common CYP2D6 alleles in Caucasians. We recommend collection of blood samples for future studies of this type. Despite the many advantages and convenience of collection of buccal cells in epidemiologic studies, the choice for source of DNA should be driven by the characteristics of the gene of interest that may require special laboratory techniques for valid and consistent genotype results, bearing in mind that most available laboratory protocols have been designed to be used with blood tissue. Additionally, since blood samples provide more DNA of better quality than other biological specimen, we recommend collection of blood specimen when the complexity of a gene is established or unclear.

Finally, we designed a simulation study to evaluate the performance of the case-only study design to detect gene-environment interactions compared to the case-control study using numerous hypothetical scenarios representing a broad range of population characteristics. We evaluated the accuracy of estimates obtained from each study design with respect to interaction effects for samples of increasingly larger sizes randomly selected from populations with known risks. We additionally compared the relative accuracy of case-control and case-only designs to detect gene-environement interaction



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

effects for various sample sizes and population characteristics. We confirmed that the case-only study produces more accurate results as long as the genotype and exposure are not highly associated among the population, an assumption that must be satisfied for validity of case-only estimates. We found, however, that accuracy of case-only estimates is not affected by violations of the assumption in samples drawn from a population in which the assumption is satisfied.

Additionally, we demonstrated that both case-control and case-only studies fail to provide accurate results under certain extreme conditions. Specifically, small sample sizes (less than 200 subjects) often lead to a substantial amount of random error in the study that may prohibit evaluation of interaction effects in a case-only study by increasing the chance of selecting a sample of cases without the gene of interest.

Thus, despite meticulous attention to avoid systematic errors caused by poorly designed studies of either type, results are often untenable if the sample size of the study is too small. The limitations of small sample sizes are especially important to consider if the frequency of genotype and/or exposure is rare in the population. When main effect determination is an additional goal in a gene-environment interaction study, requiring the use of a case-control design, we found that interaction effects are estimated with less accuracy than in a case-only study, especially given small sample sizes and low prevalences of gene and exposure. Our results suggest that elimination of systematic error (bias) may have little effect on accuracy of estimates when the effects of random error are large. While we encourage reduction of bias through careful design of a study, we strongly recommend that equal importance be placed on the size of study based on characteristics of the population. Accuracy and validity of results can only be improved



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

when both types of error are considered and minimized which often will require using large numbers of subjects in a study. Obviously, conduct of such investigations will require substantial time, financial, and resource commitments on the part of investigators and institutions. Nevertheless, we believe epidemiologic results are most reliable and can contribute more valuable scientific information when obtained from sufficiently large and well designed studies, even at the sacrifice of conducting fewer such studies.

Taken together, results of this dissertation impart some insight into the advantages and limitations of several proposed methods for epidemiologic investigations of gene-environment interactions in the development of disease.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

1. Khoury MJ. Genetic and epidemiologic approaches to the search for gene-environment interaction: the case of osteoporosis. American Journal of Epidemiology 1998;147:1-2.

2. Botto LD, Khoury MJ. Commentary: facing the challenge of gene-environment interaction: the two-by-four table and beyond. Am.J.Epidemiol. 2001;153:1016-1020.

3. Khoury MJ. Commentary: epidemiology and the continuum from genetic research to genetic testing. Am.J.Epidemiol. 2002;156:297-299.

4. Ben Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. Int.J.Epidemiol. 2002;31:285-293.

5. Millikan R. The changing face of epidemiology in the genomics era. Epidemiology 2002;13:472-480.

6. Ottman R. Gene-environment interaction: definitions and study designs. Prev.Med. 1996;25:764-770.

7. Ottman R. An epidemiologic approach to gene-environment interaction. Genet.Epidemiol. 1990;7:177-185.

8. Campbell H. Gene environment interaction. J.Epidemiol.Community Health 1996;50:397-400.

9. Andrieu N, Goldstein AM. Epidemiologic and genetic approaches in the study of gene-environment interaction: an overview of available methods. Epidemiol.Rev. 1998;20:137-147.

10. Taioli E, Zocchetti C, Garte S. Models of interaction between metabolic genes and environmental exposure in cancer susceptibility. Environ.Health Perspect. 1998;106:67-70.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

11. Khoury MJ, Little J. Human genome epidemiologic reviews: the beginning of something HuGE. Am.J.Epidemiol. 2000;151:2-3.

12. Veldman BA, Wijn AM, Knoers N, Praamstra P, Horstink MW. Genetic and environmental risk factors in Parkinson's disease. Clinical Neurology & Neurosurgery 1998;100:15-26.

13. Ben-Shlomo Y. How far are we in understanding the cause of Parkinson's disease? Journal of Neurology, Neurosurgery & Psychiatry 1996;61:4-16.

14. Checkoway H, Nelson LM. Epidemiologic approaches to the study of Parkinson's disease etiology. Epidemiology 1999;10:327-336.

15. Langston JW. Epidemiology versus genetics in Parkinson's disease: progress in resolving an age-old debate. Annals of Neurology 1998;44:S45-52.

16. Marion SA. The epidemiology of Parkinson's disease. Current issues. Adv.Neurol. 2001;86:163-172.

17. Hertzman C, Wiens M, Bowering D, Snow B, Calne D. Parkinson's disease: a case-control study of occupational and environmental risk factors. American Journal of Industrial Medicine 1990;17:349-355.

18. Tanner CM, Aston DA. Epidemiology of Parkinson's disease and akinetic syndromes. Current Opinion in Neurology 2000;13:427-430.

19. Cummings JL. Understanding Parkinson disease. JAMA 1999;281:376-378.

20. Rajput AH, Offord KP, Beard CM, Kurland LT. Epidemiology of parkinsonism: incidence, classification, and mortality. Annals of Neurology 1984;16:278-282.

21. Greenland S, Rothman KJ. Concepts of Interaction. In: Rothman KJ, Greenland S, eds. Modern Epidemiology. Philadelphia: Lippincott-Raven, 1998:3-673.

22. Pearce N. Analytical implications of epidemiological concepts of interaction. Int.J.Epidemiol. 1989;18:976-980.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

23. Smith PG, Day NE. The design of case-control studies: the influence of confounding and interaction effects. Int.J.Epidemiol. 1984;13:356-365.

24. De Michele G, Filla A, Volpe G, De Marco V, Gogliettino A, Ambrosio G, Marconi R, Castellano AE, Campanella G. Environmental and genetic risk factors in Parkinson's disease: a case-control study in southern Italy. Movement Disorders 1996;11:17-23.

25. Priyadarshi A, Khuder SA, Schaub EA, Priyadarshi SS. Environmental risk factors and Parkinson's disease: a metaanalysis. Environmental Research 2001;86:122-127.

26. Gorell JM, Johnson CC, Rybicki BA, Peterson EL, Richardson RJ. The risk of Parkinson's disease with exposure to pesticides, farming, well water, and rural living. Neurology 1998;50:1346-1350.

27. Schoenberg BS. Environmental risk factors for Parkinson's disease: the epidemiologic evidence. Canadian Journal of Neurological Sciences 1987;14:407-413.

28. Tanner CM, Chen B, Wang WZ, Peng ML, Liu ZL, Liang XL, Kao LC, Gilley DW, Schoenberg BS. Environmental factors in the etiology of Parkinson's disease. Canadian Journal of Neurological Sciences 1987;14:419-423.

29. Tanner CM. The role of environmental toxins in the etiology of Parkinson's disease. Trends in Neurosciences 1989;12:49-54.

30. Rajput AH. Environmental toxins accelerate Parkinson's disease onset. Neurology 2001;56:4-5.

31. Tuchsen F, Jensen AA. Agricultural work and the risk of Parkinson's disease in Denmark, 1981-1993. Scand J Work Environ Health 2000;26:359-362.

32. Paganini-Hill A. Risk factors for parkinson's disease: the leisure world cohort study. Neuroepidemiology 2001;20:118-124.

33. Preux PM, Condet A, Anglade C, Druet-Cabanac M, Debrock, C, Macharia W, Couratier P, Boutros-Toni F, Dumas M. Parkinson's disease and environmental



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

factors. Matched case- control study in the Limousin region, France. Neuroepidemiology 2000;19:333-337.

34. Kirkey KL, Johnson CC, Rybicki BA, Peterson EL, Kortsha, GX, Gorell JM. Occupational categories at risk for Parkinson's disease. American Journal of Industrial Medicine 2001;39:564-571.

35. Logroscino G, Marder K, Cote L, Tang MX, Shea S, Mayeux R. Dietary lipids and antioxidants in Parkinson's disease: a population-based, case-control study. Annals of Neurology 1996;39:89-94.

36. Le Couteur DG, McLean AJ, Taylor MC, Woodham BL, Board PG. Pesticides and Parkinson's disease. Biomedicine & Pharmacotherapy 1999;53:122-130.

37. Corrigan FM, Wienburg CL, Shore RF, Daniel SE, Mann D. Organochlorine insecticides in substantia nigra in Parkinson's disease. Journal of Toxicology & Environmental Health 2000;59:229-234.

38. Thiruchelvam M, Richfield EK, Baggs RB, Tank AW, Cory-Slechta DA. The Nigrostriatal Dopaminergic System as a Preferential Target of Repeated Exposures to Combined Paraquat and Maneb: Implications for Parkinson's Disease. Journal of Neuroscience 2000;20:9204-9214.

39. Corrigan FM, Wienburg CL, Shore RF, Daniel SE, Mann D. Organochlorine insecticides in substantia nigra in Parkinson's disease. Journal of Toxicology and Environmental Health 2000;59:229-234.

40. Taylor MC, Le Couteur DG, Mellick GD, Board PG. Paraoxone polymorphisms, pesticide exposure and Parkinson's disease in a Caucasian population. Journal of Neural Transmission 2000;107:979-983.

41. Lockwood AH. Pesticides and Parkinsonism: Is there an etiological link? Current Opinion in Neurology 2000;13:687-690.

42. Friedrich MJ. Pesticide study aids Parkinson research. JAMA 1999;282:2200.

43. Helmuth L. Pesticide causes Parkinson's in rats. Science 2000;290:1068.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

44. Uversky VN, Li J, Fink AL. Pesticides directly accelerate the rate of alpha-synuclein fibril formation: a possible factor in Parkinson's disease. FEBS Letters 2001;500:105-108.

45. Bhatt MH, Elias MA, Mankodi AK. Acute and reversible parkinsonism due to organophosphate pesticide intoxication: five cases. Neurology 1999;52:1467-1471.

46. Tanner CM, Ben-Shlomo Y. Epidemiology of Parkinson's disease. Advances in Neurology 1999;80:153-159.

47. Tanner CM, Ottman R, Goldman SM, Ellenberg J, Chan P, Mayeux R, Langston JW. Parkinson disease in twins: an etiologic study. JAMA 1999;281:341-346.

48. Calne S, Schoenberg B, Martin W, Uitti RJ, Spencer P, Calne DB. Familial Parkinson's disease: possible role of environmental factors. Canadian Journal of Neurological Sciences 1987;14:303-305.

49. Ecobichon DJ, Joy RM. Pesticides and Neurological Diseases. Second. Boca Raton, Ann Arbor, Boston, London: CRC Press, 1993.

50. Smith CA, Gough AC, Leigh PN, Summers BA, Harding AE, Maraganore DM, Sturman SG, Schapira AH, Williams AC, et a. Debrisoquine hydroxylase gene polymorphism and susceptibility to Parkinson's disease. Lancet 1992;339:1375-1377.

51. Kroemer HK, Eichelbaum M. It's the genes, stupid: molecular bases and clinical consequences of genetic cytochrome P450 2D6 polymorphism. Life Sciences 1995;56:2285-2298.

52. Nebert DW. Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist? American Journal of Human Genetics 1997;60:265-271.

53. Wolf CR, Smith G. Cytochrome P450 CYP2D6. IARC Sci.Publ. 1999;209-229.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

54. Langston JW. The etiology of Parkinson's disease with emphasis on the MPTP story. Neurology 1996;47:S153-S160.

55. Kurth MC, Kurth JH. Variant cytochrome P450 CYP2D6 allelic frequencies in Parkinson's disease. American Journal of Medical Genetics 1993;48:166-168.

56. Goldstein AM, Falk RT, Korczak JF, Lubin JH. Detecting gene-environment interactions using a case-control design. Genet.Epidemiol. 1997;14:1085-1089.

57. Hamajima N, Yuasa H, Matsuo K, Kurobe Y. Detection of gene-environment interaction by case-only studies. Jpn.J.Clin.Oncol. 1999;29:490-493.

58. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. Stat.Med. 1994;13:153-162.

59. Hwang SJ, Beaty TH, Liang KY, Coresh J, Khoury MJ. Minimum sample size estimation to detect gene-environment interaction in case-control designs. American Journal of Epidemiology 1994;140:1029-1037.

60. Garcia-Closas M, Rothman N, Lubin J. Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size. Cancer Epidemiol.Biomarkers Prev. 1999;8:1043-1050.

61. Rothman N, Garcia-Closas M, Stewart WT, Lubin J. Chapter 9. The impact of misclassification in case-control studies of gene-environment interactions. IARC Sci.Publ. 1999;89-96.

62. Garcia-Closas M, Lubin JH. Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches. Am.J.Epidemiol. 1999;149:689-692.

63. Garcia-Closas M, Thompson WD, Robins JM. Differential misclassification and the assessment of gene-environment interactions in case-control studies. Am.J.Epidemiol. 1998;147:426-433.

64. Lubin JH, Gail MH. On power and sample size for studying features of the relative odds of disease. Am.J.Epidemiol. 1990;131:552-566.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

65. Khoury MJ, Beaty TH, Hwang SJ. Detection of genotype-environment interaction in case-control studies of birth defects: how big a sample size? Teratology 1995;51:336-343.

66. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. Am.J.Epidemiol. 1989;129:125-137.

67. Thomas DC, Greenland S. The efficiency of matching in case-control studies of risk-factor interactions. J.Chronic.Dis. 1985;38:569-574.

68. Elbaz A, Alperovitch A. Bias in association studies resulting from gene-environment interactions and competing risks. Am.J.Epidemiol. 2002;155:265-272.

69. Albert PS, Ratnasinghe D, Tangrea J, Wacholder S. Limitations of the case-only design for identifying gene-environment interactions. Am.J.Epidemiol. 2001;154:687-693.

70. Schmidt S, Schaid DJ. Potential misinterpretation of the case-only study to assess gene-environment interaction. Am.J.Epidemiol. 1999;150:878-885.

71. Morabia A, Ten Have T, Landis JR. Interaction fallacy. J.Clin.Epidemiol. 1997;50:809-812.

72. Weed DL, Selmon M, Sinks T. Links between categories of interaction. Am.J.Epidemiol. 1988;127:17-27.

73. Suissa S. The case-time-control design: further assumptions and conditions. Epidemiology 1998;9:441-445.

74. Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence intervals for measures of interaction. Epidemiology 1996;7:286-290.

75. Parkinson J. An essay on the shaking palsy. 1. London: Sherwood, Neely, and Jones, 1817.

76. Payami H, Zareparsi S. Genetic epidemiology of Parkinson's disease. Journal of Geriatric Psychiatry & Neurology 1998;11:98-106.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

77. Siderowf A. Parkinson's disease: clinical features, epidemiology and genetics. Neurologic Clinics 2001;19:565-578.

78. Lang AE, Lozano AM. Parkinson's disease. Second of two parts. New England Journal of Medicine 1998;339:1130-1143.

79. Olanow CW, Tatton WG. Etiology and pathogenesis of Parkinson's disease. Annual Review of Neuroscience 1999;22:123-144.

