

**THE ASSESMENT AND CONTROL OF NON-DIFFERENTIAL AND  
DIFFERENTIAL EXPOSURE MISCLASSIFICATION IN A CASE-CONTROL  
STUDY OF BREAST CANCER**

by

CHRISTINE VALERIE NEWBURN-COOK

M.S.N., The University of British Columbia, 1978

B.Sc.N., Queen's University, 1976

B.A.Sc., Queen's University, 1974

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(INTERDISCIPLINARY STUDIES)

We accept this thesis as conforming  
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 1995

© Christine Valerie Newburn-Cook, 1995

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at The University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

INTERDISCIPLINARY STUDIES: Health Care and Epidemiology,  
Medicine, Research Design and Methodology, Sociology  
and Statistics

The University of British Columbia  
2075 Wesbrook Place  
Vancouver, Canada  
V6T 1W5

Date: 7 December 1995

## ABSTRACT

This research study was designed to investigate some of the methodological issues involved in the design, conduct and analysis of a case-control study. The overall objective was to determine the reliability and validity of exposure data collected in a nested case-control study of breast cancer (N=1,177). The study was designed specifically to determine if the retrospective (post-diagnostic) reports of exposure provided by the cases and the controls were systematically different, to assess the impact of any resulting exposure misclassification on the estimates of relative risk, and (most importantly) to develop and to evaluate a '*Validity Scale*' as a possible design standard for the measurement and the statistical control of differential exposure misclassification in future case-control studies. To answer these questions, exposure information was collected prospectively and retrospectively by means of a self-administered questionnaire.

When the retrospective (post-diagnostic) and prospective (pre-diagnostic) exposure assessments were compared, the reported levels of exposure were assessed to be both reliable and consistent, although some inconsistencies (i.e., random exposure misclassification) were noted. The data provided no strong and conclusive evidence that the knowledge of diagnosis (i.e., case versus control status) resulted in the differential reporting of past exposure and antecedent events by the cases and the controls.

To determine the impact of exposure misclassification on the estimates of association, the prospective and retrospective odds ratios and their 95% confidence intervals were compared. Both the pre- and post-diagnostic odds ratios were found to be comparable. Therefore, the odds ratio (OR) estimates for the various study factors had not been biased towards or away from the null value (OR=1.00) by either the systematic overreporting or the underreporting of exposure by the cases and the controls. Furthermore, these data did not provide empirical evidence for the existence of either non-differential or differential exposure misclassification.

This was apparently the first study to explore directly the impact of different control groups on the estimates of association, and in particular, whether or not a particular control group would have a tendency to bias odds ratio estimates. Two control groups were recruited for the case-control comparisons -- *healthy controls* (i.e., women with a normal mammogram) and *anamnestically equivalent controls* (i.e., women with an abnormal mammogram but no breast cancer). Correlation, Kappa and McNemar analyses reported similar levels of agreement and inconsistency between the prospective and retrospective reports of exposure among the three study comparison groups. The results suggested that no advantage was obtained by using a control group which was anamnestically equivalent to the cases (except for diagnosis) that is, had experienced the same trauma (an abnormal mammogram), had experienced the same diagnostics to determine a diagnosis, and the same motivation to participate in the research study, and to report their past exposures both completely and reliably.

In addition, within the context of this research study, an '*Exposure Data Validity Scale*' as conceptualized by Raphael (1987) was designed, implemented and evaluated as a design strategy for the measurement and the control of differential exposure misclassification (i.e., recall bias). Overall, the validity scale appeared to be an effective means of assessing the propensity of the cases and the controls to report past exposures differently, and whether or not the estimates of effect have been subject to distortion (bias) as a result of differential exposure misclassification (i.e., recall bias). Replication studies will be required to determine both the utility and effectiveness of an '*Exposure Data Validity Scale*' as a design strategy to be included routinely in future case-control studies.

## TABLE OF CONTENTS

Chapter	Page
Abstract	ii
Table of Contents	iv
List of Tables and Figures	vii
List of Appendices	x
Acknowledgments	xi
 Chapter 1 - Introduction	
1.1    Epidemiology of Breast Cancer: The Magnitude of the Problem in Canada and British Columbia	1
1.2    Research Problem and the Significance of the Study	4
1.3    Overview of the Dissertation	10
 Tables for Chapter 1	13
 Chapter 2 - Background to the Research Problem	
2.1    Methodological Limitations of Case-Control Studies	18
2.2    The Measurement of Exposure: The Problem of Misclassification and Its Impact on Risk Estimates	22
2.3    Factors Affecting Exposure Reporting and Recall Accuracy: The Respondent as a Source of Measurement Error	27
2.3.1    Encoding, Retrieval and Judgment (Inferential) Errors: An Overview	30
2.3.2    Cognitive Perspectives on Coding and Retrieval of Autobiographical Memory	32
2.3.3    Impact of the Properties of Memory on the Retrieval of Autobiographical Facts	46
2.3.4    Judgment of the Appropriate Answer: Inferential Processing Strategies and Associated Errors	53
2.3.5    Respondent Rule Effects	63
2.3.6    Nonattitudes and Acquiescent Response Behavior	66
2.3.7    Sociodemographic Correlates of Respondent Error	69
2.4    Factors Affecting Exposure Reporting and Recall Accuracy: The Interviewer as a Source of Measurement Error	76

Chapter		Page
2.5	Factors Affecting Exposure Reporting and Recall Accuracy: Task Variables as a Source of Measurement Error	87
2.6	Review of the Literature: Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)	99
2.7	Raphael's Proposal for the Measurement and the Control of Recall Bias: The Development and Implementation of an 'Exposure Data Validity Scale'	101
2.8	Methodological Considerations for the Design of an 'Exposure Data Validity Scale'	104
2.9	Summary	105
Tables for Chapter 2		107

### Chapter 3 - Research Design and Methods

3.1	Study Objectives	151
3.2	Study Design	152
3.3	Subject Recruitment	153
3.4	Case Definition and Selection	156
3.5	Control Groups: Definition and Selection	158
3.6	Matching Criteria	161
3.7	Sample Size and Power Calculations	162
3.8	Data Collection Procedures	
	3.8.1 Questionnaire One	163
	3.8.2 Questionnaire Two	164
3.9	Procedure for Handling Non-Respondents	165
3.10	Data Handling and Analysis	166
	3.10.1 Data Coding and Entry	166
	3.10.2 Analysis of Non-Respondents	167
	3.10.3 Test-Retest Reliability Analysis	167
	3.10.4 Kappa Analysis	169
	3.10.5 McNemar Analysis	170
	3.10.6 Prospective and Retrospective Relative Risk Assessments	170
3.11	Stages in the Development of an 'Exposure Data Validity Scale'	172
	3.11.1 Search for the Candidate Variables	173
	3.11.2 Selection of the Validity Scale Exposure Variables and Assignment of Weighting Factors	174

Chapter	Page
3.11.3 Administration and Analysis of the Validity Scale	177
3.11.4 Evaluation of the 'Exposure Data Validity Scale'	178
3.12 Summary	179
 Chapter 4 - Results	
4.1 Response Rate	181
4.2 Description of the Study Groups	183
4.3 Analysis of the Non-Respondents	192
4.4 Test-Retest Reliability: Agreement Between the Prospective and the Retrospective Exposure Reports	195
4.5 Kappa Analysis	198
4.6 McNemar Analysis for Directional Discordance	200
4.7 Retrospective Versus Prospective Reports of Exposure: A Comparison of the Exposure-Disease Odds Ratio Estimates	204
4.8 The 'Exposure Data Validity Scale': Development and Analysis	209
4.9 Summary	228
 Tables and Figures for Chapter 4	232
 Chapter 5 - Discussion	
5.1 The Evidence for Non-Differential and Differential Exposure Misclassification	305
5.2 Evaluation of the Effectiveness of the 'Exposure Data Validity Scale'	313
5.3 Study Validity	319
5.3.1 Internal Validity	320
5.3.2 External Validity	324
5.4 Limitations of the Study	326
5.5 Recommendations for Future Research	328
5.6 Conclusions	330
 Table for Chapter 5	334
 Bibliography	336
 Appendices 1-9	370

## LIST OF TABLES AND FIGURES

Table 1:	Risk Factors for Breast Cancer
Table 2:	Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)
Table 3:	Estimated Response Rates for the Three Study Comparison Groups
Table 4:	Reasons Given for Non-Response
Table 5:	Comparison of the Study Population to the Screening Mammography First-time and Returning Participants Regarding Various Demographic Factors and the Risk Factor Profile
Table 6:	Demographic, Medical, Reproductive and Lifestyle Characteristics of the Study Groups
Table 7:	Demographic, Medical, Reproductive and Lifestyle Characteristics of the Study Respondents and Non-Respondents
Table 8:	Pearson Product Moment and Spearman Rank Correlations for Prospective and Retrospective Reports of Exposure for Study Variables and Conditions Possibly Related to Breast Cancer Risk - A Summary by Magnitude
Table 9:	The Agreement Between Prospective and Retrospective Reports of Exposure as Measured by Kappa Values - A Summary by Magnitude
Table 10.1:	A Comparison of Prospective and Retrospective Reports of Exposure by Breast Cancer <u>Cases</u> : Level of Agreement Between Reports (Kappa) and Directional Discordance (McNemar's Test)
Table 10.2:	A Comparison of Prospective and Retrospective Reports of Exposure by <u>Controls (Group One)</u> : Level of Agreement Between Reports (Kappa) and Directional Discordance (McNemar's Test)
Table 10.3:	A Comparison of Prospective and Retrospective Reports of Exposure by <u>Controls (Group Two)</u> : Level of Agreement Between Reports (Kappa) and Directional Discordance (McNemar's Test)
Table 11:	Exposure-Disease Odds Ratio Estimates for Prospective (Pre-Diagnostic) and Retrospective (Post-Diagnostic) Reports of Exposure for Study Factors Possibly Related to Breast Cancer Development