80. Schrag A, Good CD, Miszkiel K, Morris HR, Mathias CJ, Lees AJ, Quinn NP. Differentiation of atypical parkinsonian syndromes with routine MRI. Neurology 2000;54:697-702.

81. Jankovic J, Rajput AH, McDermott MP, Perl DP. The evolution of diagnosis in early Parkinson disease. Parkinson Study Group. Archives of Neurology 2000;57:369-372.

82. Danysz W. Neurotoxicity as a mechanism for neurodegenerative disorders: basic and clinical aspects. Expert Opinion on Investigational Drugs 2001;10:985-989.

83. Jenner P, Olanow CW. Oxidative stress and the pathogenesis of Parkinson's disease. Neurology 1996;Suppl 3:S161-S170.

84. Foley P, Riederer P. Influence of neurotoxins and oxidative stress on the onset and progression of Parkinson's disease. Journal of Neurology 2000;247 Suppl 2:II82-94.

85. Nakamura K, Bindokas VP, Kowlessur D, Elas M, Milstien S, Marks JD, Halpern HJ, Kang UJ. Tetrahydrobiopterin scavenges superoxide in dopaminergic neurons. Journal of Biological Chemistry 2001;276:34402-34407.

86. Zecca L, Pietra R, Goj C, Mecacci C, Radice D, Sabbioni E. Iron and other metals in neuromelanin, substantia nigra, and putamen of human brain. J.Neurochem. 1994;62:1097-1101.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

87. Kienzl E, Puchinger L, Jellinger K, Linert W, Stachelberger H, Jameson RF. The role of transition metals in the pathogenesis of Parkinson's disease. J.Neurol.Sci. 1995;134 Suppl:69-78.

88. Sherer TB, Betarbet R, Greenamyre JT. Pathogenesis of Parkinson's disease. Curr.Opin.Investig.Drugs 2001;2:657-662.

89. Turner C, Schapira AH. Mitochondrial dysfunction in neurodegenerative disorders and ageing. Adv.Exp.Med.Biol. 2001;487:229-251.

90. Zhang Y, Dawson VL, Dawson TM. Oxidative stress and genetics in the pathogenesis of Parkinson's disease. Neurobiol.Dis. 2000;7:240-250.

91. Ames BN, Shigenaga MK, Hagen TM. Mitochondrial decay in aging. Biochim.Biophys.Acta 1995;1271:165-170.

92. Wechsler LS, Checkoway H, Franklin GM, Costa LG. A pilot study of occupational and environmental risk factors for Parkinson's disease. Neurotoxicology 1991;12:387-392.

93. Kidd PM. Parkinson's disease as multifactorial oxidative neurodegeneration: implications for integrative management. Altern.Med.Rev. 2000;5:502-529.

94. Orth M, Schapira AH. Mitochondrial involvement in Parkinson's disease. Neurochem.Int. 2002;40:533-541.

95. Swerdlow RH, Parks JK, Miller SW, Tuttle JB, Trimmer PA, Sheehan JP, Bennett JP, Jr., Davis RE, Parker WD, Jr. Origin and functional consequences of the complex I defect in Parkinson's disease. Ann.Neurol. 1996;40:663-671.

96. Di Monte DA, Chan P, Sandy MS. Glutathione in Parkinson's disease: a link between oxidative stress and mitochondrial damage? Ann.Neurol. 1992;32 Suppl:S111-S115.

97. Rajput AH. Frequency and cause of Parkinson's disease. Canadian Journal of Neurological Sciences 1992;19:103-107.

98. Bernstein K. The Parkinson's Web. http://pdweb.mgh.harvard.edu 1995.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

99. Zhang Z, Roman GC. Worldwide occurence of Parkinson's disease: an updated review. Neuroepidemiology 1993;12:195-208.

100. Muthane U, Jain S, Gururaj G. Hunting genes in Parkinson's disease from the roots. Medical Hypotheses 2001;57:51-55.

101. Lux WE, Kurtzke JF. Is Parkinson's disease acquired? Evidence from a geographic comparison with multiple sclerosis. Neurology 1987;37:467-471.

102. Lanska DJ. The geographic distribution of Parkinson's disease mortality in the United States. Journal of the Neurological Sciences 1997;150:63-70.

103. Thiessen B, Rajput AH, Laverty W, Desai H. Age, environments, and the number of substantia nigra neurons. Advances in Neurology 1990;53:201-206.

104. Duvoisin R. A Brief History of Parkinsonism. In: Cedarbaum JM, Gancher ST, eds. Neurologic Clinics. Harcourt Brace Jovanovich, Inc., 1992:301-316.

105. Marder K, Logroscino G, Alfaro B, Mejia H, Halim A, Louis E, Cote L, Mayeux R. Environmental risk factors for Parkinson's disease in an urban multiethnic community. Neurology 1998;50:279-281.

106. Golbe L. Parkinson's disease: nature meets nurture. Lancet 1998;352:1328-1329.

107. Elbaz A, Grigoletto F, Baldereschi M, Breteler MM, Manubens-Bertran JM, Lopez-Pousa S, Dartigues JF, Alperovitch A, Tzourio C, Rocca WA. Familial aggregation of Parkinson's disease: a population-based case-control study in Europe. Neurology 1999;52:1876-1882.

108. Rostami-Hodjegan A, Lennard MS, Woods HF, Tucker GT. Meta-analysis of studies of the CYP2D6 polymorphism in relation to lung cancer and Parkinson's disease. Pharmacogenetics 1998;8:227-238.

109. Bharucha NE, Stokes L, Schoenberg BS, Ward C, Ince S, Nutt JG, Eldridge, R, Calne DB, Mantel N, Duvoisin R. A case-control study of twin pairs discordant for Parkinson's disease: a search for environmental risk factors. Neurology 1986;36:284-288.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

110. Marttila RJ, Kaprio J, Koskenvuo M, Rinne UK. Parkinson's disease in a nationwide twin cohort. Neurology 1988;38:1217-1219.

111. Vaughan JR, Davis MB, Wood NW. Genetics of Parkinsonism: a review. Annals of Human Genetics 2001;65:111-126.

112. Tan EK, Khajavi M, Thornby JI, Nagamitsu S, Jankovic J, Ashizawa T. Variability and validity of polymorphism association studies in Parkinson's disease. Neurology 2000;55:533-538.

113. Markopoulou K, Langston JW. Candidate genes and Parkinson's disease: where to next? Neurology 1999;53:1382-1383.

114. Nicholl DJ, Bennett P, Hiller L, Bonifati V, Vanacore N, Fabbrini G, Marconi R, Colosimo C, Lamberti P, Stocchi F, Bonuccelli U, Vieregge P, Ramsden DB, Meco G, Williams AC. A study of five candidate genes in Parkinson's disease and related neurodegenerative disorders. European Study Group on Atypical Parkinsonism. Neurology 1999;53:1415-1421.

115. Mouradian MM. Recent advances in the genetics and pathogenesis of Parkinson disease. Neurology 2002;58:179-185.

116. Harada S, Fujii C, Hayashi A, Ohkoshi N. An association between idiopathic Parkinson's disease and polymorphisms of phase II detoxification enzymes: glutathione S-transferase M1 and quinone oxidoreductase 1 and 2. Biochem.Biophys.Res.Commun. 2001;288:887-892.

117. Maccioni RB, Munoz JP, Barbeito L. The molecular bases of Alzheimer's disease and other neurodegenerative disorders. Arch.Med.Res. 2001;32:367-381.

118. Checkoway H, Costa LG, Woods JS, Castoldi AF, Lund BO, Swanson PD. Peripheral blood cell activities of monoamine oxidase B and superoxide dismutase in Parkinson's disease. J.Neural Transm.Park Dis.Dement.Sect. 1992;4:283-290.

119. Mellick GD, Buchanan DD, Silburn PA, Chan DK, Le Couteur DG, Law LK, Woo J, Pang CP. The monoamine oxidase B gene GT repeat polymorphism and Parkinson's disease in a Chinese population. Journal of Neurology 2000;247:52-55.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

120. McGuire V, Nelson LM, Koepsell TD, Checkoway H, Longstreth WT, Jr. Assessment of occupational exposures in community-based case- control studies. Annual Review of Public Health 1998;19:35-53.

121. Tanner CM, Chen B, Wang W, Peng M, Liu Z, Liang X, Kao, LC, Gilley DW, Goetz CG, Schoenberg BS. Environmental factors and Parkinson's disease: a case-control study in China. Neurology 1989;39:660-664.

122. Menegon A, Board PG, Blackburn AC, Mellick GD, Le Couteur DG. Parkinson's disease, pesticides, and glutathione transferase polymorphisms. Lancet 1998;352:1344-1346.

123. Stroombergen MC, Waring RH. Determination of glutathione S-transferase mu and theta polymorphisms in neurological disease. Hum.Exp.Toxicol. 1999;18:141-145.

124. Arand M, Muhlbauer R, Hengstler J, Jager E, Fuchs J, Winkler L, Oesch F. A multiplex polymerase chain reaction protocol for the simultaneous analysis of the glutathione S-transferase GSTM1 and GSTT1 polymorphisms. Analytical Biochemistry 1996;236:184-186.

125. Rahbar A, Kempkes M, Muller T, Reich S, Welter FL, Meves S, Przuntek H, Bolt HM, Kuhn W. Glutathione S-transferase polymorphism in Parkinson's disease. J.Neural Transm. 2000;107:331-334.

126. Smargiassi A, Mutti A, De Rosa A, De Palma G, Negrotti A, Calzetti S. A case-control study of occupational and environmental risk factors for Parkinson's disease in the Emilia-Romagna region of Italy. Neurotoxicology 1998;19:709-712.

127. De Palma G, Mozzoni P, Mutti A, Calzetti S, Negrotti A. Case-control study of interactions between genetic and environmental factors in Parkinson's disease. Lancet 1998;352:1986-1987.

128. Riedl AG, Watts PM, Jenner P, Marsden CD. P450 enzymes and Parkinson's disease: the story so far. Movement Disorders 1998;13:212-220.

129. Christensen PM, Gotzsche PC, Brosen K. The sparteine/debrisoquine (CYP2D6) oxidation polymorphism and the risk of Parkinson's disease: a meta-analysis. Pharmacogenetics 1998;8:473-479.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

130. Le Couteur DG, McCann SJ. P450 enzymes and Parkinson's disease. Movement Disorders 1998;13:851-852.

131. McCann SJ, Le Couteur DG, Green AC, Brayne C, Johnson AG, Chan D, McManus ME, Pond SM. The epidemiology of Parkinson's disease in an Australian population. Neuroepidemiology 1998;17:310-317.

132. McCann SJ, Pond SM, James KM, Le Couteur DG. The association between polymorphisms in the cytochrome P-450 2D6 gene and Parkinson's disease: a case-control study and meta- analysis. Journal of the Neurological Sciences 1997;153:50-53.

133. Barbeau A, Cloutier T, Roy M, Plasse L, Paris S, Poirier J. Ecogenetics of Parkinson's disease: 4-hydroxylation of debrisoquine. Lancet 1985;2:1213-1216.

134. Tsuneoka Y, Matsuo Y, Iwahashi K, Takeuchi H, Ichikawa Y. A novel cytochrome P450IID6 gene associated with Parkinson's disease. Journal of Biochemistry 1993;114:263-266.

135. Lucotte G, Turpin J, Gerard N, Panserat S, Krishnamoorthy R. Mutation frequencies of the cytochrome CYP2D6 gene on Parkinson's disease patients and their families. American Journal of Medical Genetics 1996;67:361-365.

136. Atkinson A, Singleton AB, Steward A, Ince PG, Perry RH, McKeith IG, Fairbairn AF, Edwardson JA, Daly AK, Morris CM. CYP2D6 is associated with Parkinson's disease but not with dementia with Lewy Bodies or Alzheimer's disease. Pharmacogenetics 1999;9:31-35.

137. Bon MA, Jansen Steur EN, de Vos RA, Vermes I. Neurogenetic correlates of Parkinson's disease: apolipoprotein-E and cytochrome P450 2D6 genetic polymorphism. Neuroscience Letters 1999;266:149-151.

138. Sams C, Mason HJ, Rawbone R. Evidence for the activation of organophosphate pesticides by cytochromes P450 3A4 and 2D6 in human liver microsomes. Toxicology Letters 2000;116:217-221.

139. Stefanovic M, Topic E, Ivanisevic AM, Relja M, Korsic M. Genotyping of CYP2D6 in Parkinson's disease. Clinical Chemistry & Laboratory Medicine 2000;38:929-934.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

140. Sachse C, Brockmoller J, Bauer S, Roots I. Cytochrome P450 2D6 variants in a Caucasian population: allele frequencies and phenotypic consequences. American Journal of Human Genetics 1997;60:284-295.

141. Wong JYY, Seah ES, Lee EJD. Pharmacogenetics: the molecular genetics of CYP2D6 dependent drug metabolism. Ann Acad Med Singapore 2000;29:401-406.

142. Stuven T, Griese EU, Kroemer HK, Eichelbaum M, Zanger, UM. Rapid detection of CYP2D6 null alleles by long distance- and multiplex-polymerase chain reaction. Pharmacogenetics 1996;6:417-421.

143. Armstrong M, Daly AK, Cholerton S, Bateman DN, Idle JR. Mutant debrisoquine hydroxylation genes in Parkinson's disease. Lancet 1992;339:1017-1018.

144. Kosel S, Lucking C, Egensperger R, Mehraein P, Graeber M. Mitochondrial NADH dehydrogenase and CYP2D6 genotypes in Lewy-body parkinsonism. J.Neurosci.Res. 1995;44:174-183.

145. Wang S, Lai MD, Lai ML, Huang JD. R296C and other CYP2D6 mutations in Chinese. Pharmacogenetics 1995;5:385-388.

146. Bordet R, Broly F, Destee A, Libersa C, Lifitte JJ. Lack of relation between genetic polymorphism of cytochrome P-450IID6 and sporadic idiopathic Parkinson's disease. Clinical Neuropharmacology 1996;19:213-221.

147. Diederich N, Hilger C, Goetz CG, Keipes M, Hentges F, Vieregge P, Metz H. Genetic variability of the CYP 2D6 gene is not a risk factor for sporadic Parkinson's disease. Annals of Neurology 1996;40:463-465.

148. Bandmann O, Vaughan J, Holmans S, Marsden CD, Wood NW. Association of slow acetylator genotype for N-acetyltransferase 2 with familial Parkinson's disease. Lancet 1997;350:1136-1139.

149. Steen VM, Andreassen OA, Daly AK, Tefre T, Borresen A, Idle JR, Gulbrandsen A. Detection of the poor metabolizer-associated CYP2D6(D) gene deletion allele by long-PCR technology. Pharmacogenetics 1995;5:215-223.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

150. Roberts R, Joyce P, Kennedy MA. Rapid and comprehensive determination of cytochrome P450 CYP2D6 poor metabolizer genotypes by multiplex polymerase chain reaction. Human Mutation 2000;16:77-85.

151. Vieregge P, von Maravic C, Friedrich HJ. Life-style and dietary factors early and late in Parkinson's disease. Canadian Journal of Neurological Sciences 1992;19:170-173.

152. Aquilonius SM, Hartvig P. A Swedish county with unexpectedly high utilization of anti-parkinsonian drugs. Acta Neurol.Scand. 1986;74:379-382.