Table 12:	A Comparison of the Prospective and Retrospective Exposure-Disease Odds Ratios: Study Factors for Which the Odds Ratio Estimates Differed
Table 13:	Estimation of Exposure-Disease Odds Ratios for the Selected Exposures and Conditions Included in the Validity Scale
Table 14.1:	Exposure Data Validity Scale Analysis: The Determination of Study Group Differences for Individual Scale Exposure Variables and the Aggregate Validity Scale Summary Score (Part 1)
Table 14.2:	Exposure Data Validity Scale Analysis: The Determination of Study Group Differences for Individual Scale Exposure Variables and the Aggregate Validity Scale Summary Score (Part 2)
Table 14.3:	Exposure Data Validity Scale Analysis: Exposure Factors Identified by Breast Cancer Patients as Being Relevant For Disease Occurrence (Part 3)
Table 15:	Contingency Table of Subject Responses Classified by Risk Factor and Perceived Level of Risk (Etiologic Importance)
Table 16:	Definition of Risk Factor Codes
Table 17:	Response Frequency Expressed as Percentages of Marginal Row Totals
Table 18:	Analysis of the Frequency of Response Percentages: Factors Identified as "Plausible" Risk Factors for Breast Cancer Development Using the Subjective Inclusion Rule
Table 19:	Numerical Output from the Correspondence Analysis: The Principal Inertia (Eigenvalues) and Total Inertia, the Percentages of Inertia and Cumulative Percentages
Table 20:	The Analysis of Row and Column Coefficients: Absolute Contributions (CTR), Squared Correlations (COR), Distance of the Profiles from the Origin (FACT T), Profile Masses (MASS) and the Quality of the Representation of the Row and Column Profiles (QLT)
Table 21:	Definitions of the Column and Row Coefficients
Table 22:	Assessment of the Accuracy of the Two-Dimensional Graphical Representation: QLT Analysis
Table 23:	The Evaluation of the Accuracy of the Two-Dimensional Graphical Display: CTR (Contributions to Inertia) and COR (Contributions to the Principal Axis) Analyses

Table 24: A Comparison of the Prospective Risk Estimates of this Study with the Retrospective Estimates of Association Reported in Other Case-Control Studies

Figure 1: A One-Dimensional Display of Row (Risk Factor) and Column (Levels of Risk) Profiles

Figure 2.1: A Two-Dimensional Display of Row and Column Points: A Simultaneous Display

Figure 2.2: A Two-Dimensional Display of Row and Column Points: A Simultaneous Display Increased by Magnification

## **LIST OF APPENDICES**

- Appendix 1: Validity Scale Development - Selection of Exposure Variables and Weighting Factors: Letter of Information to Potential Subjects
- Appendix 2: The Validity Scale Questionnaire Used to Evaluate the Plausibility and Perceived Etiologic Importance of Several Exposure Variables and Conditions Being Considered for Possible Inclusion in Questionnaire Two
- Appendix 3: The Enrolment Screening Questionnaire Administered to Mammography Clients by the Screening Mammography Program of British Columbia. (This Questionnaire is referred to in this thesis as Questionnaire One (Q1). Pre-diagnostic exposure data were generated from responses to Q1 items).
- Appendix 4: Letter of Introduction Sent to Potential Subjects from the Executive Director of the Screening Mammography Program of BC. (This letter introduces the 'purpose' of the research study as well as the researcher who would be calling the potential subjects to determine their willingness to participate.
- Appendix 5: Letter of Information Accompanying the Study Questionnaire (Questionnaire Two)
- Appendix 6: The Study Questionnaire - Questionnaire Two (Q2). This Questionnaire was used to Collect Retrospective (Post-diagnostic) Reports of Exposure as well as Subject Responses to Included Validity Scale Items
- Appendix 7: Reminder Letter Sent to Non-Respondents
- Appendix 8: Codebook Developed for Questionnaire One and Two Data Processing
- Appendix 9: Power Calculations

## ACKNOWLEDGMENTS

I am grateful to the members of my thesis committee for their individual and collective contributions to the completion of my research and dissertation. I am indebted to my research supervisor, Dr. Robert Conry, who provided ongoing guidance, support and encouragement, as well as constructive criticism and suggestions during the final editing of the manuscript. I am grateful as well to Dr. Gregory Hislop, the Senior Epidemiologist at the British Columbia Cancer Agency, and also a member of my thesis committee, for assisting me to gain access to the Screening Mammography Program of British Columbia for the conduct of this study. He provided invaluable assistance to the development of the study protocol, and preparation of the study questionnaire, and important input regarding changes to the research design prior to data collection.

The other members of my thesis committee were Dr. Michael Schulzer, Dr. Martin Schechter, Dr. Nancy Waxler-Morrison and Dr. Gordon Page. I wish to express to them my sincere gratitude for participating on my thesis committee, for their interest in the study and also for providing me with useful comments and suggestions at the different stages of the dissertation.

I also thank Dr. Linda Warren and Ms. Lisa Kahn, the Executive Director and research coordinator respectively of the Screening Mammography Program of British Columbia for assisting me in the recruitment of the cases and controls, and for providing access to the SMP BC data which was required for this thesis project.

There are several other people who generously provided assistance with this thesis, and to whom I am indebted. Dr. Jonathan Berkowitz was an invaluable resource for numerous questions regarding data analysis. I would also like to thank Reverend Amethyst Campbell for her assistance in the coding of data and Mrs. Sherry Mihamoto for data entry. I especially wish to thank the generous women who participated in this study. As

well, I acknowledge the assistance of the University of British Columbia for financial support provided by means of a graduate fellowship.

My family and friends have provided continuous support and encouragement. In particular, I would like to thank Commander Robert Blakely who encouraged me at a crucial stage in this endeavour to rethink my priorities and to complete this dissertation. And last but not least, to my husband and '*best friend*', Brian Cook, I thank you for your love, your belief in me and my abilities, your constant support and encouragement, as well as your numerous editorial comments.

*To the Memory of My Grandmother - Mary Newburn-Redhead*

## Chapter 1

### INTRODUCTION

#### 1.1 Epidemiology of Breast Cancer: The Magnitude of the Problem in Canada and British Columbia

The incidence of breast cancer is rising around the world. It is a major public health problem for women in the more developed countries, especially in North America and western Europe (Miller and Bulbrook, 1986). Until approximately 1985, breast cancer was the principal cause of cancer deaths in Canadian women. However, while breast cancer mortality rates have remained stable over the past decade, mortality rates for lung cancer have increased rather dramatically. Consequently, lung cancer has surpassed breast cancer as the most frequent and fatal neoplasm in women in the industrialized western countries (National Cancer Institute of Canada, 1995). Furthermore, the National Cancer Institute of Canada (1995) projects that 5,800 Canadian women will die from lung cancer and 5,400 from breast cancer in 1995.

The 1995 age-standardized breast cancer mortality rates for Canada and British Columbia are 31 per 100,000 and 27 per 100,000 respectively. Therefore, it is estimated that of the 5,400 Canadian women who will die from breast cancer in 1995, 830 will be women in British Columbia (National Cancer Institute of Canada, 1995).

In Canada, breast, colorectal and lung cancers are responsible for at least 55% of the new cases of cancer in women. The age-standardized incidence rate (ASIR) for cancer of the female breast in Canadian women is 103 cases per 100,000, whereas the ASIR in British Columbia is 117 cases per 100,000. The BC rate is comparable to the Canadian age-standardized incidence rate. Given the national

and provincial ASIRs, it is estimated that in 1995, there will be 17,700 new cases of breast cancer diagnosed in Canada, and 2,600 women will be from British Columbia (National Cancer Institute of Canada, 1995).

In 1989, the number of new cases of breast cancer in Canada was 12,300. A comparison of 1989 and 1995 figures demonstrates clearly that breast cancer incidence continues to rise, and more new cases are expected each year. The average annual percent change in age-standardized breast cancer incidence from 1983-1990 in Canada has increased by 1.3%. Only the incidence of female lung cancer is rising faster, at a rate of 3.7% (National Cancer Institute of Canada, 1989; Friedenreich, 1990; National Cancer Institute of Canada, 1995). Miller and Bulbrook (1986) projected that if the incidence rates continue to increase in young women under the age of 50 years worldwide, "the number of breast cancer cases will increase from 541,000 (in 1975) to over 800,000 by the year 2000, and this figure could even exceed one million cases" (p.173). In addition, "over half of these cases will be diagnosed in countries where breast cancer is not currently the most frequent cancer in women" (p.173) -- in Asian countries including Japan and Singapore, central Europe, and some South American countries (Miller and Bulbrook, 1986, p.173; Friedenreich, p.5). Breast cancer is an epidemic which is responsible for more morbidity than any other disease (Papaioannou, 1974; Wallis, 1991).

The observed trends of a decrease in mortality, along with an increase in incidence of female breast cancer over the past decade may be related to several factors, including: earlier detection, mammographic examinations (used since the mid-1980s), more sensitive diagnostic techniques, and improvements in cancer registration (Wigle et al., 1986; Friedenreich, 1990; National Cancer Institute of Canada, 1995).

The National Cancer Institute of Canada (1995) noted that although cancer (for all sites) is primarily a disease of elderly Canadians, female breast cancer is "more frequent at earlier ages with about one third of all cases occurring in women aged 40-59 years, and another third in women aged 70 years and older" (p.38). In fact, breast cancer is the leading cause of death for women aged 35-50 years (Paffenbarger et al., 1980).