153. Rybicki BA, Johnson CC, Uman J, Gorell JM. Parkinson's disease mortality and the industrial use of heavy metals in Michigan. Mov Disord. 1993;8:87-92.

154. Granieri E, Carreras M, Casetta I, Govoni V, Tola MR, Paolino E, Monetti VC, De Bastiani P. Parkinson's disease in Ferrara, Italy, 1967 through 1987. Arch.Neurol. 1991;48:854-857.

155. Svenson LW, Platt GH, Woodhead SE. Geographic variations in the prevalence rates of Parkinson's disease in Alberta. Can.J.Neurol.Sci. 1993;20:307-311.

156. Rajput AH, Uitti RJ, Stern W, Laverty W, O'Donnell K, O'Donnell D, Yuen, WK, Dua A. Geography, drinking water chemistry, pesticides and herbicides and the etiology of Parkinson's disease. Canadian Journal of Neurological Sciences 1987;14:414-418.

157. Hodgson E, Levi PE. Pesticides: an important but underused model for the environmental health sciences. Environmental Health Perspectives 1996;104 Suppl 1:97-106.

158. Semchuk KM, Love EJ, Lee RG. Parkinson's disease and exposure to agricultural work and pesticide chemicals. Neurology 1992;42:1328-1335.

159. Butterfield PG, Valanis BG, Spencer PS, Lindeman CA, Nutt JG. Environmental antecedents of young-onset Parkinson's disease. Neurology 1993;43:1150-1158.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

160. Semchuk KM, Love EJ, Lee RG. Parkinson's disease and exposure to rural environmental factors: a population based case-control study. Canadian Journal of Neurological Sciences 1991;18:279-286.

161. Ho SC, Woo J, Lee CM. Epidemiologic study of Parkinson's disease in Hong Kong. Neurology 1989;39:1314-1318.

162. Hertzman C, Wiens M, Snow B, Kelly S, Calne D. A case-control study of Parkinson's disease in a horticultural region of British Columbia. Movement Disorders 1994;9:69-75.

163. Hubble JP, Cao T, Hassanein RE, Neuberger JS, Koller WC. Risk factors for Parkinson's disease. Neurology 1993;43:1693-1697.

164. Chan DK, Woo J, Ho SC, Pang CP, Law LK, Ng PW, Hung WT, Kwok T, Hui, E, Orr K, Leung MF, Kay R. Genetic and environmental risk factors for Parkinson's disease in a Chinese population. Journal of Neurology, Neurosurgery & Psychiatry 1998;65:781-784.

165. Golbe LI, Farrell TM, Davis PH. Follow-up study of early-life protective and risk factors in Parkinson's disease. Movement Disorders 1990;5:66-70.

166. Seidler A, Hellenbrand W, Robra BP, Vieregge P, Nischan P, Joerg J, Oertel WH, Ulm G, Schneider E. Possible environmental, occupational, and other etiologic factors for Parkinson's disease: a case-control study in Germany. Neurology 1996;46:1275-1284.

167. Liou HH, Tsai MC, Chen CJ, Jeng JS, Chang YC, Chen SY, Chen RC. Environmental risk factors and Parkinson's disease: a case- control study in Taiwan. Neurology 1997;48:1583-1588.

168. Stern M, Dulaney E, Gruber SB, Golbe L, Bergen M, Hurtig H, Gollomp S, Stolley P. The epidemiology of Parkinson's disease. A case-control study of young-onset and old-onset patients. Archives of Neurology 1991;48:903-907.

169. Wong GF, Gray CS, Hassanein RS, Koller WC. Environmental risk factors in siblings with Parkinson's disease. Archives of Neurology 1991;48:287-289.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

170. Koller W, Vetere-Overfield B, Gray C, Alexander C, Chin T, Dolezal J, Hassanein R, Tanner C. Environmental risk factors in Parkinson's disease. Neurology 1990;40:1218-1221.

171. Jimenez-Jimenez FJ, Mateo D, Gimenez-Roldan S. Exposure to well water and pesticides in Parkinson's disease: a case-control study in the Madrid area. Movement Disorders 1992;7:149-152.

172. Simon DK, Mayeux R, Marder K, Kowall NW, Beal MF, Johns, DR. Mitochondrial DNA mutations in complex I and tRNA genes in Parkinson's disease. Neurology 2000;54:703-709.

173. Giasson BI, Lee VM. A new link between pesticides and Parkinson's disease. Nature Neuroscience 2000;3:1227-1228.

174. Betarbet R, Sherer TB, MacKenzie G, Garcia-Osuna M, Panov AV, Greenamyre JT. Chronic systemic pesticide exposure reproduces features of Parkinson's disease. Nat.Neurosci. 2000;3:1301-1306.

175. Gaedigk A. Interethnic differences of drug-metabolizing enzymes. Int.J.Clin.Pharmacol.Ther. 2000;38:61-68.

176. Forsberg L, de Faire U, Morgenstern R. Oxidative stress, human genetic variation, and disease. Arch.Biochem.Biophys. 2001;389:84-93.

177. McElroy S, Sachse C, Brockmoller J, Richmond J, Lira M, Friedman D, Roots I, Silber BM, Milos PM. CYP2D6 genotyping as an alternative to phenotyping for determination of metabolic status in a clinical trial setting. AAPS.PharmSci. 2000;2:E33.

178. Collins A, Lonjou C, Morton NE. Genetic epidemiology of single-nucleotide polymorphisms. Proc.Natl.Acad.Sci.U.S.A 1999;96:15173-15177.

179. Payami H, Lee N, Zareparsi S, Gonzales McNeal M, Camicioli R, Bird TD, Sexton G, Gancher S, Kaye J, Calhoun D, Swanson PD, Nutt J. Parkinson's disease, CYP2D6 polymorphism, and age. Neurology 2001;56:1363-1370.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

180. Harty LC, Shield PG, Winn DM, Caporaso NE, Hayes RB. Self-collection of oral epithelial cell DNA under instruction from epidemiologic interviewers. American Journal of Epidemiology 2000;151:199-205.

181. Garcia-Closas M, Egan KM, Abruzzo J, Newcomb PA, Titus-Ernstoff L, Franklin T, Bender PK, Beck JC, Le Marchand L, Lum A, Alavanja M, Hayes RB, Rutter J, Buetow K, Brinton LA, Rothman N. Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. Cancer Epidemiol.Biomarkers Prev. 2001;10:687-696.

182. Sachse C, Brockmoller J, Hildebrand M, Muller K, Roots, I. Correctness of prediction of the CYP2D6 phenotype confirmed by genotyping 47 intermediate and poor metabolizers of debrisoquine. Pharmacogenetics 1998;8:181-185.

183. Subcommittee E48.02 on Characterization and Identification of Biological Systems ASfTaMACEoB. Standard Guide for Detection of Nucleic Acid Sequences by the Polymerase Chain Reaction Technique. 1997.

184. Kuller LH. The etiology of breast cancer--from epidemiology to prevention. Public Health Rev. 1995;23:157-213.

185. Daudt A, Alberg AJ, Helzlsouer KJ. Epidemiology, prevention, and early detection of breast cancer. Curr.Opin.Oncol. 1996;8:455-461.

186. Guo S-W. Gene-environment interaction and the mapping of complex traits: some statistical models and their implications. Human Heredity 2000;50:286-303.

187. Viboud C, Boelle PY, Kelly J, Auquier A, Schlingmann J, Roujeau JC, Flahault A. Comparison of the statistical efficiency of case-crossover and case-control designs: application to severe cutaneous adverse reactions. J.Clin.Epidemiol. 2001;54:1218-1227.

188. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. Am.J.Epidemiol. 1980;112:467-470.

189. Mucci LA, Wedren S, Tamimi RM, Trichopoulos D, Adami HO. The role of gene-environment interaction in the aetiology of human cancer: examples from cancers of the large bowel, lung and breast. J.Intern.Med. 2001;249:477-493.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

190. Ishibe N, Kelsey KT. Genetic susceptibility to environmental and occupational cancers. Cancer Causes Control 1997;8:504-513.

191. Scriver CR, Clow CL. Phenylketonuria and other phenylalanine hydroxylation mutants in man. Annu.Rev.Genet. 1980;14:179-202.

192. Hirvonen A, Husgafvel-Pursiainen K, Anttila S, Vainio H. The GSTM1 null genotype as a potential risk modifier for squamous cell carcinoma of the lung. Carcinogenesis 1993;14:1479-1481.

193. Dean G. Porphyria variegata. Acta Derm.Venereol.Suppl (Stockh) 1982;100:81-85.

194. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene- environment interaction: case-control studies with no controls. American Journal of Epidemiology 1996;144:207-213.

195. Sturmer T, Brenner H. Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. Am.J.Epidemiol. 2002;155:593-602.

196. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. Stat.Med. 1983;2:243-251.

197. Yang Q, Khoury MJ, Flanders WD. Sample size requirements in case-only designs to detect gene- environment interaction. American Journal of Epidemiology 1997;146:713-720.

198. Begg CB, Zhang ZF. Statistical analysis of molecular epidemiology studies employing case-series. Cancer Epidemiol.Biomarkers Prev. 1994;3:173-175.

199. Hosmer DW, Lemeshow S. Applied Logistic Regression. Second. John Wiley & Sons, Inc., 2000.

200. Wang J, Liu Z, Chen B. Association between cytochrome P-450 enzyme gene polymorphisms and Parkinson's disease. Zhonghua Yi.Xue.Za Zhi. 2000;80:585-587.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

201. Wang J, Liu Z, Chan P. Lack of association between cytochrome P450 2E1 gene polymorphisms and Parkinson's disease in a Chinese population. Mov Disord. 2000;15:1267-1269.

202. Takakubo F, Yamamoto M, Ogawa N, Yamashita Y, Mizuno Y, Kondo I. Genetic association between cytochrome P450IA1 gene and susceptibility to Parkinson's disease. J.Neural Transm.Gen.Sect. 1996;103:843-849.

203. Daly AK, Brockmoller J, Broly F, Eichelbaum M, Evans WE, Gonzalez FJ, Huang JD, Idle JR, Ingelman-Sundberg M, Ishizaki T, Jacqz-Aigrain E, Meyer UA, Nebert DW, Steen VM, Wolf CR, Zanger UM. Nomenclature for human CYP2D6 alleles. Pharmacogenetics 1996;6:193-201.

204. Ritz B, Yu F. Parkinson's disease mortality and pesticide exposure in California 1984-1994. Int.J.Epidemiol. 2000;29:323-329.

205. Broly F, Marez D, Sabbagh N, Legrand M, Millecamps S, Lo Guidice JM, Boone P, Meyer UA. An efficient strategy for detection of known and new mutations of the CYP2D6 gene using single strand conformation polymorphism analysis. Pharmacogenetics 1995;5:373-384.

206. Scarlett LA, Madani S, Shen DD, Ho RJY. Development and characterization of a rapid and comprehensive genotyping assay to detect the most common variants in cytochrome P450 2D6. Pharmaceutical Research 2000;17:242-246.

207. Hersberger M, Marti-Jaun J, Rentsch K, Hanseler E. Rapid detection of the CYP2D6*3, CYP2D6*4, and CYP2D6*6 alleles by tetra-primer PCR and of the CYP2D6*5 allele by multiplex long PCR. Clinical Chemistry 2000;46:1072-1077.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# APPENDICES



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# Appendix A

## Consent Form

**COLORADO STATE UNIVERSITY**
**INFORMED CONSENT TO PARTICIPATE IN A RESEARCH PROJECT**

**TITLE OF PROJECT**: Effects of gene-environment interaction in the pathogenesis of Parkinson's disease

**NAME OF PRINCIPAL INVESTIGATOR**: John S. Reif, DVM

**NAME OF CO-INVESTIGATORS**: Susan Hariri, MPH
Marie Legare, DVM, PhD

**CONTACT NAME AND PHONE NUMBER FOR QUESTIONS/PROBLEMS**: Susan Hariri (970) 491-2891

**SPONSOR OF PROJECT**: Colorado State University, College of Veterinary Medicine and Biomedical Sciences, College Research Council

**PURPOSE OF THE RESEARCH**: Parkinson's disease (PD) is a neurological illness for which there is no known cause. It is increasingly evident that the disease is caused by multiple factors involving both heredity and environment. The group of chemicals known as pesticides has been repeatedly shown to be associated with development of PD, however, the exact mechanism and levels of exposure required to produce disease are not known. Moreover, there is variability among individuals in the genes responsible for the proper breakdown of chemicals such as pesticides once a person is exposed, possibly making some individuals more susceptible to developing PD. It is our theory that the interactions between environmental agents, specifically, pesticides, in the presence of certain gene mutations (CYP2D6 PM) increase a person's risk of developing PD. There is very little information available to support our hypothesis currently, however, and more research is needed to determine the effects of gene and environment interactions. This research project is a pilot study to establish the protocols necessary to examine the relationship between a specific gene mutation and exposure to agricultural chemicals and the risk of PD.

**PROCEDURES/METHODS TO BE USED**:
You will be personally interviewed at your home by a member of the research team to determine past environmental exposures that may be relevant to your condition. The interview will consist of a structured questionnaire that will take approximately 30 minutes to complete. You will be asked questions related to exposure to pesticides in your occupation and residence. In addition, lifestyles and family history will be assessed. After the completion of the interview, you will be asked to collect some cell samples from the inside of your cheek using a cotton swab, which will be provided to you. The procedure involves vigorous rubbing of a cotton swab up and down the inside wall of your cheek. The cheek swab containing your genetic material (DNA) will be transferred in a sterile tube to our laboratory at Colorado State University where it will be analyzed for a specific gene mutation (CYP2D6 PM). Once the analysis is completed, the remainder of your sample will be discarded and no portion will be left for future analyses. The collection of your cheek swab will take about 15 minutes to complete for a total of approximately 45 minutes of your time.

**RISKS INHERENT IN THE PROCEDURES**:
There are no known risks by participation in this study. No hazard or physical stress is anticipated from either the interview or the cheek swab cell collection. Cheek swabs are self-collected and can be stopped in case of any discomfort. Questionnaire is self-reported and non-invasive. Interview is voluntary and may be terminated at any time. Participants have the choice to refuse answers to any questions with which they feel uncomfortable. There are no risks associated with any form of the gene under investigation. It is not possible to identify all potential risks in research procedures, but the researcher(s) have taken reasonable safeguards to minimize any known and potential, but unknown, risks.

**BENEFITS**:
There are no known direct benefits to participants.



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**CONFIDENTIALITY**:
All information obtained in this project will be confidential. Questionnaires and cell samples will be coded and matched, after which all identifying information will be removed. All data will be kept in locked files at Colorado State University Department of Environmental Health, and will be accessible only to authorized project staff. Cell samples will not be used for anything other than purposes of this study. Your name or results will not be released to any private party, employer, or insurance company. Your individual results will be provided to you upon request. Results from this study will be released as group data only, and there will be no means of identifying you personally in any published reports.

**LIABILITY**:
The Colorado Governmental Immunity Act determines and may limit Colorado State University's legal responsibility if an injury happens because of this study. Claims against the University must be filed within 180 days of the injury.

Questions about subjects' rights may be directed to Celia S. Walker at (970) 491-1563.

**PARTICIPATION**:
Your participation in this research is voluntary. If you decide to participate in the study, you may withdraw your consent and stop participating at any time without penalty or loss of benefits to which you are otherwise entitled.

Your signature acknowledges that you have read the information stated and willingly sign this consent form. Your signature also acknowledges that you have received, on the date signed, a copy of this document containing _3_ pages.

________________________________
Participant name (printed)

________________________________ ________________
Participant signature Date

________________________________ ________________
Witness to signature (project staff) Date



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Test results request form**

I would like to be informed of my CYP2D6 gene status. I understand that the results are not interpretable at this time.
Yes
No

Your signature acknowledges that you have read the information stated and willingly sign this consent form. Your signature also acknowledges that you have received, on the date signed, a copy of this document containing __3__ pages.

________________________________
Participant name (printed)

________________________________ __________________
Participant signature Date

________________________________ __________________
Witness to signature (project staff) Date



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## Appendix B

## Questionnaire

## PESTICIDE EXPOSURE AND PARKINSON'S DISEASE



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Before we begin, I'd like to remind you that all information will be strictly confidential. Your answers will not be used for any purposes other than this study and will not be shared with anyone.

**PERSONAL INFORMATION:**

I'd like to start by asking you some questions about yourself.

**What is your date of birth?**

Month

Day

Year

**What is your current marital status?**

- □Now married
- □Widowed
- □Separated or divorced
- □Never married

**What is the highest level of school that you completed?**

- □1-8 years
- □Some high school
- □High school graduate
- □GED (high school equivalency)
- □Vocational training beyond high school
- □Some college
- □College graduate
- □Graduate or professional school
- □Other (specify)

**Which of these income categories best describes your family's total annual income before taxes, from all sources?**

If retired, pre-retirement income before taxes?

□Below $20,000 □$21,000-$40,000

□$41,000-$80,000 □$Over 81,000

**Are you of Hispanic, Latino or Spanish origin?**

- □Yes
  - □Mexican, Mexican American
  - □Chicano
  - □Puerto Rican
  - □Cuban
  - □Other
- □No
- □Don't Know

**Which of the following groups best describes your ethnic origin?**

- □White
- □African American/Black
- □Native American or Alaska Native
- □Asian or Pacific Islander
- □Other (specify________________)
- □Don't Know

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**RESIDENTIAL HISTORY**

In this section, I'm going to ask you about places you have lived.

**Have you ever lived at a residence where your drinking water came from a well?**

- ☐Yes
  - ☐Dug
  - ☐Drilled
  - ☐Don't Know

  How old were you when you started to live in the house with well water?
  - ☐Less than one year old
  - ☐2-5 years old
  - ☐6-10 years old
  - ☐11-20 years old
  - ☐Over 20 years old

  How many years did you drink well water?

- ☐No
- ☐Don't Know

**Have you ever lived in a rural area (population less than 1,000 per square mile in core area; population less than 500 per square mile in surrounding area)?**

- ☐Yes

  How old were you when you began to live in a rural area?
  - ☐Less than one year
  - ☐2-5 years
  - ☐6-10 years
  - ☐11-20 years
  - ☐More than 20 years

  How many years did you live in a rural area?

- ☐No
- ☐Don't Know

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**Have you ever lived on a farm?**

☐Yes

How old were you when you began to live on a farm?

☐Less than one year old
☐2-5 years old
☐6-10 years old
☐11-20 years old
☐Over 20 years old

How many years did you live on this farm?

☐Less than one year
☐2-5 years
☐6-10 years
☐11-20 years
☐Over 20 years

How many acres were planted on this farm?

☐Less than 5 acres
☐5-49 acres
☐50-499 acres
☐More than 500 acres
☐None

What was/were MAJOR crops or animals raised on this farm?

☐Cash grains

(e.g., wheat, rice, corn, soybeans)

☐Field crops

(e.g., cotton, tobacco, sugarcane, sugar beets)

☐Vegetables
☐Fruits and nut trees
☐Ornamental plants and nursery flowers
☐No crops or plants grown on the farm
☐Livestock (e.g., beef cattle, hogs, sheep)
☐Dairy farm
☐Poultry and eggs
☐Other animals
☐No animals raised on farm
☐Other (specify)

☐No

☐Don't Know **Repeat for each farm lived on**

**Have you personally ever kept a garden?**

☐Yes

How large was the garden (repeat for each garden)?

How old were you when you began to garden?

How many years did you garden?

☐Less than one year
☐2-5 years
☐6-10 years
☐11-20 years
☐More than 20 year

Did you ever use pesticides on the garden(s)?

☐Yes

Type of pesticide(s) (insecticide, herbicide, fungicide)
Name of pesticide(s)

______________________

☐Don't Know

How old were you when you began?
How old were you when you stopped?
_______hours/day
_______days/week

☐No
☐Don't Know

☐No
☐Don't Know



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**MILITARY HISTORY**

**Have you ever served in the military?**

□Yes □No

What were the dates of your military service? (Age joined/age left)

Which branch of the armed forces were you in?

□Army □Air Force
□Navy □National Guard
□Marines □Coast Guard
□Reserves

Did you serve in any wars?

□Yes □No

| Which war (s) | Dates served (age) |
|---|---|
| | |

Were you exposed to any nerve agents? (e.g., Sarin, mustard gas, phosgene) □Yes □No

| Name of nerve agent | Date (s) exposed (age) | Type of contact | Length of exposure | Medical attention |
|---|---|---|---|---|
| | | □Dermal<br>□Inhalation<br>□Ingestion | ______Minutes<br>______Hours<br>______Days | □Yes<br>□No |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**OCCUPATIONAL HISTORY**
This section is about your work history.

**What was your main occupation most of your life?**

**What were your job responsibilities/what tasks did you perform?**

**Have you ever worked in any of the following industries for one year or longer?**

☐Agriculture (crop farming, grain farming, orchard work, market gardening, pesticide application)
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Horticulture (e.g., lawn service, tree trimming, landscaping)
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Forestry
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Extermination (home or industrial chemical spraying for termites, roaches, rodents)
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Petroleum, plastic, rubber work
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



☐Paint, lacquer, furniture finishing work
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Mechanics (jet, auto repair, or other engine)
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Machinist
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Printing industry (typography, lithography)
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Chemist
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Textile work, weaving (dry cleaning, shoe industry, leather work)
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Paper mill work
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

☐Wood work
How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
☐Full-time ☐Part-time
_______hours/day
_______days/week

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



□Metal work

How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
□Full-time □Part-time
_______hours/day
_______days/week

□Welding

How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
□Full-time □Part-time
_______hours/day
_______days/week

□Mining

How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
□Full-time □Part-time
_______hours/day
_______days/week

□Refrigerating system work

How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
□Full-time □Part-time
_______hours/day
_______days/week

□Planer mill work

How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
□Full-time □Part-time
_______hours/day
_______days/week

□Other job where you were routinely in contact with chemicals

How old were you when you began?
How old were you when you stopped?
How many years did you work in this area?
□Full-time □Part-time
_______hours/day
_______days/week

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**Have you ever worked on a farm?**

- □Yes (Go to next question)
- □No (Go to next section)
- □Don't Know (Go to next section)

**Over your lifetime, how many years have you worked on a farm?**

- □Less than 5 years
- □5-10 years
- □11-20 years
- □21-30 years
- □Over 30 year

**In your lifetime, have you personally used any pesticides (including insecticides, herbicides, fungicides, and fumigants)?**

- □Yes (Go to next question)
- □No (Go to next section)

**From the following list, please indicate all applications for which you used pesticides.**

- □Termite control
- □Rodent control
- □Lawn and garden
- □Greenhouse
- □Stored grain or other stored agricultural products
- □Highway weed control
- □Forestry applications
- □Aerial spraying
- □Herbicide (weed killers) application to farm crops
- □Insecticide applications to farm crops
- □Insecticide applications to farm animals/animal shelters
- □Insecticide application to pets
- □Insecticide applications in the home
- □Insecticide application in commercial buildings
- □Fungicide (chemicals for controlling disease on crops) application
- □Fumigant (gases or liquids that turn into gas when released) app
- □Other (please specify)
- □None of these

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**How many years did you personally spend mixing or applying pesticides?**

- □One year or less
- □2-5 years
- □6-10 years
- □11-20 years
- □21-30 years
- □More than 30 years
- □Never

**During those years, how many days per year did you personally mix or apply pesticides?**

- □Less than five
- □5-9 days
- □10-19 days
- □20-39 days
- □40-59 days
- □60-150 days
- □More than 150 days
- □Never

**When you mixed pesticides, what percent of the time did you do the mixing?**

- □Less than 50% of the time
- □50% or more of the time
- □I never mixed the pesticides myself

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



CHEMICAL EXPOSURE HISTORY

**Now I'm going to ask you about specific pesticides.**

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Have you ever used herbicides (pesticides used to control weeds)?<br>□**Yes** □**No** | | | | | |
| Atrex, Atranex, or other atrazine products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Lasso, Chimiclor, or other alachlor products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Merphos (cotton defoliant) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Paraquat | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Petroleum oil or petroleum distillates | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Roundup, Jury, or other glyphosate products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Treflan, Trilin, or other trifluralin products | ☐Yes<br>☐No<br>☐Don't Know | ☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐Less than five<br>☐5-9 days<br>☐10-19 days<br>☐20-39 days<br>☐40-59 days<br>☐60-150 days<br>☐More than 150 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |
| 2,4-D | ☐Yes<br>☐No<br>☐Don't Know | ☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐Less than five<br>☐5-9 days<br>☐10-19 days<br>☐20-39 days<br>☐40-59 days<br>☐60-150 days<br>☐More than 150 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |
| 2,4,5 T P (Silvex) or 2,4,5 T (neither on market) | ☐Yes<br>☐No<br>☐Don't Know | ☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐Less than five<br>☐5-9 days<br>☐10-19 days<br>☐20-39 days<br>☐40-59 days<br>☐60-150 days<br>☐More than 150 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**Have you ever used insecticides (on crops, nursery, lawn and garden, or livestock)?**
**□Yes □No**

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Azodrin, or other monocrotophos products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Ambush, Pounce, Asana or other permethrin or pyrethroid products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Bidrin, or other dicrotophos product | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Bromofos | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Chlordecone or Kepone | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Counter or other terbufos products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Cygon, Cyanarid, or other dimethoate products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Dimecron, or other phosphamidon products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Dyfonate or other fonofos products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Dylox, Dipterex, Proxol, Neguron, or other trichlorfon products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Folithion, Sumithion, or other fenitrothion products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Furadan, Curaterr, or other carbofuran products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Lindane or gamma BHC | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Lorasban, Dursban, or other chlorpyrifos products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Malathion | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Monitor, Tamaron, or other methamidophos products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Mipafox (reports of delayed neuropathy) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Noxfish or other rotenone products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Parathion (Niran, Penncap) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Phosdrin, or other mevinphos products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Phosvel or other leptophos products (no longer on market | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Sevin, Carbamine, or other carbaryl products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Spectracide, Dianon, Knox Out, or other diazinon products | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Temik or other aldicarb product | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

| **Name of Pesticide** | **Have you personally mixed or applied?** | **How many years did you mix or apply?** | **In average year, how many days did you use it?** | **In what decade did you personally first start to use this pesticide?** | **What type of contact did you have with this pesticide?** |
|---|---|---|---|---|---|
| Thimet, Rampart, or other phorate product | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Aldrin, Aldrex, Aldrite, or Octalene (no longer on market) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Benzene hexachloride, HCH, BHC (no longer on market) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Chlordane (no longer on market) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Dieldrin, Dieldrex, Dieldrite (no longer on market) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| DDT (no longer on market) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Heptachlor (no longer on market) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Toxaphene (no longer on market) | □Yes<br>□No<br>□Don't Know | □1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Less than five<br>□5-9 days<br>□10-19 days<br>□20-39 days<br>□40-59 days<br>□60-150 days<br>□More than 150 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



| Have you ever used fungicides (chemicals for controlling disease on crops and seed; may have been pre-applied to seed)?<br>☐Yes ☐No | | | | | |
|---|---|---|---|---|---|
| **Name of Pesticide** | **Have you personally mixed or applied?** | **How many years did you mix or apply?** | **In average year, how many days did you use it?** | **In what decade did you personally first start to use this pesticide?** | **What type of contact did you have with this pesticide?** |
| Benlate, Tersan, or other benomyl products | ☐Yes<br>☐No<br>☐Don't Know | ☐already applied to seed<br>☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐Preapplied to seed<br>☐None<br>☐1 day<br>☐2-5 days<br>☐6-10 days<br>☐11-20 days<br>☐More than 20 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |
| Bravo, Evade, Daconil 2787, or other chlorothalonil products | ☐Yes<br>☐No<br>☐Don't Know | ☐already applied to seed<br>☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐Preapplied to seed<br>☐None<br>☐1 day<br>☐2-5 days<br>☐6-10 days<br>☐11-20 days<br>☐More than 20 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |
| Manex, Manzate, Dithane Z-78, or other maneb or mancozeb products | ☐Yes<br>☐No<br>☐Don't Know | ☐already applied to seed<br>☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐Preapplied to seed<br>☐None<br>☐1 day<br>☐2-5 days<br>☐6-10 days<br>☐11-20 days<br>☐More than 20 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |

| Name of Pesticide | Have you personally mixed or applied? | How many years did you mix or apply? | In average year, how many days did you use it? | In what decade did you personally first start to use this pesticide? | What type of contact did you have with this pesticide? |
|---|---|---|---|---|---|
| Orthocide, Clomitane, or other captan products | □Yes<br>□No<br>□Don't Know | □already applied to seed<br>□1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Preapplied to seed<br>□None<br>□1 day<br>□2-5 days<br>□6-10 days<br>□11-20 days<br>□More than 20 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |
| Ridomil, Subdue, or other metalaxyl products | □Yes<br>□No<br>□Don't Know | □already applied to seed<br>□1 year or less<br>□2-5 years<br>□6-10 years<br>□11-20 years<br>□21-30 years<br>□More than 30 years | □Preapplied to seed<br>□None<br>□1 day<br>□2-5 days<br>□6-10 days<br>□11-20 days<br>□More than 20 days | □Before 1960<br>□In the 1960's<br>□In the 1970's<br>□In the 1980's<br>□In the 1990's | □Dermal<br>□Inhalation<br>□Ingestion |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

| **Fumigants (gases or liquids that turn to gas when released; used in enclosed spaces or to treat soil)?**<br>☐**Yes** ☐**No** | | | | | |
|---|---|---|---|---|---|
| Brom-O-Gas, Brom-O-Sol, or other methyl bromide products | ☐Yes<br>☐No<br>☐Don't Know | ☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐None<br>☐1 day<br>☐2-5 days<br>☐6-10 days<br>☐11-20 days<br>☐More than 20 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |
| Phostoxin, Gastoxin, or other aluminum phosphide products | ☐Yes<br>☐No<br>☐Don't Know | ☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐None<br>☐1 day<br>☐2-5 days<br>☐6-10 days<br>☐11-20 days<br>☐More than 20 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |
| Carbon tetrachloride, carbon disulfide (80/20 mix) | ☐Yes<br>☐No<br>☐Don't Know | ☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐None<br>☐1 day<br>☐2-5 days<br>☐6-10 days<br>☐11-20 days<br>☐More than 20 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |
| EDB, E-D-Bee, Bromofume, or other ethylene dibromide products | ☐Yes<br>☐No<br>☐Don't Know | ☐1 year or less<br>☐2-5 years<br>☐6-10 years<br>☐11-20 years<br>☐21-30 years<br>☐More than 30 years | ☐None<br>☐1 day<br>☐2-5 days<br>☐6-10 days<br>☐11-20 days<br>☐More than 20 days | ☐Before 1960<br>☐In the 1960's<br>☐In the 1970's<br>☐In the 1980's<br>☐In the 1990's | ☐Dermal<br>☐Inhalation<br>☐Ingestion |