A Canadian woman has a 1 in 3 chance of developing cancer over the course of her lifetime. In contrast, the same woman has a 1 in 9 chance of developing breast cancer, and a 1 in 24 chance of dying from a breast neoplasm. Unfortunately, the probability of developing breast cancer during one's lifetime also continues to increase (National Cancer Institute of Canada, 1995, p.42). While breast cancer was the principal cause of premature mortality in women in 1989, lung cancer is replacing it as the leading cause of premature mortality (i.e., years of life lost before age 75) (Bisch et al., 1989; Friedenreich, 1990). However, together, they still pose a major health concern to women and to health care providers and policy-makers. According to the 1995 Canadian Cancer Statistics, the potential years of life lost due to breast cancer in 1992 was estimated at 95,000, which is equivalent to 21.6% of premature mortality from all causes. "Although more men than women die from cancer every year, women generally live longer than men and many of the cancer deaths among women occur at younger ages", as is the case with breast cancer. Consequently, the loss of potential years of life is slightly higher for women (National Cancer Institute of Canada, 1995, p.47).

The five-year survival rate for breast cancer is approximately 75%. Only 63% of the breast cancer patients are alive 10 or more years after diagnosis (Wallis, 1991).

Breast cancer remains a major public health concern, and is indeed an epidemic affecting women in Western industrial countries (Paffenbarger et al.,

1980) for several reasons: its morbidity, the years of life lost due to premature death, a rising incidence rate, relative ignorance regarding its etiology, and conflicting research reports on exposure-disease associations for the risk factors believed responsible for breast cancer development.

Multiple variables have been identified as potential risk factors for breast cancer. These factors, and their influence on risk, are identified and summarized in Table 1 (pp.13-17). It must be noted that very few variables have been identified with absolute certainty as risk factors for breast cancer development. A plethora of etiologic investigations (both case-control and cohort) has been conducted, which have produced conflicting rather than supporting evidence for these putative factors. Relative risk estimates for the exposure variables differ considerably from study to study. Governments continue to invest grant monies, and researchers continue to explore the question of breast cancer etiology, in order to get a better idea about what the precipitating factors are. Limitations in research methodology may be responsible for the many contradictions, and the significant lack of progress made regarding our understanding of the etiology and natural history of breast cancer.

## **1.2 The Research Problem and Significance of Study**

Considering the fact that the incidence of breast cancer in Canada and British Columbia is among the highest in the world, and continues to increase, there is a clear need to complete well-conducted, and methodologically sound etiologic studies to determine the factors responsible for disease occurrence, and when possible, to initiate prevention programs.

The case-control design is particularly well-suited for studying a disease like breast cancer which occurs many years after exposure to the suspected etiologic factors. Unlike cohort studies and clinical trials in which subjects are

followed in a forward direction from exposure to a particular outcome or disease, case-control studies begin with the recruitment of subjects who already have a particular outcome or disease (referred to as the 'cases') and those without the disease (the 'controls'). The two groups are then compared with respect to the prevalence of the exposures and antecedent conditions thought to be associated with the development of the outcome or disease under investigation.

By employing the case-control approach to diseases like breast cancer, which have a long latency, investigators can quickly and efficiently mount and conduct a study because they begin immediately to search for and to recruit women with breast cancer (i.e., the cases). Unlike the situation with cohort studies and clinical trials, relatively little money and effort are expended on the follow-up of subjects who remain free of disease; as well, "there is no need to wait for time to elapse between an exposure and the manifestation of disease" (Schlesselman, 1982, pp.18-19). In addition, Mantel (1973) indicates that comparatively fewer subjects are required in order to test for exposure-disease associations (i.e., there is no requirement to follow a large number of subjects to get sufficient numbers of individuals who develop the particular disease); and, more than one potential risk factor can be studied at the same time. Another advantage to a retrospective study is that the standard error of the odds ratio is smaller than that found in a prospective study, or a cross-sectional study of the same size (Fleiss, 1981). The stated advantages of case-control studies make them the preferred design for the study of rare diseases and those with a long incubation period (i.e., breast cancer).

However, the ability of a case-control study to generate valid estimates of association between the risk factor(s) and disease occurrence (i.e., exposure-disease odds ratio) depends on the capacity of the cases and the controls to provide complete and accurate personal histories regarding past events and the

exposures of interest. By definition, the exposure-disease odds ratio (OR) is the ratio of the odds of exposure among the cases to the odds of exposure among the controls (Last, 1995, p.118). For rare conditions (e.g., most cancers), the odds ratio provides a valid estimate of the relative risk (RR), which is a measure of the magnitude of the association between exposure and disease, and indicates the likelihood of developing the disease in those subjects who were exposed relative to those who were not exposed (Henneckens and Buring, 1987; Miettinen, 1976). If the  $OR=1.0$ , the ratio of the odds of exposure among the cases is equal to that among the controls, and there is no association between the exposure and the disease. An  $OR>1.0$  indicates a positive association, whereas an  $OR<1.0$  represents a negative association.

Methodologists challenge the credibility of case-control research, in particular the reliability and validity of exposure data, and consequently the study's findings with respect to the relationship between an exposure and disease (i.e., OR estimates) for several reasons:

- 1) its non-experimental approach, and the "backwards directionality" of reasoning from effect (disease outcome) to cause (risk factors) - (Rothman, 1986; Kramer, 1988);
- 2) the fact that subjects are requested to provide exposure information **after** they are aware of their disease status (Mackenzie, 1986; Friedenreich, 1990);
- 3) the observation that the compared groups (i.e., those subjects with (cases) and without the disease (controls)) are selected from two separate populations; the researcher cannot thus be confident that the groups are similar with respect to extraneous risk factors and other sources of distortion (Kramer, 1988; Kleinbaum et al., 1982); and,
- 4) the number, diversity and legitimacy of variables that have been proposed as possible contributors to inequivalent and faulty recall of exposure by the cases

and controls -- i.e., the salience and emotional impact of the outcome event, the presence or absence of disease, the length of recall, "telescoping", respondent characteristics (motivation, age, sex, education, socioeconomic status) the research design employed, the type of controls used (i.e., an anamnestic equivalent, diseased referent, population controls, proxy respondents), the method of data collection (i.e., interview, self-administered questionnaire), trait desirability, the need for social approval when the requested information is either sensitive or embarrassing, as well as time, memory and judgment factors.

Critics of the case-control method believe that this research design is susceptible to several methodological problems and hidden biases (Cole, 1979). Of particular concern is the possibility of exposure misclassification -- measurement errors which occur in the process of obtaining the required exposure information from the cases and the controls -- and specifically, the differential recall of exposure by the cases and the controls (i.e., recall bias). If differential exposure misclassification is present, the odds ratio may be biased in unpredictable ways (i.e., either the underestimation or overestimation of the association between a potential risk factor and the occurrence of disease). It is also conjectured that the fluctuations in the odds ratio estimates either towards or away from the null value (i.e.,  $OR=1.0$  -- which implies no risk or association between the risk factor and the disease outcome) may result in either type I or type II errors, and partially explain the problem of discrepant and contradictory results found among etiologic case-control and cohort studies on the same topic (Hayden et al., 1982; Morabia, 1990; Austin et al., 1994).

However, the empirical evidence for such **criticisms** of the case-control method may be unjustified. In view of the large number of case-control studies completed, it is notable that very few of them have evaluated both the reliability and validity of exposure information collected and the impact of exposure

misclassification on the estimates of effect (i.e., ORs). The few studies which do exist have not provided conclusive and convincing evidence that differential exposure misclassification is as significant a problem as it is thought to be. This lack of evidence does not convince the critics, often due to the inherent methodological limitations of the research itself (Mackenzie, 1986).

The warnings about case-control methodology persist, and confidence in the case-control results is cloaked in the fear that bias does exist, and has in fact invalidated study conclusions. These doubts will persist as long as researchers fail to prove the contrary -- directly within the context of case-control research.

Case-control methodology is nevertheless an important epidemiological tool for the investigation of cause-effect relationships in situations in which neither randomized clinical trials nor cohort studies can be performed. Therefore, it is necessary in the routine conduct of case-control studies for researchers to address formally the reliability and validity of their data and the exposure estimates, and to adjust the estimates of effect when distortion exists.

Therefore, the overall objective of this study is to determine the reliability and validity of exposure information collected in a case-control study of breast cancer, and the impact of any resulting exposure misclassification on the exposure-disease odds ratios for the various study factors. This research extends the work of Klemetti and Saxen (1967), Mackenzie (1986), Friedenreich (1990), and others who have investigated the nature and impact of differential misclassification (recall bias) in case-control studies by also investigating the presence and consequences of non-differential exposure misclassification (NDEM) on estimates of effect, as well as the use of an exposure data validity scale to measure and to control differential exposure misclassification (DEM).

Rothman (1986) noted that epidemiologists have generally found it more acceptable to underestimate than to overestimate effects. This may partly explain

the focusing of attention on differential exposure misclassification in substantive research. However, this investigator believes that NDEM, which occurs when exposure "misclassification is incorrect for an equal proportion of the cases and the controls" (Rothman, 1986, p.87) must also be considered in the conduct of etiologic case-control studies. Non-differential misclassification bias and its effect on the odds ratio estimates have not received enough attention from researchers and methodologists. NDEM must be systematically and empirically investigated in case-control studies. Its presence may obscure subtle but real risk effects (i.e., the occurrence of type II error), and account for not finding significant risk effects for factors which have biological plausibility and have been proven in animal studies. This researcher contends that data validity studies must consider both type I and type II errors -- that is, the declaration that an association is significant when it is not (a type I error), and the failure to find an effect when one exists (a type II error).