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Are there other pesticides that you used regularly?**

☐Yes (specify)

☐No

**How often, if ever, have you had the following symptoms that you think may be related to your using pesticides?**

| | | | |
|---|---|---|---|
| Been excessively tired | ☐never/rarely | ☐sometimes | ☐frequently/almost always |
| Had headaches or dizziness | ☐never/rarely | ☐sometimes | ☐frequently/almost always |
| Had nausea or vomiting | ☐never/rarely | ☐sometimes | ☐frequently/almost always |
| Had skin irritation | ☐never/rarely | ☐sometimes | ☐frequently/almost always |
| Had chest discomfort | ☐never/rarely | ☐sometimes | ☐frequently/almost always |
| Felt nervous or depressed | ☐never/rarely | ☐sometimes | ☐frequently/almost always |

**Have you ever had a single episode with any type of pesticide which caused you to become ill?**

☐Yes (Go to next question)
☐No (Go to next section)
☐Don't Know (Go to next section)

**What was the name of this chemical?**

**What year did this experience take place?**

**How soon after you were exposed were you able to wash it off with soap and water?**

☐Less than 30 minutes
☐30 minutes-3 hours
☐4-6 hours
☐More than 7 hours

**Did you breathe this compound?**

☐Yes
☐No
☐Don't Know

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**Did you get any of this compound on any of the following body parts?**

☐Head or face
☐Hands or arms
☐Legs or feet
☐Other (specify)

**Did you swallow any of this chemical?**

☐Yes
☐No
☐Don't Know

**Did you receive medical attention?**

☐Yes
☐No
☐Don't Know

**How do you personally apply pesticides?**

☐Do not personally apply pesticides
☐Airblast
☐Boom on tractor, truck, trailer
  Enclosable cab?
  ☐Yes
    ☐Did you operate with windows and door
      ☐ open
      ☐ closed
  ☐No
☐Hand spray gun
☐Backpack sprayer
☐Mist blower/fogger
☐Aerial (aircraft application)
☐In furrow or banded
☐Seed treatment
☐Distribute granules or tablets
☐Pour fumigant from bucket
☐Gas canister
☐Row fumigation
☐Powder duster
☐Inject animals
☐Dip animals
☐Spray animals
☐Ear tags
☐Dust or pour on animals
☐None of these
☐Other (Specify)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**What type of protective equipment do you generally wear when you personally handle pesticides?**

☐Never use protective equipment
☐Cartridge respirator or gas mask
☐Face shields or goggles
☐Disposable outer clothing (e.g., Tyvek)
☐Fabric or leather gloves
☐Chemically resistant gloves (e.g., neoprene, nitrile)
☐Other (Specify—e.g., boots, apron)

**OTHER CHEMICAL EXPOSURE:**

**Have you worked with any of the following chemicals for one year or longer (if yes, number of years)?**

☐Carbon monoxide
☐Manganese
☐Mercury
☐Aluminum
☐Zinc
☐Paint
☐Lacquer
☐Gasoline
☐Acetone
☐MEK
☐n-Hexane
☐Cyclo-hexane
☐Hepthane
☐Ethyl acetate
☐Freon
☐Toluene
☐Trichloroethylene

**How often have you applied natural fertilizers (manure)?**

☐Never
☐1-5 days
☐6-25 days
☐26-50 days
☐More than 50 days

**How often have you applied chemical fertilizers?**

☐Never
☐1-5 days
☐6-25 days
☐26-50 days
☐More than 50 days

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**PERSONAL HEALTH HISTORY:**
This section includes questions about your present health and about specific illnesses or health problems that you may have had.

**How old were you when you first began to experience symptoms that were later determined to be the first signs of Parkinson's disease?**

**What were the symptoms you were experiencing at that time?**

**How old were you when you were first diagnosed with Parkinson's disease by a neurologist?**

**What prescription medications are you currently taking for the Parkinson's?**
How long have you been taking each?

**Has a doctor ever told you that you have any of the following illnesses or health problems?**

☐Pneumonia (viral or bacterial)
☐Viral encephalitis
☐Epilepsy or seizures
☐Alzheimer's disease
☐Amyotrophic lateral sclerosis (Lou Gehrig's disease)
☐Motor neuron disease
☐Nervous disorder
☐Multiple sclerosis
☐Depression
☐Shingles
☐Mononucleosis
☐Lupus
☐Heart disease
☐Stroke
☐Diabetes
☐Kidney disease
☐Goiter
☐Thyroid disease
☐Brain cancer
☐Other cancer
☐Pesticide poisoning
☐Solvent poisoning
☐Lead poisoning
☐Head injury requiring medical attention

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**Have you ever undergone general anesthesia?**

☐Yes

How many times did you have general anesthesia before you were diagnosed with Parkinson's disease?

☐No

☐Don't Know

**Do you have any amalgam tooth fillings?**

☐Yes

How many amalgam tooth fillings did you have before you were diagnosed with PD?

☐No

☐Don't Know

**During your lifetime, have you smoked at least 100 cigarettes?**

☐Yes

☐No

☐Don't Know

How old were you when you started to smoke cigarettes?

Do you smoke cigarettes now?

☐Yes

☐No

☐Don't Know

On average, how much do you or did you smoke each day?

☐10 or less

☐11-20

☐21-40

☐More than 40

What is the total number of years you smoked cigarettes?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Have you used any of the following tobacco products for six months or longer?**

☐Cigars
☐Cigarillos
☐Pipe
☐Smokeless tobacco
☐None

**At what age did you start using this tobacco product?**

**Are you currently using any tobacco product?**

☐Yes
☐No (age when quit)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**FAMILY INFORMATION:**
Now I'd like to ask you some information about members of your family.

**On your FATHER'S side of your family, in what country was your grandfather born?**

**On your FATHER'S side of your family, in what country was your grandmother born?**

**In what country was your father born?**

**Was your father ever diagnosed with Parkinson's disease?**

- ☐Yes
  - How old was he when he was diagnosed?
- ☐No
- ☐Don't Know

**Did your father ever have a problem with uncontrollable tremor or shaking when trying to do an activity?**

- ☐Yes
  - How old was he when he experienced these symptoms?
- ☐No
- ☐Don't Know

**Is your father alive?**

- ☐Yes
  - Age
- ☐No

At what age did your father pass away?

What was the cause of death?

**What was your father's occupation most of his life?**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**On your MOTHER'S side of your family, in what country was your grandfather born?**

**On your MOTHER'S side of your family, in what country was your grandmother born?**

**In what country was your mother born?**

**Was your mother ever diagnosed with Parkinson's disease?**

☐Yes

How old was she when she was diagnosed?

☐No

☐Don't Know

**Did your mother ever have a problem with uncontrollable tremor or shaking when trying to do an activity?**

☐Yes

How old was she when she experienced these symptoms?

☐No

☐Don't Know

**Is your mother alive?**

☐Yes

Age

☐No

At what age did your mother pass away?

What was the cause of death?

**What was your mother's occupation most of her life?**

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Family Health History**:

**Do you have any blood-related brothers or sisters (living or deceased)?**

☐Yes

How many brothers or sisters do you have?

Age/gender of each

Have any of your brothers or sisters been diagnosed with Parkinson's disease (which one)?

☐Yes (age diagnosed/gender)

☐No

☐Don't Know

☐No

☐Don't Know

**Do you have any children?**

☐Yes

How many children do you have?

Age/gender of each

Have any of your children ever been diagnosed with Parkinson's disease?

☐Yes (age diagnosed/gender)

☐No

☐Don't Know

☐No

☐Don't Know

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**As far as you know, have any of your relatives ever been diagnosed with PD?**

☐Yes

How is this person related to you?

☐No

☐Don't Know

**As far as you know, do any of your relatives have a problem with uncontrollable tremor or shaking when they are trying to do an activity**?

☐Yes

How is this person related to you?

**As far as you know, has this person's tremor or shaking problem ever been diagnosed by a doctor?**

What was the medical diagnosis?

☐No

☐Don't Know

This completes our interview. Thank you for taking time to answer these questions. I will send you the results of the study as soon as it is completed. The results will be presented as a summary without any personal information about you. Do you have any comments you would like add?

# Appendix C

## Relative Accuracy of Case-only and Case-control Estimates of Gene-Environment Interactions



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

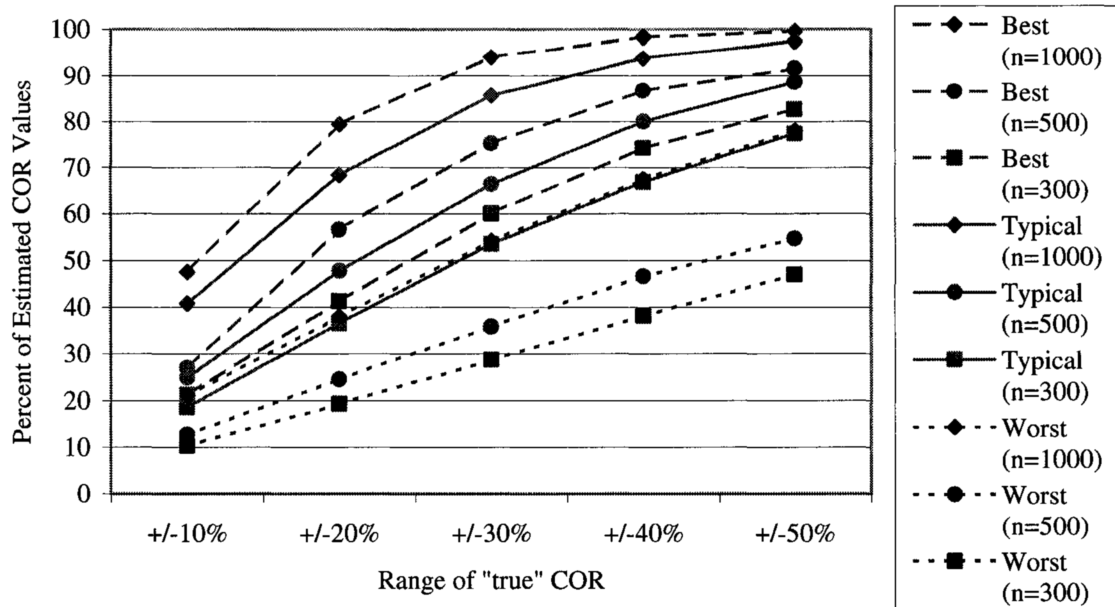



**Figure C-1. Percentage of estimated COR values within given range of "true" COR (COR=3) in a case-control study given worst- (Population 1), typical- (Typical Population), and best- (Population 32) case scenarios. Graph legend indicates scenarios and sample sizes.**

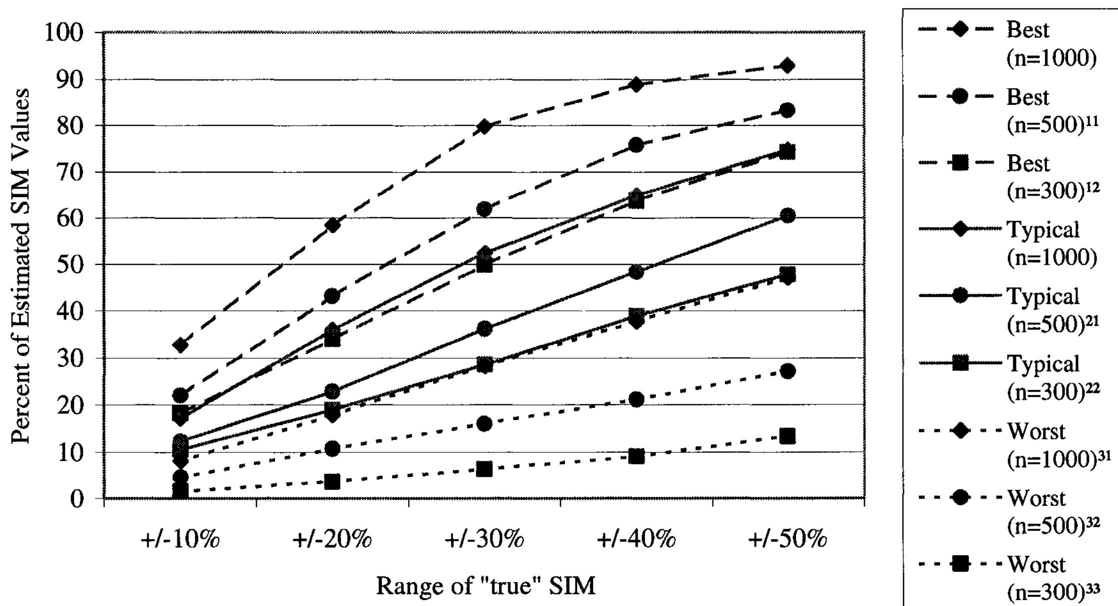



**Figure C-2. Percentage of estimated SIM values within given range of "true" SIM (SIM=3) in a case-control study given worst- (Population 1), typical- (Typical Population), and best- (Population 32) case scenarios. Percentages of replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [11]97.9%, [12]94.2%, [21]95.3%, [22]86.1%, [31]83.9%, [32]58.1%, [33]42.6%. Graph legend indicates scenarios and sample sizes.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

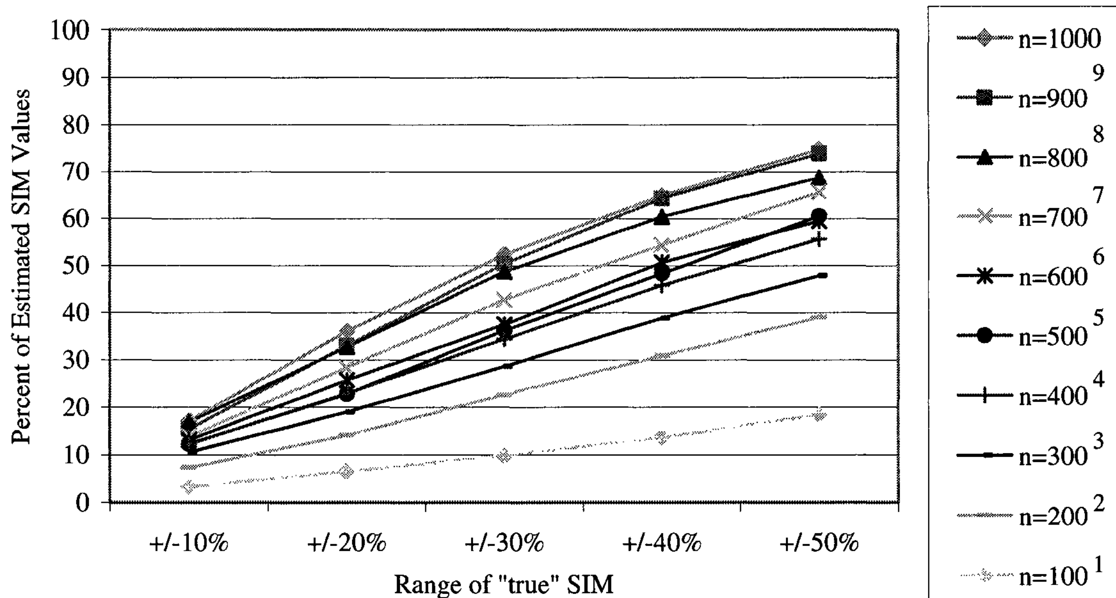



**Figure C-3. Typical Population (Interaction Effects)--Percentage of estimated SIM values within given range of "true" SIM (SIM=3) in a case-control study. Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [1]48.6%, [2]72.5%, [3]86.1%, [4]92.2%, [5]95.3%, [6]96.2%, [7]97.3%, [8]98.0%, [9]98.9%.**

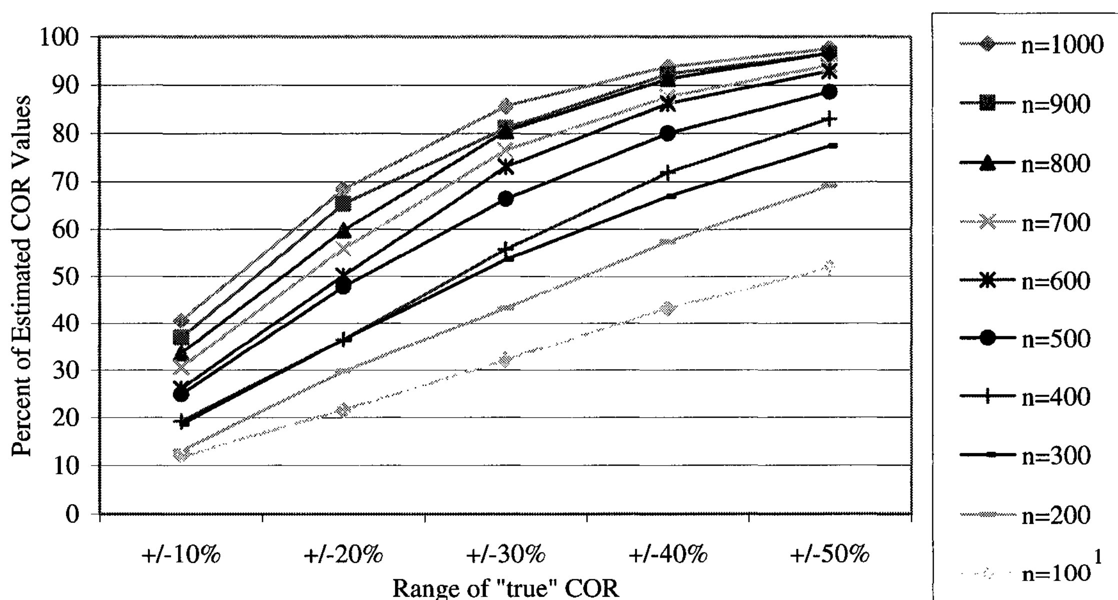



**Figure C-4. Typical Population (Interaction Effects)--Percentage of estimated COR values within given range of "true" COR (COR=3) in a case-only study. Percentage of replicates included (fitted logistic model converged or OR < 15) for each sample size is as follows: [1]97.5%.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

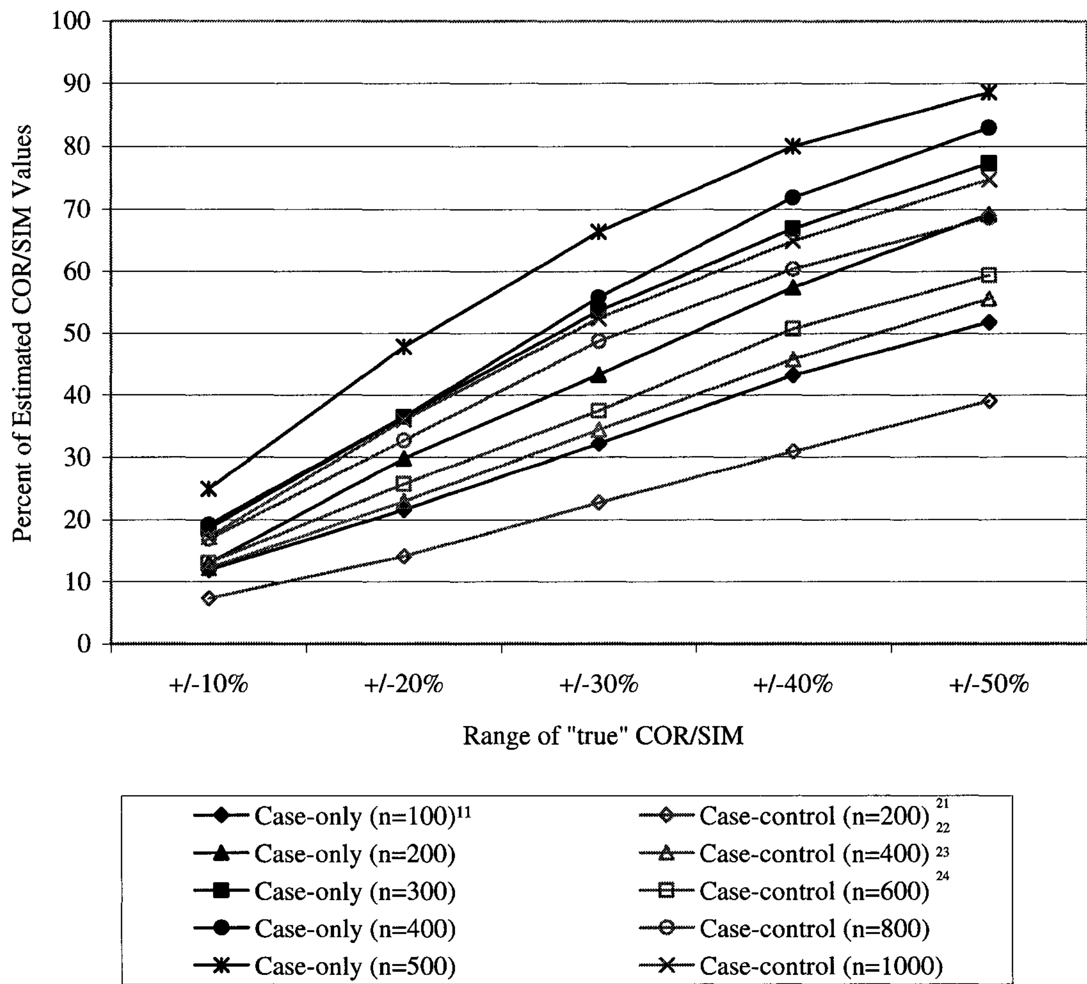



**Figure C-5. Typical Population (Interaction Effects)--Percentage of estimated COR/SIM values within given range of "true" COR/SIM (COR/SIM=3) in a case-only versus a case-control study given equal number of cases. Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [11]2.5%, [21]27.5%, [22]7.8%, [23]3.8%, [24]2.0%.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

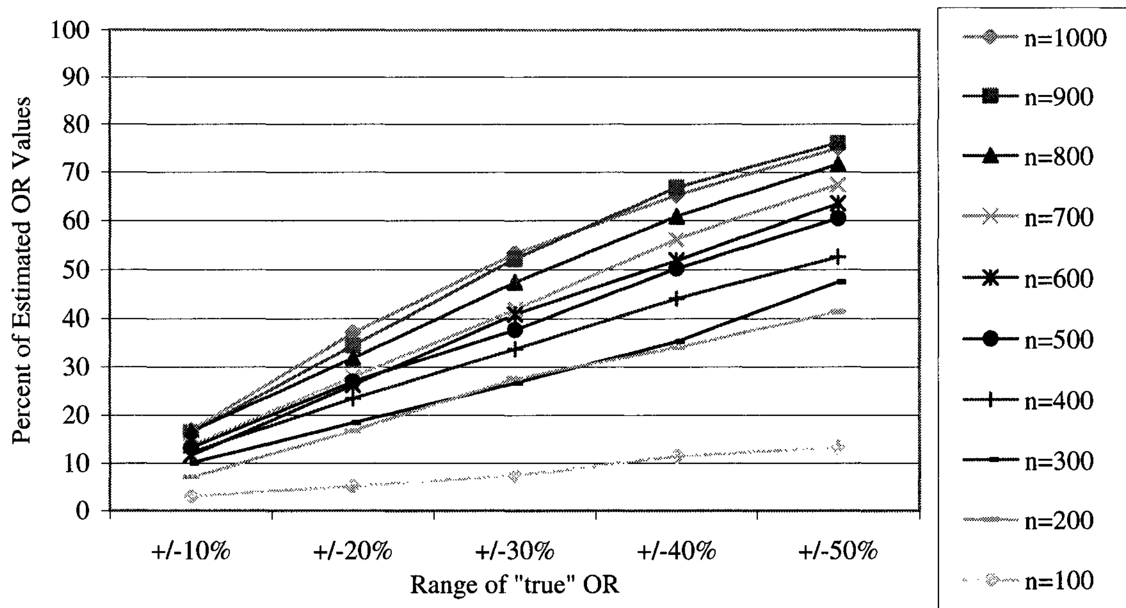



**Figure C-6. Typical Population (Independence Assumption)--Percentage of estimated OR values among controls within given range of "true" OR (OR=1) in a case-control study. Fitted logistic models converged or OR < 15 for all 1000 replicates for each sample size.**

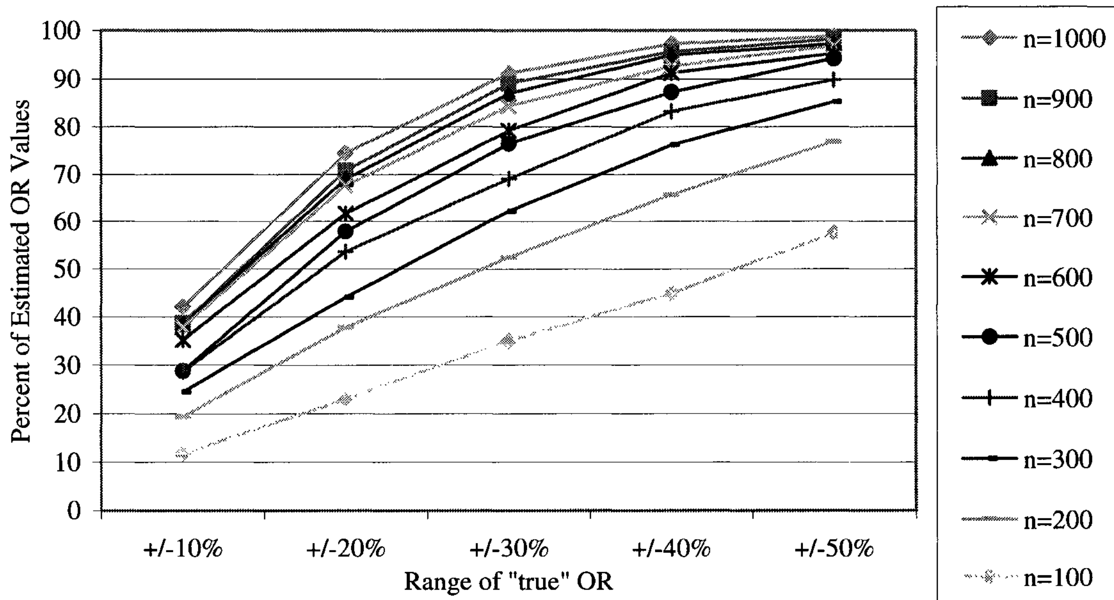



**Figure C-7. Typical Population (Type 2 GEI Assumption)--Percentage of estimated OR values within given range of "true" OR (OR=1) in a case-control study. Fitted logistic models converged or OR < 15 for all 1000 replicates for each sample size.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

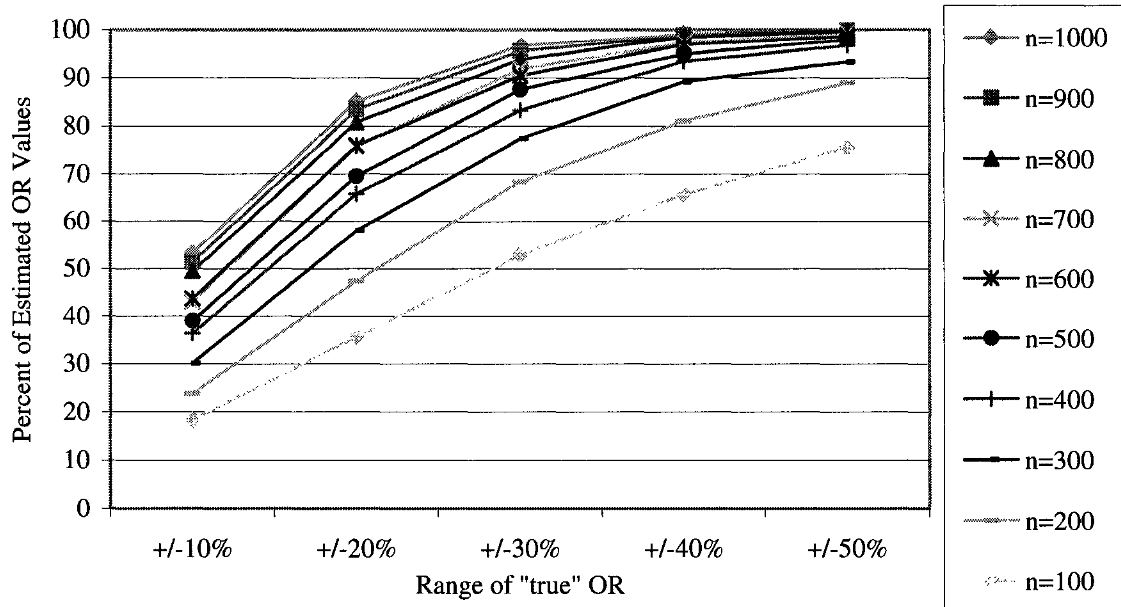



**Figure C-8. Typical Population (Effects of Exposure Alone)--Percentage of estimated OR values within given range of "true" OR (OR=2) in a case-control study. Fitted logistic model converged or OR < 15 for all 1000 replicates for each sample size.**

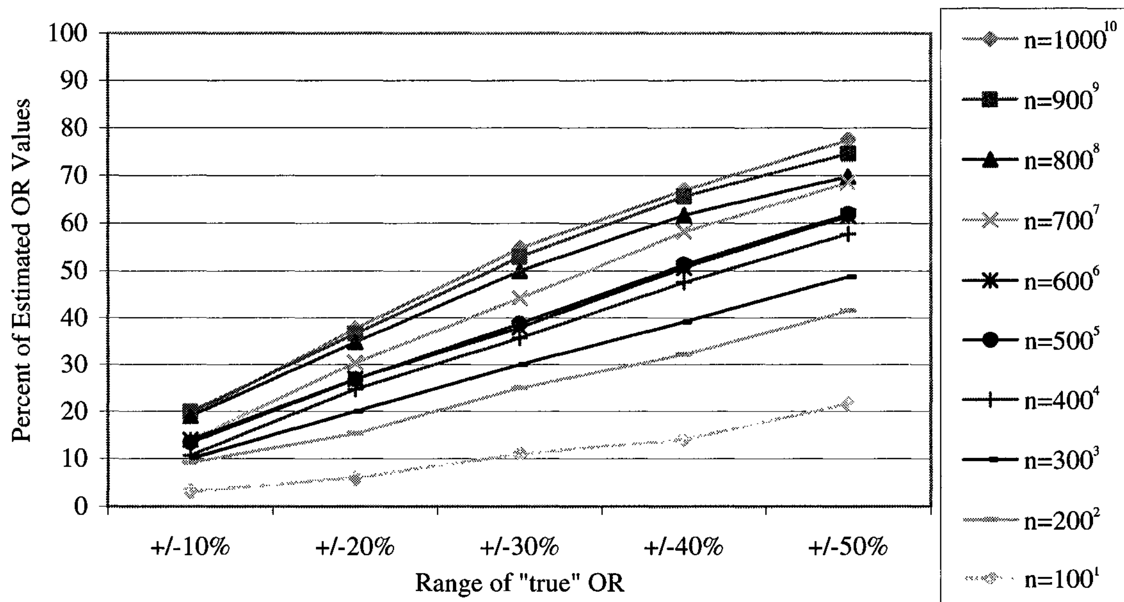



**Figure C-9. Typical Population (Multiplicative Interaction)--Percentage of estimated OR values within given range of "true" OR (OR=6) in a case-control study. Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [1]48.6%, [2]72.1%, [3]80.8%, [4]82.0%, [5]85.6%, [6]89.5%, [7]92.0%, [8]93.3%, [9]94.3%, [10]94.8%.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