The significance of this study also rests in its attempts to design, implement and evaluate a validity scale as a design strategy for use in case-control studies to measure and to control for exposure misclassification. If successful, the validity scale concept could be adopted and modified for routine inclusion in a case-control study. Thus, researchers could (in every case) verify directly and empirically the epidemiological adequacy of the exposure data on which their research depends, or even show that exposure measurement error and/or bias exists: and if it were present, whether or not it poses a plausible threat to the study's conclusions. Information on the magnitude and the direction of exposure misclassification could then be used statistically to correct estimates of association (odds ratios), so that valid study findings could be generated -- conclusions in which both the researcher and the scientific community could have confidence. This will increase the design's credibility and enhance its acceptance as an

important research strategy for the study of chronic and rare disease etiology, while furthering our understanding of the natural history and the etiologic basis of disease. This study is also significant in its application of an infrequently used French multivariate statistical technique -- correspondence analysis (developed by the French analyst - Jean-Paul Benzécri in the early 1960s) -- for the construction of the 'exposure data validity scale', and specifically, the selection of the exposure items for inclusion in the scale and their relevant weighting factors.

### **1.3 Overview of the Dissertation**

This dissertation is divided into five chapters. Chapter One provides the general background to this study. It begins with a brief discussion of the epidemiology of breast cancer to demonstrate the magnitude of this problem, and its impact on women's health. The multiple risk factors identified in breast cancer etiology are then summarized, followed by comments about the lack of progress in understanding the natural history of breast cancer, and about the existence of conflicting evidence for the etiologic importance of the various risk factors; these are provided to substantiate the rationale for this study, and for the selection of the specific research questions. Finally, there is brief argument supporting the significance of this study for the evaluation and interpretation of previously conducted cancer case-control studies, as well as for the refinement and improvement of future etiologic studies.

Chapter Two discusses the limitations of case-control methodology, and examines the problem of exposure misclassification, with its impact on the estimation of relative risk. The review of the literature in Chapter Two deals specifically with two critical content areas:

- 1) the processes of human memory responsible for the encoding, retrieval and reporting of exposure and other health related events, as well as the factors which affect recall accuracy; and,
- 2) the substantive research which has examined the reliability of exposure data, and the presence, direction and magnitude of differential exposure misclassification (recall bias).

Chapter Three provides a detailed discussion of the research methods employed in this study to address the research questions. The areas discussed include: the choice of a nested case-control study design, the recruitment and the selection of the study groups, the use of multiple control groups and anamnestic equivalents, as well as the specific procedures used to collect and to analyze the data. This chapter also describes the stages in the construction of the 'exposure data validity scale' which was to be evaluated in this study as a possible design strategy to measure and to control for recall bias.

Chapter Four reports the findings of the study. It includes a description and comparison of the study groups, and an analysis of the non-respondents regarding demographic, medical, reproductive, exposure, anthropometric and lifestyle variables. The results of the data analytic procedures are presented to address several key issues: the reliability of prospective versus retrospective exposure data, whether or not differential exposure misclassification exists, the impact of exposure misclassification on the exposure-disease odds ratios, and the validity of study conclusions regarding the association of study factors to the occurrence of breast cancer. Chapter Four concludes with the results of the 'exposure data validity scale', and an evaluation of its effectiveness.

Chapter Five examines and interprets the results of this study with respect to the agreement between prospective and retrospective exposure data, and to the extent and impact of non-differential and differential exposure misclassification

on risk estimates. As well, it provides recommendations and conclusions regarding the use of a validity scale to measure exposure case-control differences in reporting accuracy. The validity and generalizability of the findings are discussed, along with the strengths and limitations of this study. The chapter concludes with important recommendations for future studies.

Table 1: Risk Factors for Breast Cancer

Risk Factor	Influence on Breast Cancer Risk	Comments
1. Female sex	Increase risk	Most important risk factor
2. Age at menarche (early vs late)	Inversely related to risk Protective - late menarche	Early menarche: menarche before age 12 has twice the risk relative to menarche after age 12  Late menarche - $\geq 15$ yrs  [Relative risk (RR) - 1.1 to 1.9] <sup>2</sup>
3. Age at menopause (late vs early)	Directly related to risk	Late menopause (after age 45) results in twice the risk relative to early menopause  [RR - 1.1 to 1.9] <sup>2</sup>
4. Age at First Full Term Pregnancy (late vs early)	Directly related to risk	Late first full term pregnancy (older than 30 years) or no pregnancy (nulliparous) results in 2- 3x the risk relative to early first pregnancy (i.e., younger than 22 years) - which appears to be protective  [RR - 2.2 to 4.0] <sup>2</sup>
5. Marital Status	High risk: never married Low risk: ever married	Probably related to childbearing practices [RR - 1.1 to 1.9] <sup>2</sup>
6. Country of Residence	High risk: industrialized countries (North America and Northern Europe)  Low risk: Asia, Africa	[RR > 4.0] <sup>2</sup>
7. Socioeconomic Status	High risk: upper class Low risk: lower class	[RR - 1.1 to 2.0] <sup>3</sup>

Risk Factor Data were taken from the following sources:

1. Vihko R, Apter D. Endogenous steroids in the pathophysiology of breast cancer. **Critical Reviews in Oncology/Hematology** 1989; 9: 1-16.
2. Kelsey JL, Hildreth NG. **Breast and Gynecologic Cancer Epidemiology**. Boca Raton, Florida: CRC Press, 1983.
3. Kelsey, JL. Breast Cancer Epidemiology: Summary and Future Directions. **Epidemiologic Reviews** 1993; 15: 256-263, p. 257.
4. Dunn B, Hislop TG, Anthony V. **Breast Cancer Risk and Prognosis** (Handout - BCCA).
5. Petrakis NL, Ernster VL, King MC. In **Cancer Epidemiology and Prevention** (D Schottenfeld and J Fraunrence, eds.). Philadelphia: WB Saunders, 1982, pp. 855-870.

Table 1 (continued): Risk Factors for Breast Cancer

Risk Factor	Influence on Breast Cancer Risk	Comments
8. Oophorectomy	Decreases risk	<p>Protection is inversely related to age at oophorectomy</p> <p>Low risk (protection) if surgery occurs before age 40 years</p> <p>[RR - 2.0 to 4.0]<sup>2</sup></p>
9. Ionizing radiation	Increases risk if the radiation to the chest is in moderate to high doses <sup>3</sup>	<p>Dose response is probably linear. Sensitivity varies with age</p> <p>The greatest sensitivity occurs during puberty and mammary development</p> <p>[RR - 2.0 to 4.0]<sup>2</sup></p>
10. Benign breast disease	Risk is increased with fibrocystic disease	<p>Proliferative lesions most important</p> <p>Women with benign breast disease may have increased risk up to 5x with demonstrated severe hyperplasia with atypia</p> <p>[RR - 2.0 to 4.0]<sup>2</sup></p>
11. Family history of breast cancer	Increases risk	<p>Having first degree relatives (mother, sister, grandmother) with breast cancer gives 2x to 3x the relative risk. Risk is increased (higher) if relative has had early breast cancer (premenopausal) and/or bilateral breast cancer</p> <p>[RR &gt; 4.0]<sup>2</sup></p>
12. Early abortion	Increases risk	Inconclusive evidence in research studies
13. Body build/ Weight/Obesity (Anthropometric)	<p>Obese vs normal weight (thin) may increase risk of postmenopausal cancer</p> <p>Breast cancer at ≥ 50 yrs - (high risk: obese; low risk: thin)</p> <p>Breast cancer at &lt; 50 years (high risk: thin; low risk: obese)</p>	<p>Concerns only postmenopausal women; may be opposite premenopausally. Weak evidence that premenopausal breast cancer related to excess body weight. Negatively associated with premenopausal cancer</p> <p>[RR - 2.0 to 4.0]<sup>2</sup></p>

Table 1 (continued): Risk Factors for Breast Cancer

Risk Factor	Influence on Breast Cancer Risk	Comments
14. Race/Ethnicity	High risk: white Low risk: black  Breast cancer at $\geq 45$ yrs: (high risk - white; low risk - Hispanic/Asian)  Breast cancer at $< 40$ yrs: (high risk - black; low risk - Hispanic/Asian)	This relationship exists for women over age 40 years. Under 40 years, black women have a higher risk for breast cancer  [RR - 1.1 to 1.9] <sup>2</sup>
15. Place of residence	High risk: urban Low risk: rural	Effect mediated through several other factors  [RR - 1.1 to 1.9] <sup>2</sup>
16. History of primary cancer in ovary or endometrium	High risk: yes Low risk: no	[RR - 2.0 to 4.0] <sup>2</sup>
17. Age	Increases risk  High risk: old Low risk: young	Effect starts at puberty and diminishes after menopause  Breast cancer is extremely uncommon in women under age 25. For women under 30, the risk is about 1%, but once women are in their 30s, they are in a 15% risk category. After age 40, women enter the period in which 85% of breast cancers occur <sup>4</sup>  [RR > 4.0] <sup>3</sup>
18. Lactation	High risk: no Low risk: yes (protective)	Research evidence is inconclusive. Very weak relationship, if any
19. Parity	Inversely related to risk	Effect of additional pregnancies small compared to age at first full term birth
20. Exogenous hormone use:  a. Estrogen replacement therapy (ERT)  b. Oral contraceptives if taken under age 45 years	Nil risk. Evidence inconclusive or controversial	The overall opinion is that estrogen used for contraception or for the treatment of menopausal symptoms is not associated with an increased risk of breast cancer. This is a controversial area and has proponents on both sides  Prolonged use of ERT may increase risk  Birth control pills taken after age 45 years may increase risk