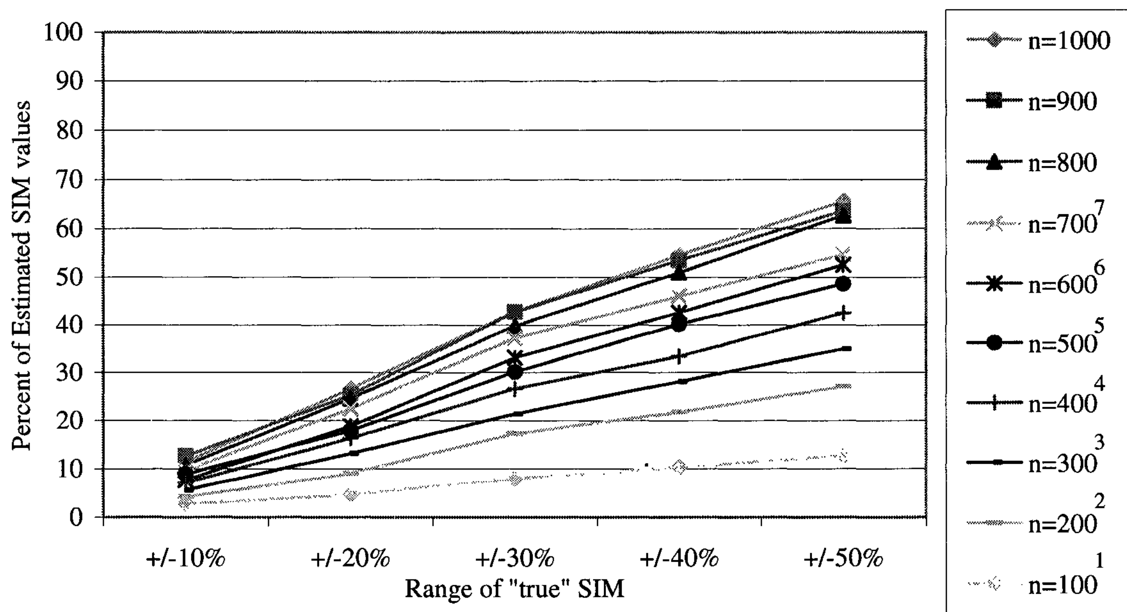



**Figure C-10. Population 33 (Interaction Effects)--Percentage of estimated SIM values within given range of "true" SIM (SIM=1.2) in a case-control study. Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [1]50.0%, [2]73.7%, [3]87.3%, [4]93.2%, [5]96.4%, [6]97.8%, [7]98.8%.**

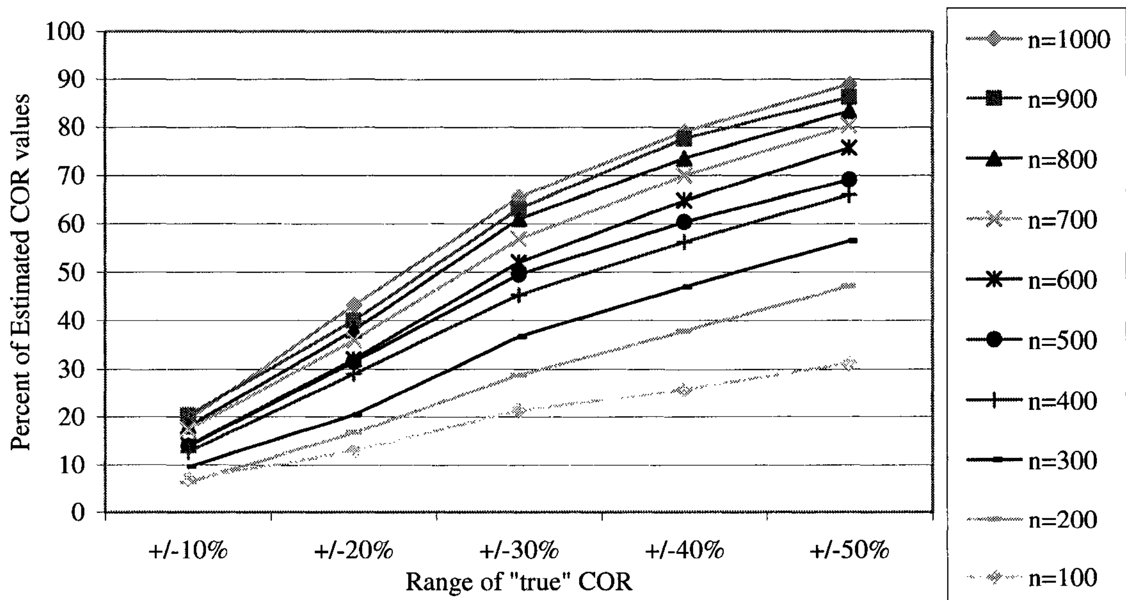



**Figure C-11. Population 33 (Interaction Effects)--Percentage of estimated COR values within given range of "true" COR (COR=1.2) in a case-only study. Fitted logistic models converged or OR < 15 for all replicates for each sample size.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

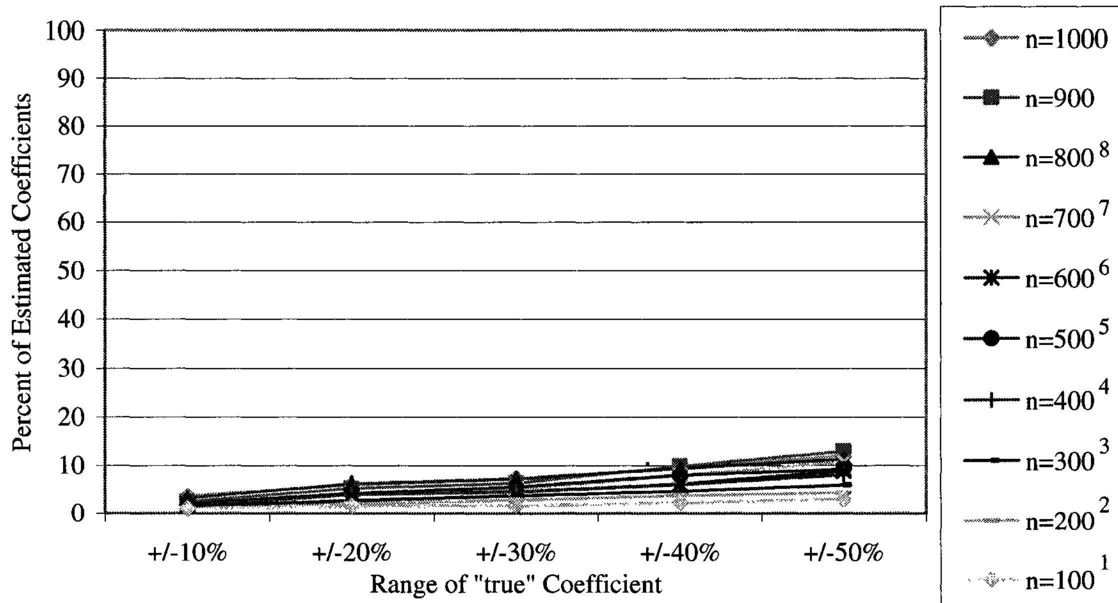



**Figure C-12. Population 33 (Interaction Effects)--Percentage of estimated coefficients within given range of "true" coefficient (beta=0.18) in a case-control study. Percentages of 1000 replicates included (fitted logistic model converged or -20 < beta < 20) for each sample size are as follows: [1]31.4%, [2]62.2%, [3]84.2%, [4]93.2%, [5]91.7%, [6]95.7%, [7]97.9%, [8]98.9%.**

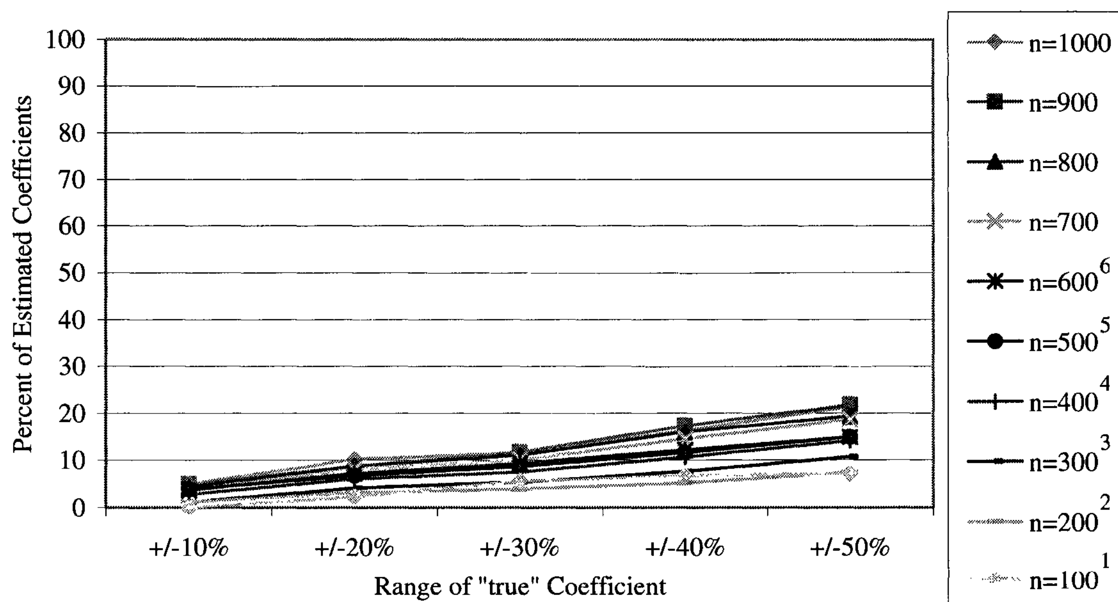



**Figure C-13. Population 33 (Interaction Effects)--Percentage of estimated coefficients within given range of "true" coefficient (beta=0.18) in a case-only study. Percentages of 1000 replicates included (fitted logistic model converged or -20 < beta < 20) for each sample size are as follows: [1]63.5%, [2]85.6%, [3]95.8%, [4]93.2%, [5]98.5%, [6]99.1%.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

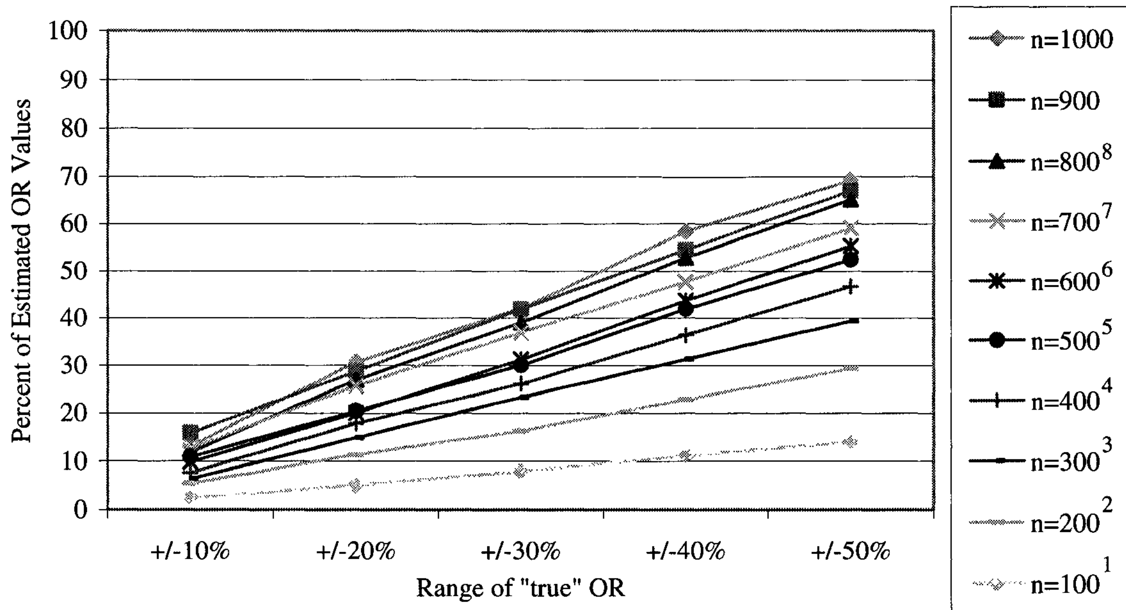



**Figure C-14. Population 33 (Multiplicative Interaction)--Percentage of estimated OR values within given range of "true" OR (OR=1.5) in a case-control study. Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [1]50.0%, [2]73.8%, [3]87.3%, [4]93.2%, [5]96.4%, [6]97.7%, [7]98.7%, [8]99.1%.**

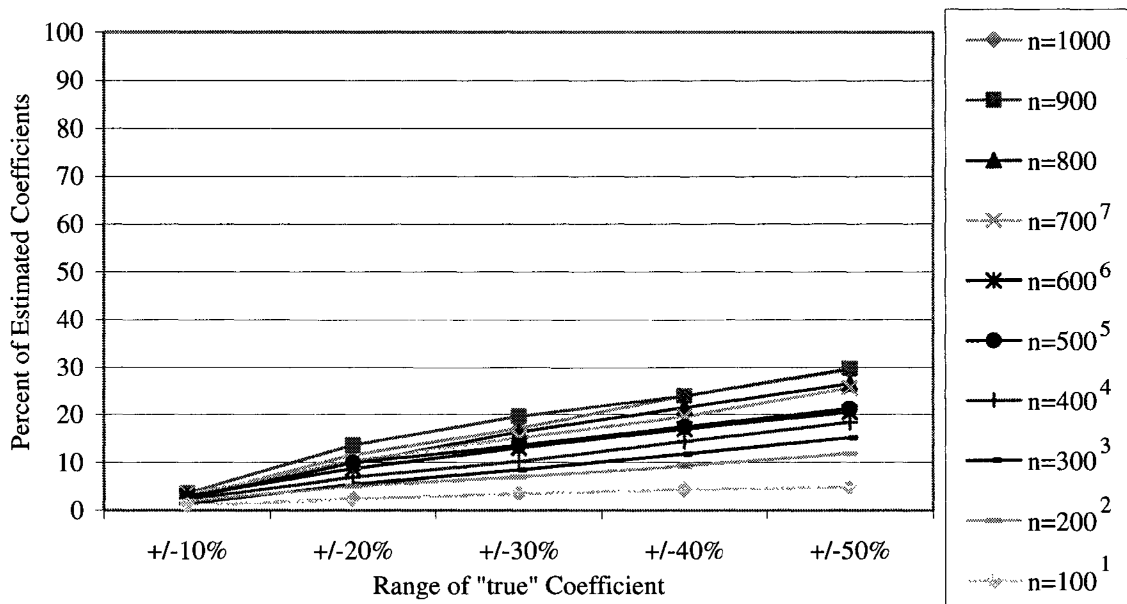



**Figure C-15. Population 33 (Multiplicative Interaction)--Percentage of estimated coefficients within given range of "true" coefficient (beta=0.41) in a case-control study. Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [1]31.2%, [2]62.2%, [3]84.2%, [4]91.7%, [5]95.6%, [6]97.9%, [7]98.9%.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