Table 1 (continued): Risk Factors for Breast Cancer

Risk Factor	Influence on Breast Cancer Risk	Comments
21. Lifestyle factors:  a. Smoking  b. Alcohol Consumption  c. Emotional stress  d. Exercise	No overall association with breast cancer  Increased risk with alcohol consumption  Inconsistent results. Nil risk  Inconsistent results confounded by diet and obesity effects	Inconclusive and conflicting research evidence. Dose-response data are inconsistent
22. Diet	Fat: saturated fat positively associated with risk in postmenopausal patients  Protein: no association after adjustment for fat  Carbohydrates: no association after adjustment for fat  Total calories: no association after adjustment for fat  Fiber/Fruits/Vegetables: Negative association -- strongest in postmenopausal women	High fat diet probably increases risk
23. Atypical epithelial cells in nipple aspirate fluid <sup>3</sup>	Increases risk	High risk: cells present Low risk: no fluid produced  [RR > 4.0] <sup>3</sup>
24. Nodular densities on the mammogram <sup>3</sup>	Increases risk	High risk: densities occupying > 75% of breast volume  Low risk: parenchyma composed entirely of fat  [RR - 2.0 to 4.0] <sup>3</sup>

Table 1 (continued): Risk Factors for Breast Cancer

Risk Factor	Influence on Breast Cancer Risk	Comments
25. Hyperplastic epithelial cells without atypia in nipple aspirate fluid <sup>3</sup>	Increases risk	High risk: cells present Low risk: no fluid produced  [RR - 2.0 to 4.0] <sup>3</sup>
26. Religion	High risk: Jewish Low risk: Seventh-day Adventist, Mormon <sup>3</sup>	[RR - 1.1 to 2.0] <sup>3</sup>

## Chapter 2

### BACKGROUND TO THE RESEARCH PROBLEM

A case-control study is classified as a retrospective, observational and non-experimental research design, in which two groups are studied: one with a particular disease or outcome ('cases'), and the other without the disease or outcome ('controls' or 'referents'). The two study groups are then compared regarding the prevalence of existing prior exposures, characteristics and conditions hypothesized as putative (risk) factors for the outcome event. In other words, case-control methodology seeks to compare exposure frequencies between diseased and non-diseased study groups; and, the inclusion of one or more control groups provides an estimate of the frequency of exposure expected in subjects free of the disease (Schlesselman, 1982).

In a discussion of the historical development of case-control methodology, McFarlane et al. (1986) noted that this research design was developed by epidemiologists to examine cause-effect relationships in situations where experimental randomized clinical trials and cohort studies could not be conducted due to lack of feasibility, or to logistical, ethical, and cost limitations.

#### 2.1 Methodological Limitations of Case-Control Studies

There are several significant advantages to using the case-control design -- the simultaneous study of multiple risk factors; the relative simplicity of design, implementation and evaluation; its statistical efficiency (relatively few subjects required, no extended follow-up or loss of subjects-to-follow-up); its cost-efficiency; and, the lack of harm to subjects. Combining these virtues with the

development of, and improvements in, statistical procedures for the handling of case-control data has resulted in widespread use and increased acceptance of case-control methodology (Cole, 1979; Breslow, 1982; Schlesselman, 1982; Rothman, 1986). Sackett (1979) noted a fourfold increase in the number of case-control studies being completed and published in medical journals. In fact, case-control methodology remains the design of choice for the study of rare diseases (i.e., congenital anomalies) and chronic diseases such as cancer, where the incubation or latency period between exposure and disease outcome extends over many years (Mackenzie, 1986).

However, the disadvantages of using case-control methodology, and its associated deficiencies, have provoked intense scrutiny of this research design, dividing the research community into supporters and opponents of this investigative tool. The deficiencies and or limitations of case-control methodology include:

- 1) The possibility of unreliable and incomplete exposure data being collected. In a case-control study, the reliability and validity of exposure data are challenged because the researcher must rely on subjects' recall or on records for information on past exposure. Records are not always available equally for both the cases and controls, and exposure data may be missing or recorded in a format that is not useful in etiologic studies. Problems also exist because the exposure data are collected **after** diagnosis (group status) has been determined (Mackenzie, 1986).

Mackenzie (1986) and Friedenreich (1990) also note that cases and controls differ in many ways because of their disease experience; these differences may affect recall accuracy. The recall of past events and exposures is susceptible to substantial human errors (Kleinbaum et al., 1982; Cole, 1979). For example, diagnosis and treatment may impede the memory processes of case subjects, preventing complete and accurate recall of past exposures; the controls may be

less motivated to remember and to report exposures because the outcome event is not important to them; the cases (on the other hand) may be more motivated because of the salience of the outcome event (disease diagnosis), and the need to understand why this has happened to them, and what has caused their disease. There may be increased stimulation of 'search for cause' cognitive processes within the case subjects (Raphael, 1987; Raphael and Cloitre, 1994). Consequently, with the *causal search model*, one would expect that "causes or prior exposures connected plausibly with a disorder should be reported more completely by the cases than the controls" (Raphael, 1994, p.555). Furthermore, when subjects are aware of the exposure-disease associations being investigated, or that the exposure information being requested is either threatening or embarrassing to them, differences in recall may occur. Lastly, differences in hospitalization and diagnostics may help cases to remember and report antecedent exposures more completely and reliably because they have been prompted through increased questioning and examinations by several health care practitioners;

2) The validation of exposure information is difficult or sometimes impossible because case-control research must rely on either the recall of the cases, controls or proxy respondents, or exposure data which is recorded in hospital, physician or pharmacy records (Schlesselman, 1982);

3) The temporal direction of the investigation (directionality of inference testing) is from effect (disease) back to cause (risk factors). This 'backwards directionality' is the focal point of case-control controversy because it is the most significant methodological difference from the classical experimental approach, which investigates disease causation in a forward direction from cause to effect, through a process of deductive reasoning. It is assumed that under these circumstances, complete and accurate exposure histories for the cases and controls

cannot be acquired (Schneiderman and Levin, 1973; Rothman, 1986; Mackenzie, 1986; Kramer, 1988); and,

4) The fact that comparison groups are selected from two separate populations (i.e., those with the disease (cases), and those without (controls). In this case, the researcher cannot be certain that the two comparison groups are comparable regarding "extraneous risk factors and other sources of distortion" (confounders) (Kleinbaum et al., 1982, p.70).

Cole (1979) and Sackett (1979) conclude that the retrospective approach of case-control methodology, and the deficiencies previously discussed leave case-control studies subject to a wide range of sampling and measurement biases (including exposure misclassification and recall bias) and to methodological problems which may bias estimates of association between exposures and disease. If significant, these biases may create or obscure effect estimates, and consequently invalidate the study's findings.

Research into design deficiencies and possible biases is clearly justifiable given both the wide use and importance of case-control methodology in the study of rare and chronic diseases (including cancer), and its important limitations. Methodologists such as Feinstein (1979a), Cole (1979) and Ibrahim and Spitzer (1979) have called for the systematic and empirical investigation of case-control design and its validity as a paradigm for the determination of etiologic associations. Dorn (1959) summarized this need when he stated that there was an ongoing requirement for researchers to ensure study validity by improving upon the design, execution and analysis phases of case-control studies, and most importantly by providing strategies to assess for, to minimize or to control for the errors and biases that they are susceptible to.

Sackett (1979) observed that both case-control and cohort analytical studies are susceptible to bias, but of the two, the case-control design is both affected by

more sources of bias and less able to guard against them. He advocated that "the continued development and refinement of methodological standards for case-control studies becomes a high priority, especially in view of their increasingly frequent execution and appearance in the scientific literature" (p.59).

This dissertation seeks to do what Dorn (1959) and Sackett (1979) advocate: it addresses the appropriateness of case-control methodology for the etiologic investigation of breast cancer; it assesses the reliability and validity of retrospectively collected exposure data, and considers the nature and effect of any exposure misclassification on the estimates of effect. Finally, it proposes the design, implementation and evaluation of an 'exposure data validity scale' for the measurement and the control of differential exposure misclassification.

## 2.2 The Measurement of Exposure: The Problem of Misclassification and Its Impact on Risk Estimates

In a case-control study, once the cases and controls have been recruited, a primary task of the investigator is to collect exposure information from the study groups for comparison. Rothman (1986) noted that the collection of exposure information from the cases and the controls may be subject to error which results in "information bias", and the distortion (biasing) of the estimates of effect (p.84). Rothman also differentiated between two types of information bias -- non-differential and differential misclassification as well as their consequences on exposure-disease odds ratios. The basis for distinguishing between these two types of misclassification error according to Rothman (1986) is "whether the classification error on one axis of classification (either exposure or disease) is independent of the classification on the other axis. The existence of classification errors that are not independent of the other axis is referred to as **differential misclassification**, whereas the existence of classification errors for either exposure

or disease that are independent of the other axis is considered **non-differential misclassification**" (p.84).

Differential exposure misclassification bias is regarded as a major threat to the validity of case-control studies because it can result in either an exaggeration or an underestimation of an exposure-disease association. Differential exposure misclassification is sometimes referred to in epidemiological texts as information or response bias (Checkoway et al., 1989). As discussed in Section 2.1, there are multiple factors which may influence recall accuracy, and result in exposure misclassification by the cases and the controls. The differences in the completeness and accuracy of exposure histories provided by the cases and controls may be random, systematic, or both random and systematic.

Raphael (1987) distinguishes between simple memory failure in which recall is equivalently poor (non-differential) among cases and controls, and 'anamnestic inequivalence' (i.e., differential memory failure) in which the cases and controls differ with respect to the completeness and accuracy of exposure recall. Section 2.3 will detail the various factors that may account for 'anamnestic inequivalence' and the possible differences in the recall accuracy of the cases and the controls.

In the case of **non-differential exposure misclassification** (NDEM), the prevalence of exposure reporting errors (misclassification) is similar for the cases and the controls. Furthermore, exposure misclassification errors are independent of the case-control (disease) status of the study subjects. Both the cases and controls experience 'memory failure', and are unable to remember or report exposures either accurately or completely. This type of exposure misclassification results in measurement error and usually a loss of statistical power (Raphael, 1987, p.167). As a consequence of NDEM, the exposure-disease odds ratios are biased in a predictable way, towards the null value (i.e., an assessment of no association

between an exposure and the disease/outcome event) (Copeland et al., 1977). In other words, the presence of NDEM results in the weakening or the masking of an association, and a type II error. Methodologists have noted that non-differential exposure misclassification cannot obscure a relationship between exposure and disease, nor can it create a statistically significant association when none exists (Gullen et al., 1968; Marshall et al., 1981; Mackenzie, 1986; Rothman, 1986; Chu et al., 1989).

Conversely, **differential exposure misclassification** (DEM) due to recall bias occurs when the probability of exposure misclassification is different for the cases and the controls. The prevalence of false positive and false negative reports of exposure (patterned error misclassification) differs systematically by group (case vs control). DEM occurs as a result of differential memory failure (Raphael, 1987). Here, the systematic misclassification of exposure is related to outcome (disease status): the resulting bias leads either to an underestimation or overestimation of the strength of the association between the outcome and the hypothesized risk variables. In other words, DEM results in a systematic departure from the 'truth', and biases the odds ratios either towards or away from the null value, in an unpredictable manner (Checkoway et al., 1989; Coughlin, 1990). Several studies which have attempted to study the nature and effects of DEM have incorrectly conceptualized what **recall bias** is (i.e., assuming it to be the overreporting of exposure by cases and the underreporting of exposure by controls), and therefore have only looked for odds ratio distortions that are biased away from the null.

Any study which hopes to determine if differential exposure misclassification must be taken into account will have to determine both the prevalence of false positive and false negative reports which are a function of the sensitivity and specificity of exposure classification, and subsequently, whether or

not the exposure-disease odds ratios are biased towards or away from the null value.

The effect of non-differential exposure misclassification on estimates of effect has been discussed by Bross (1954) and Copeland et al. (1977). Rothman (1986) and Austin et al. (1994) commented on the fact that NDEM, and not just DEM (i.e., recall bias) is also a threat to study validity. Rothman (1986, p.86) stated that "non-differential misclassification has generally been considered less a threat to validity than differential exposure misclassification, since the bias introduced by non-differential misclassification is always in a predictable direction: toward the null condition [Bross, 1954; Copeland et al., 1977]". For example, if in reality there was a significant difference between the two comparison groups with respect to smoking, the existence of NDEM could result in a certain proportion of the truly exposed cases being misclassified as unexposed, while at the same time the same proportion of the truly unexposed controls could be misclassified as exposed. The result of the exposure misclassification in different directions, but in the same proportion for both the cases and controls could decrease the real differences between the two groups (cases and controls) regarding exposure to nicotine exposure through smoking, and result in an estimated  $OR=1.0$ , when in fact, the estimated  $OR>1.0$ . Here, the researcher would fail to find a significant association. This example demonstrates that NDEM may be responsible for a type II error -- a failure to detect a subtle and weak association between a risk factor (nicotine exposure) and the disease. Unfortunately, studies on exposure misclassification routinely exclude an examination of exposure data for NDEM. Rothman (1986) further observed that researchers are more concerned about erroneously claiming a significant association when one does not exist (i.e., type I error), than they are about underestimating an odds ratio, and failing to find a significant effect (i.e., type II error). He further noted (p.88) that NDEM will be

present in every epidemiological study, and that investigators should show greater concern for the consequences of NDEM, especially in studies that indicate no effect so that they may be able "to determine to what extent a real effect might have been obscured" (p.88). NDEM may result in the obscuring of real effects, especially if they are weak or subtle (Rothman, 1986; Austin et al., 1994). Elwood (1988) noted "that the greater the error the more 'noise' there is in the system, and therefore the more difficult it will be to detect a true difference in the factor being assessed between the groups being compared" (p.60). Consequently, it will be more difficult to detect a true difference in exposure prevalence and to find the cause of an effect.

When compared to DEM, NDEM is considered less serious because it masks marginal associations mainly, but major exposure-disease associations will be detected even in the presence of NDEM. It is also acknowledged that type I error (which is associated with DEM) is more serious than type II error (which is associated with NDEM). This explains why researchers sometimes concentrate exclusively on DEM and its effect on study conclusions. However, it goes without saying that researchers must ensure the reliability and validity of the exposure data, and consider the nature and impact of both NDEM and DEM, and then correct their estimates of effect for any resulting distortions due to exposure misclassification.

In summary, exposure misclassification is a potential problem in every study, regardless of design, but especially in case-control research. Therefore, it is very important to assess for its presence, to determine if it is non-differential or differential, and to estimate its magnitude and direction so that estimates of effect may be statistically adjusted.

### **2.3 Factors Affecting Exposure Reporting and Recall Accuracy: Respondent as a Source of Measurement Error**

The gathering of information from and about survey respondents through interviews and questionnaires has a long history and is widely used in both the social sciences and health care disciplines (Moss and Goldstein, 1975; Fienberg and Tanur, 1983). In fact, the sample survey has been described as the "single most important information gathering invention of the social sciences" (Adams et al., 1982, p.64). Many disciplines have come to depend on survey data for explanation of disease etiology, for evaluation of treatment protocols, for input to policy-making, for governmental and business administration, as well as for basic and applied social sciences research. However, from a methodological perspective, this research design is susceptible to problems of reliability and validity.

Because a significant proportion of survey data, both factual and attitudinal, is derived from self-reports, which ask respondents to recall past events or attitudes, researchers know that such reports may be highly inaccurate. They have identified several response bias variables as possible sources of invalid conclusions for studies which rely on retrospective, self-reported data. These include: trait desirability, the need for social approval, the salience and emotional impact of the event, the method of data collection, the questionnaire format and context, the interview situation, respondent motivation, as well as time, memory and judgment factors. Hauser (1969) emphasized the potential magnitude of the error associated with survey data when he questioned if "...more misinformation than information had been gathered on subjects by means of survey methodology".

Some methodologists believe that the response bias variables act as a form of systematic bias which significantly distorts the relationship observed between the independent and dependent variable(s). Several social scientists have realized

that although the literature indicates such distortion may occur, it has not been demonstrated conclusively (Gove and Geerken, 1977; Sudman and Bradburn, 1974). Sackett (1979) commented that methodologists must go beyond the mere cataloging of biases. He also stated that there was an urgent requirement for "the empiric elucidation of the dynamics and results of these biases. Methodologists have too long ignored their responsibility to measure the occurrence and magnitude of bias" (p.59). Many social scientists and health care researchers agree that there should be research on the biases themselves.

Accuracy of reporting, specifically the precision and validity of retrospective recall data, is a serious methodological concern. It affects the research endeavours in many disciplines including survey research, health care research, sociology, psychology, demography, market research and statistics. Although there is a body of methodological work assessing the validity and precision of survey methods, the resulting information has not been integrated across the disciplines. Researchers are not fully aware of what is available in subject areas other than their own. For instance, although an experimental psychologist may know very little about collecting retrospective data in surveys, her study of human memory has a direct bearing on survey data because the accuracy of such data relies crucially on a respondent's memory -- the retrieval and communication of recalled responses to questions posed. On the other hand, survey researchers must be scientifically self-critical, overcoming any complacency, such as assuming that if large numbers of people give apparently definite answers to a straightforward recall question, the combined result must be treated as valid and accurate (Moss and Goldstein, 1975).

The purpose of this section of the dissertation is to review the pertinent cognitive and social sciences literature with respect to the problems associated with the collection and interpretation of retrospective recall survey data. It will

attempt to summarize and integrate much of what is known about each of the error sources and response bias variables through the writings in three research domains: cognitive and experimental psychology, social psychology, and survey research methodology. It will first discuss relevant observations about human memory, using a schema of encoding, retrieval and judgment, as elaborated by such writers as Alba and Hasher (1983). Next, it shall review the literature of social psychology which deals with three distinctive sources of survey error -- the respondent, the interviewer and the "task" -- and the known response bias variables pertinent to these sources which affect the accuracy of reporting, after information has been retrieved from memory, and which may jeopardize the validity of study conclusions. In this latter regard the researcher is interested in the impact of threatening or embarrassing questions on the truthfulness of the respondents' answers, the characteristics and behavior of the interviewer and respondent, the questionnaire design, the interview situation, as well as the motivation of respondents to participate and provide precise answers.

Thirdly, the literature in survey research methodology delineates the many sources of response bias; as well, it identifies those respondent groups and tasks most susceptible to such errors.

The framework for discussion of response effects or errors is as follows:

- 1) the respondent as a source of measurement error;
- 2) the interviewer as a source of measurement error; and,
- 3) the task variables associated with measurement error ("task variables" refer to the conditions under which the required information is given by the respondent to the interviewer). This framework is derived from the work of Groves (1989) and from Sudman and Bradburn (1974).

Of the three sources of response effects, Sudman and Bradburn (1974) believe that the "task variables" are the most important source of response effects

in survey data. In epidemiology, however, respondent variables are considered to have a considerable effect upon recall accuracy. As Mackenzie notes (1986), "given the number and variety of influences on reporting (and their potential to interact with each other), it is not difficult to construct scenarios in which cases and referents **might** be differentially influenced, and might produce reports of differing validity. However, there is little experimental evidence to support or refute any of the possibilities" (p.14).