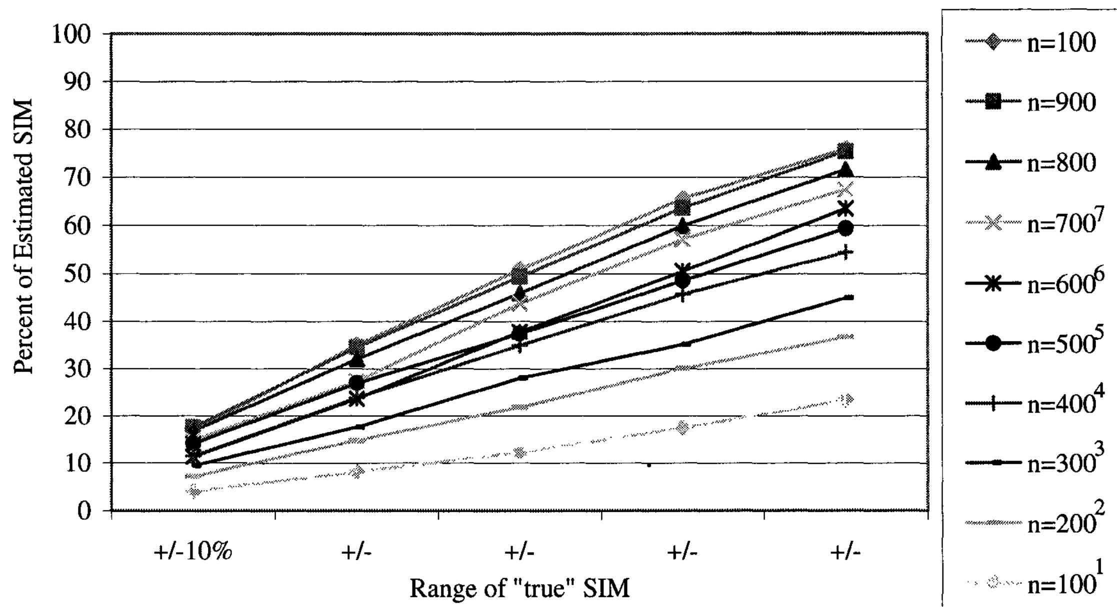



**Figure C-16. Population 34 (Interaction Effects)--Percentage of estimated SIM values within given range of "true" SIM (SIM=3) in a case-control study with independence assumption violated (OR among controls =1.5). Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [1]59.2%, [2]81.6%, [3]92.1%, [4]94.4%, [5]95.9%, [6]97.2%, [7]98.1%.**

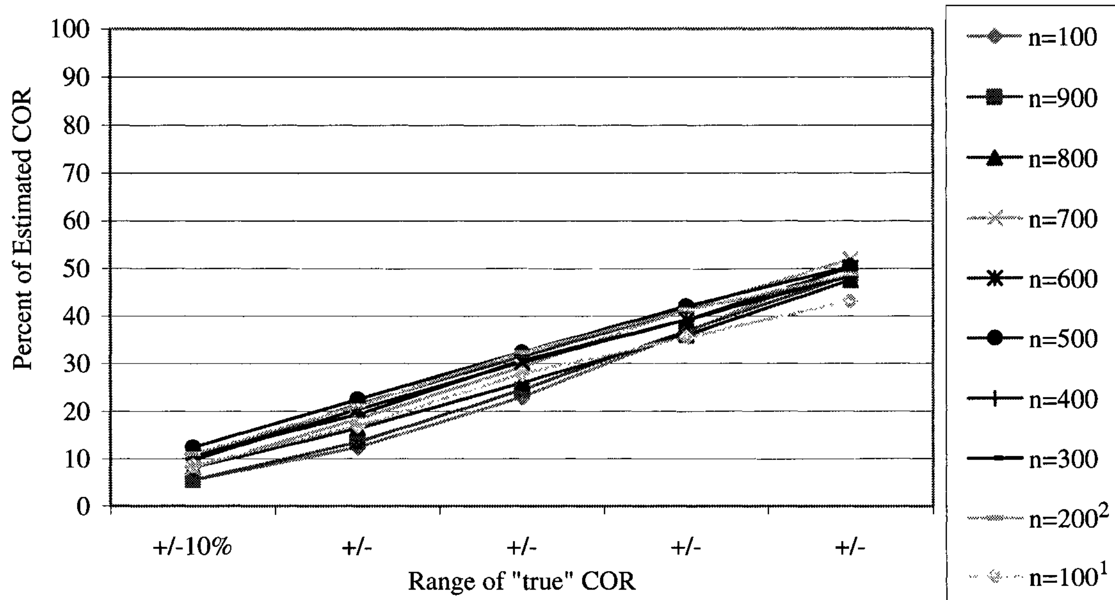



**Figure C-17. Population 34 (Interaction Effects)--Percentage of estimated COR values within given range of "true" COR (COR=3) in a case-only study with independence assumption violated (OR among controls =1.5). Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [1]95.0%, [2]98.6%.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

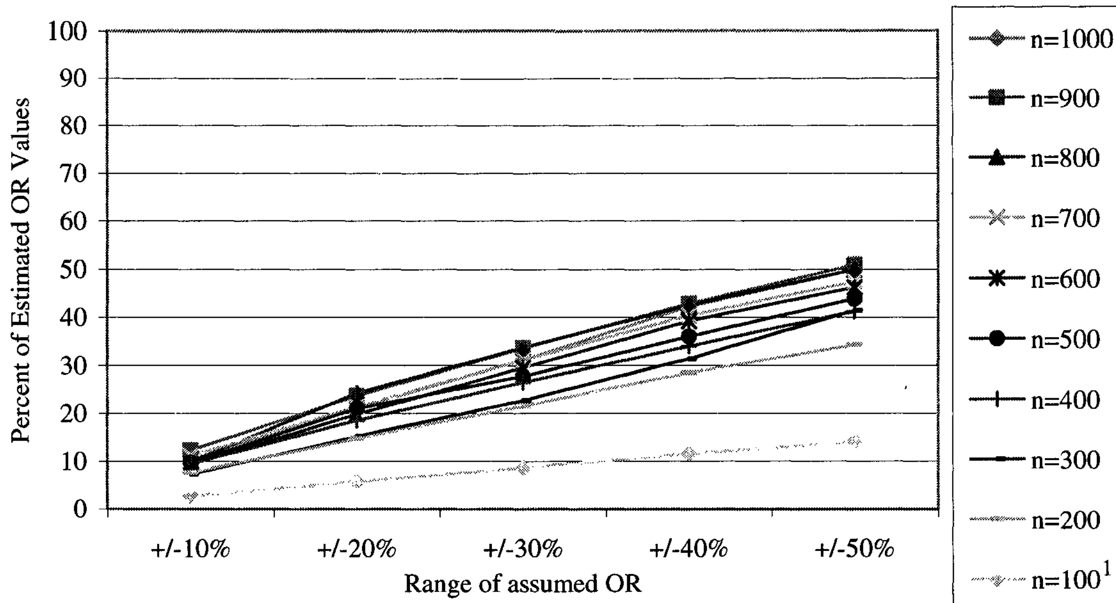



**Figure C-18. Population 34 (Independence assumption)--Percentage of estimated OR values within given range of assumed OR (OR=1) in a case-control study where "true" OR = 1.5. Percentage of 1000 replicates included (fitted logistic model converged or OR < 15) for sample size 100 is as follows: [1]98.9%.**

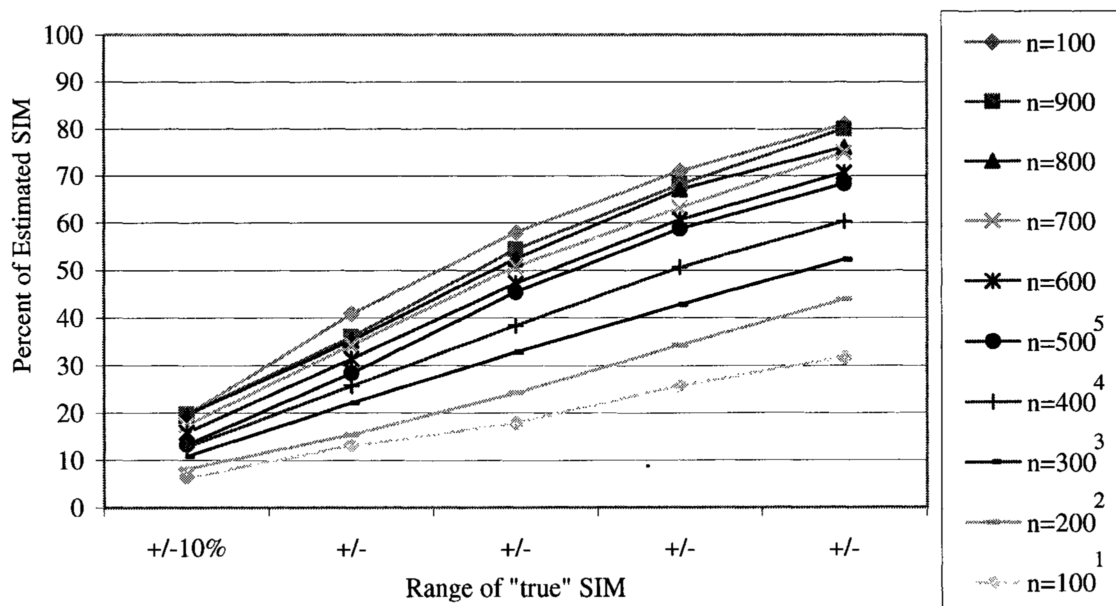



**Figure C-19. Population 35 (Interaction Effects)--Percentage of estimated SIM values within given range of "true" SIM (SIM=3) in a case-control study with independence assumption violated (OR among controls =3). Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [1]71.6%, [2]89.3%, [3]94.8%, [4]98.0%, [5]98.5%.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

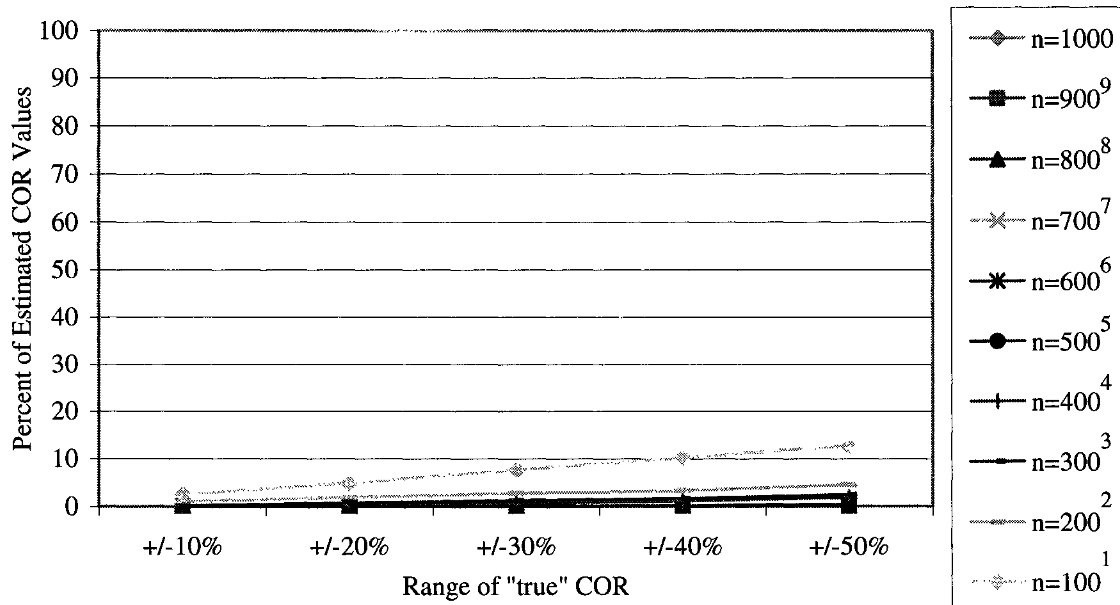



**Figure C-20. Population 35 (Interaction Effects)--Percentage of estimated COR values within given range of "true" COR (COR=3) in a case-only study with independence assumption violated (OR among controls =3). Percentages of 1000 replicates included (fitted logistic model converged or OR < 15) for each sample size are as follows: [1]73.5%, [2]83.1%, [3]88.6%, [4]92.4%, [5]96.6%, [6]97.6%, [7]98.2%, [8]98.7%, [9]98.9%.**

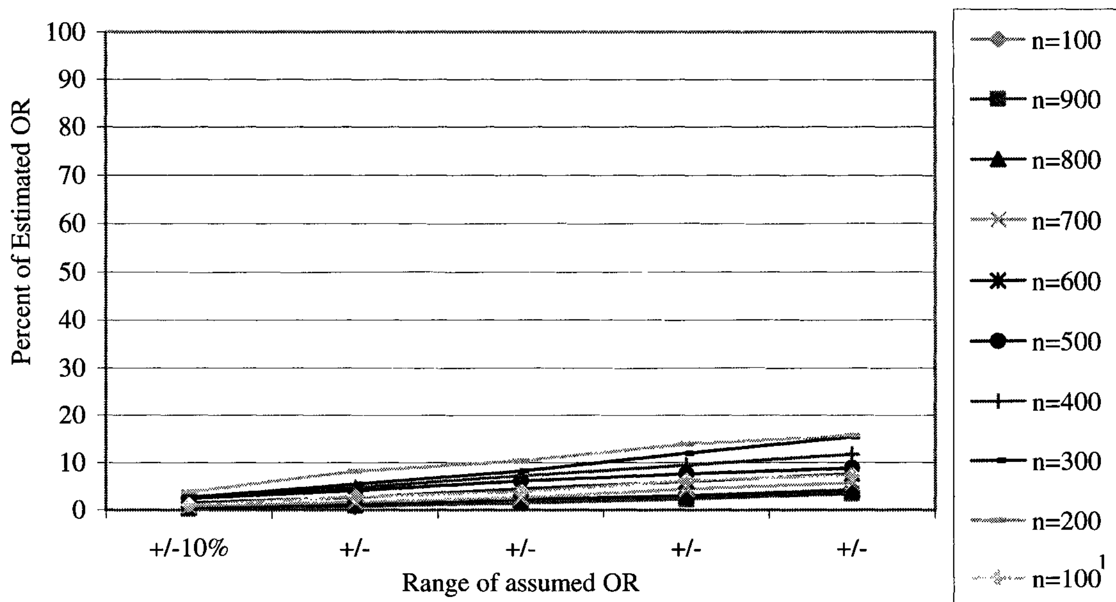



**Figure C-21. Population 35 (Independence assumption)--Percentage of estimated OR values within given range of assumed OR (OR=1) in a case-control study where "true" OR = 3. Percentage of 1000 replicates included (fitted logistic model converged or OR < 15) for sample size 100 is as follows: [1]94.5%.**



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.