### **2.3.1 Encoding, Retrieval and Judgment (Inferential) Errors: An Overview**

Retrospective, non-experimental, survey research often questions respondents for both qualitative and quantitative facts about prior events, behaviors and attitudes. Examples include: How old were you when you first had intercourse? Age at menarche? When was the last time you visited your doctor for a PAP smear? Would you describe your childhood as happy? What was your maternal grandmother's death caused by?

Respondents' attributes and actions may affect the quality of the data collected. These questions do not just require simple recall of unambiguous facts from one's memory. They actually require sophisticated mental (information) processing. Due to limits on their ability to recall and enumerate specific autobiographical information, individuals often find these questions difficult to answer. It is a fact that respondents forget details associated with specific events, and often combine similar incidents into a single generalized memory (Linton, 1982). In these cases, respondents rely on inferences based on partial recall of information from memory to construct their answers (Bradburn et al., 1987).

Cognitive psychology can achieve insight into these sources of error. Specifically, how do subjects encode information into their memories; how do

they later retrieve it; and, how do they combine the recalled information into a single integrated response by inferential processing?

As Groves (1989) has observed, practitioners of the new cognitive science perspectives on survey response, such as Hastie and Carlston (1980), have identified five sources of survey response measurement error:

- 1) the respondent does not possess the knowledge required to respond to the survey questions (i.e., the information was never encoded, or the respondent has simply forgotten the details required);
- 2) the respondent does not engage in the appropriate cognitive activities (i.e., there is failure of the retrieval processes) at the time of response. Memory studies (Alba and Hasher, 1983) indicate that exact copies of personal events are never stored in memory. Furthermore, retrieval results in the recall of partial information, which can then be either reconstructed accurately or distorted and left incomplete;
- 3) the respondent does not understand the intended meaning of the survey question. Groves (1989) notes that the meanings of questions are not "fixed properties constant over all persons in a population" (p.419). The meaning assigned by the respondent to the question depends on the behaviour and characteristics of both the interviewer and respondent, consistent with the perspectives of symbolic interactionism (Stryker and Statham, 1985), on the context and form of the questions, and on the characteristics of the interview environment;
- 4) the respondent does not attend to the request for information and lacks the motivation to engage in the deep cognitive processing required to produce complete and accurate responses; and
- 5) the respondent does not communicate the appropriate response, once the information is retrieved. Many psychological processes affect what is actually

articulated in response to the survey questions. These include the perceived sensitivity or threat of the questions being asked, the social desirability of the respondent's answers, the perceived expectations of the interviewer, and the question itself in the context of the respondent's general knowledge and understanding of the purpose of the study.

The first source of error cannot be addressed by design strategies. Nothing can be done to overcome the loss of information which has never been retained, or subsequently has been forgotten. The researcher can only deal with the situation whereby the respondent has encoded and retained the required information, but is having some difficulty retrieving it. If the survey researcher understands cognitive theory about how knowledge and events are stored in memory -- by information processing -- she can construct survey procedures to access this information. Here, the emphasis is on the retrieval process and finding the cues (relevant schema) necessary to evoke the memory.

### **2.3.2 Cognitive Perspectives on Encoding and Retrieval of Autobiographical Memory**

Freeman et al. (1987) examined the matter of respondent accuracy of recall within the framework of the principles of memory organization in cognitive theory. They concluded that both forgetting and false recall are not random; they found systematic bias, which seems to lie somewhere in the cognitive processes of the respondent, somewhere within memory, between perception and recall. [The following discussion relied heavily upon the analysis in Freeman]

Contemporary cognitive theory provides no general overall explanation of the storage and retrieval of experience and information. Psychologists have come to accept that humans may be able to change memory storage and retrieval

strategies to suit differing demands. Nevertheless, a set of five fairly general principles has emerged in the research literature on memory:

1. Human memory is organized; humans create mental structures that impose patterns on information;
2. The organization in a mental structure is revealed in free recall. A categorical form of organization is usually imposed somewhere between stimulus and recall. Once a person has established a structure to organize a class of experiences, any new experience is then perceived and processed in terms of expectations imposed by that structure;
3. The organization of memory is based on experience. Mandler (1979) has said that "the mind creates order and structure out of a welter of stimulation, seeks for and finds regularities, and comes to expect them in the future" (p.260). But individuals vary in experiences. They differ in their exposure to, and knowledge about, the regularities exhibited in and among the elements found in a class of events;
4. The ability of a person to recall an element that occurred within an event depends on two factors: the amount of elaboration of his mental structure; and, the degree to which the element is "typical" in, and of, the event being examined.
5. The tendency of a person to falsely recall an element that did not actually occur depends on two factors: the amount of elaboration of the person's mental structure, and the degree to which the element is typical in events of the kind being studied. Although elaborate mental structures aid in recall, they may also have a cost in accuracy; what seems to happen is that with increased experience and increased mental structure, there occurs also an increasing tendency for "default" processing of those typical elements. People with well-developed mental structures will process incoming information about the typical elements of an event only superficially; their attention focuses on the untypical elements; they

will see what they expect to see of insignificant elements, based on prior experience. In such cases, a request for retrieval can not be met with genuine recall of the elements; rather, there will be a constructive process that taps into the general structure, as well as into specific memory. If the structure and event do not match exactly, false recall occurs. However, to the degree that the "normal" elements in an event are statistically typical, use of the model embedded in the cognitive structure as a substitute for actual perception will introduce very few errors.

Freeman describes this as the "organizational" view of human memory, which suggests a way to reconstruct event details from the data provided by informants. High knowledge respondents (those with well developed mental structures) forget little but they do create errors in recall by reporting typical elements which did not actually occur in the particular case. However, as their errors tend toward the long term pattern, their collective judgment about a particular event should provide the best possible index of that pattern.

The cognitive literature on memory emphasizes the fact that although one's memory of complex events and autobiographical information is sometimes very accurate, it is also frequently incomplete and highly distorted (Groves, 1989; Alba and Hasher, 1983).

### **Schema Theory**

The following discussion is found in Alba and Hasher (1983), who analyzed the strengths and weaknesses of the influential schema theory.

There is no single, well-accepted cognitive theory which provides a totally satisfactory explanation of such response measurement errors. However, "schema theories" (proposed originally by Bartlett, 1932) appear to give excellent insight into the nature of such errors, as well as to account for accuracy in recall.

Schema-guided encoding theory describes four central processes for the encoding of complex events/experiences -- selection, abstraction, interpretation and integration -- and one central retrieval process of reconstruction. These, taken together, can explain the inaccuracy, incompleteness or distortion of survey data. The theory posits that **incompleteness in recall** is attributed to the failure of retrieval processes, whereas **distortions** are due to associative encoding processes (Alba and Hasher, 1983).

By definition, 'schemata' are "sets of interrelated memories organized so that relationships are represented among attributes of events or pieces of information". A schema refers to the general knowledge that an individual possesses about a particular domain; and it is the vehicle which permits the encoding, storage and retrieval of information related to the specific domain (Groves, 1989, p.410). It is proposed that what is encoded or stored in memory is determined by a guiding schema or 'knowledge framework' that selects the central elements and modifies the experience in order to arrive at "a coherent, unified, expectation-conforming and knowledge-consistent representation" of an experience (Alba and Hasher, 1983, p.203).

The four central encoding processes (selection, abstraction, interpretation and integration) are responsible for schema-guided encoding of complex information, behaviour and attitudes. Explication of each of the four stages of the multistage encoding process follows to show how 'schema theory' accounts for the potential accuracy of memory, incompleteness in recall, and distortions in the data recalled and reported.

### **(1) Selection**

Only some of the incoming information of an event or experience will be encoded and stored as part of the memory representation of that event/experience.

Three factors will determine what information is selected:

- (1) the existence of a relevant schema or 'knowledge frame';
- (2) activation of that schema at the time of encoding; and,
- (3) the importance of the incoming information with respect to the activated schema.

The first condition requires the presence of existing relevant information. Prior knowledge, whether semantic or structural, increases the probability that new information will be encoded. In other words, specific domain-related prior knowledge is required for the acquisition of new domain-related information. The encoding of new information can be seen as a 'mapping process' of new information onto old; it depends on a sufficiently well-developed knowledge base or schema. The amount of new information which can be assimilated depends not only upon the amount of prior relevant domain-related knowledge, but also on the degree to which the incoming information matches the existing knowledge structure. In the absence of domain-related prior knowledge, there is no schema into which this new information can be readily integrated or subsumed. Thereupon the information is quickly lost or distorted. In this condition, memory is poor. For example, an end-stage kidney patient who has received dialysis three times per week over the past 5 years, and been subjected to frequent and diverse diagnostic workups, will be better able to assimilate information about a new diagnostic procedure than is a patient hospitalized for the first time for a diagnostic workup which includes this new procedure. The experienced kidney patient has a well-developed, extensive schema for such diagnostic procedures. He can more easily integrate information about a new procedure. It is not

unlikely that his recall of information about diagnostic tests, including the new one, will be more accurate than the information provided by the inexperienced subject.

The second factor emphasizes that the possession of prior domain-specific knowledge is not sufficient in itself to guarantee the encoding and storage of the new domain-related information. There is a further requirement that the relevant schema or knowledge frame be concurrently activated at the time of encoding. Experimental evidence shows the importance of schema activation during the encoding process.

Bransford and Johnson (1973) postulate that when knowledge structures are inactive during the encoding process, new knowledge cannot be integrated easily: the "absence of the appropriate semantic context can seriously affect the acquisition process" (p.397). Furthermore, Bransford and Nitsch (1978) speculate that less experienced subjects (e.g., the patient hospitalized for the first time for a diagnostic workup) will have greater difficulty than more experienced subjects (e.g., the chronic renal patient who has been subjected to frequent diagnostics) in determining the situational cues that can lead to the activation of an appropriate schema.

Anderson and Pichert (1978) provided experimental evidence that respondents preferentially recalled information and events which were congruent with their perspective at the time they were encoded; that is, ideas important to an activated schema are more likely to have a selection advantage for storage. The research was designed to study the independent effects of the nature of cues/schema on recall by using two distinct cueing strategies for the same text of a story. The results corroborated the hypothesis that recall of information was consistent with the perspective taken at the time the information was encoded. Analysis of the interview protocols suggested that the shift in perspective led

respondents to invoke a different schema which provided implicit cues for different categories of story information. In other words, the schema which was activated would determine the kinds of material recalled.

Anderson and Pichert (1978) concluded that these data clearly show that **retrieval** processes are independent of **encoding** processes, and that apparently forgotten material can be remembered through a shift in perspective. Also, different schemata operate at retrieval to influence what is recalled (Groves, 1989). Information irrelevant to the activated schema may never be permanently encoded; or it may be encoded but processed less elaborately than more relevant information (Anderson and Pichert, 1978; Alba and Hasher, 1983).

Even when information agrees with two different schemata, only one schema is activated during encoding; and recall is more accurate and complete for the information that is consistent with the activated schema (Alba and Hasher, 1983).

Tulving and Thomson (1973) stress the importance of the "encoding specificity principle" (p.353). The specific information stored is a function of what is perceived and how it is encoded. Further, what is stored determines the specific retrieval cues that can be effective in accessing the stored information. In other words, the memory representation and the properties of effective retrieval cues are determined by the specific encoding processes used on the incoming stimuli.

The implication of this research is that implementation of schema-driven questioning strategies (i.e., diverse recall cueing strategies) by survey researchers is important for eliciting responses relevant to the purpose of their studies. Tulving and Psotka (1971) showed that multiple cueing at the time of recall improved the quality of the material recalled. Therefore, when a survey is asking respondents to remember detailed or complex information, it would seem advantageous for the interviewer to ask the respondent to recall the required data

from different perspectives in an attempt to trigger different schema in order to build up a complete and accurate recall of the information being requested.

The experimental work of Richardson and Gropper (1964) also suggests that if respondents must recall complex details/events, they should be asked to recall the **same** information on **successive** attempts in order to improve the quality of recall. It has been shown experimentally in studies on 'bounded recall' that response is improved when subjects are interviewed on more than one occasion and are reminded of what was recalled in previous sessions (Neter and Waksberg, 1964b). These strategies may not always be feasible given an analysis of the cost-benefit ratio of additional interview time against the amount and precision of the additional information obtained. Schema theory suggests that research should be directed at learning which schema are used to organize the information sought.

The third factor which influences encoding in human information processing is the relative importance of the incoming stimuli in relation to the activated schema. Only important elements of the incoming stimulus are focused on for encoding. Because more attention is devoted to these elements than to less important ones, it is these same elements which are likely to be learned and represented accurately in the individual's memory. Traditional schema theory predicts that only the relevant information will be encoded; the remainder will be either rejected and lost, or distorted to fit the existing schema (Owens et al., 1979). This would account, in some measure, for the incomplete and inaccurate reporting of events.

Two selection principles have been postulated to account for the information that is selected and encoded. First, it is proposed that the ideas most important to the theme of the information, and which cannot be derived from previously encoded information, will be given special attention during encoding and subsequently will be recalled best (Owens et al., 1979; Spiro, 1980b).

The second selection principle is derived from script theory and the work of Schank and Abelson (1977) and is discussed in Alba and Hasher (1983). It focuses on how subjects process information related to frequently occurring events (e.g., visiting the doctor, eating out). A "script" is a temporarily ordered set of detailed memories containing the normal sequence of actions performed during the event under the usual circumstances (Groves, 1989, p.411). This means that not every detail of the experienced event is stored. Only the distinguishing features or 'atypical information' will be selected and encoded in memory. The theory predicts that the memory traces representing the highly typical events of the particular episode will be forgotten, or simply will not be encoded. These details can be derived by recalling that a scripted event occurred, and then by recalling highly probable elements from the "prototypical script" (Schank and Abelson, 1977).

As a result of the selection process, the memory trace of any event is likely to be incomplete. Thus, it is impossible for a respondent to reproduce a carbon copy of an event, even when under full motivation to do so (Alba and Hasher, 1983). Instead, she will try to reconstruct the event from the information which has been encoded and recalled. Bartlett (1932) stated that event reconstruction will consist of the recall of stored information plus "probable detail" from general schematic knowledge. Research suggests that reconstruction using script information (i.e., probable detail) can lead to poorer, imprecise reporting. For instance, Bower et al. (1979) found in a series of experiments with students concerning stories about routine activities (e.g., visiting a doctor or eating in a restaurant) that the respondents tended to recall attributes of the stories that were never communicated. They recalled and reported 'typical' actions that never happened. Instead, the 'scripts' or the respondent's prior personal knowledge regarding the activity (i.e., going to the doctor) provided additional details for the

memory trace of the particular story. In other words, subjects confused what was said with what the script strongly implied when remembering script-based texts. Distortions of the original event will occur when the probable detail generated during reconstruction is not actually part of the original event.

Alba and Hasher (1983) then note that Spiro (1977, 1980b) has showed that the reconstruction process is most likely to result in distorted/imprecise recall when the respondent encounters additional schema-relevant knowledge contradicting the encoded schema. Subsequent recall will depend on these two sources of information -- one correct, the other wrong. In this situation, the recalled information includes additional, incorrect information which is a byproduct of the reconstruction process that is attempting to resolve the inconsistencies between the two sources of information. Recall will contain additional information that was not a part of the original event.

In summary, selection process theory suggests that a significant proportion of the original event is not encoded nor represented in an individual's memory. Therefore, the selection process can account for incomplete respondent recall during survey research, whereas inaccuracy and distortion in recall can be attributed to failure of the reconstruction process which is believed to operate in retrieval (Alba and Hasher, 1983).

## **(2) Abstraction**

During the abstraction stage of encoding, the information selected by the activated schema is further reduced. Only the semantic content or meaning will be abstracted. As a result, the surface structure will be lost. Thus, only an abstracted memory trace of the original stimulus is stored. Because significant detail is lost during abstraction, this process can account for a respondent's incomplete recall and distortion of complex events/experiences.

The further concern, however, is to explain **accurate** recall. Alba and Hasher (1983), note that psycholinguistic research findings help schema theorists do this with the hypothesis that speakers of a language share preferred means of expressing information. If both speaker and listener have the same preferences or biases, the listener's reproduction may later seem to be accurate but is really only the imposition of the shared language structure. On the other hand, "distortions" would result from the abstracting process when the sender and receiver do not speak the same language, do not share the same biases. This distortion is a common occurrence in medicine. For a newly hospitalized patient with little previous health care experience, what he abstracts about the diagnostic workup may vary greatly from what the doctor intended him to understand -- even when the doctor has tried to forestall this problem through the use of careful, apparently non-medical terms and painstaking explanations.

However, distortion is better explained by two other schema-theory processes, interpretation and integration. Abstraction can be tangentially linked to recall distortion because it is a precondition for these last two stages in the encoding process (Alba and Hasher, 1983, p.209).

### **(3) Interpretation**

The discussion to this point has attributed the distortions in the recall of events/experiences to the encoding processes (selection and abstraction) which **reduce** the information encoded and stored. This loss of information is partially compensated at recall by reconstruction and the addition of 'probable detail'.

Distortions also occur because the semantic information encoded is in fact only an interpretation of the explicitly presented stimuli, albeit one that is consistent with the activated domain-specific schema. Distortions due to faulty interpretation are referred to as constructive errors because additional information

is added to the explicit information during or shortly after encoding. As a direct consequence of interpretation, there will be an elaboration of the memory trace of a complex experience/event.

Harris and Monaco (1978) note that respondents' interpretations are typically inferences of two general types. The first, 'pragmatic implication', involves transforming explicit information into its probable underlying intent. The second involves inferences made during comprehension when there is a need to (1) concretize vague information; (2) provide missing detail; and, (3) simplify complex and detailed information.

The possibility of distorting an original experience/event occurs during this stage of encoding because the respondent is able to add or change the information that is conveyed by the stimulus (Alba and Hasher, 1983). For example, when a patient is told by his doctor that he has a tumor which requires biopsy, he may later recall this discussion as his doctor actually telling him that he has cancer.

#### **(4) Integration**

The information which remains after selection, abstraction and interpretation will then be integrated with the previously acquired, related information activated during the current encoding episode. A single integrated memory representation is created. Individual detail exists only as a part of a complex semantic whole.

Integration processes occur either when a new schema is formed, or when an existing schema is modified. Gentner and Loftus (1979) give experimental evidence that once integration occurs and prior knowledge has been modified or updated, accurate recall of the original stimulus/information becomes highly unlikely. New information will be integrated into the old knowledge frames; thereafter, distinct traces of a 'to-be-remembered' event do not exist; only an

integrated memory trace remains. In the experiments of Gentner and Loftus (1979), subjects were shown either a film or slides of various traffic situations. Afterwards, a question was posed relative to the traffic scene that either implied the presence of additional information which never actually was present or which contradicted prior existing information. In subsequent memory tests, the subjects often misrecognized new slides containing the additional or contradictory information. The new and contradictory data replaced the individual's knowledge of the original traffic scene, resulting in a single, integrated memory of the scene. This distortion has received much research attention in relation to eyewitness testimony in court cases.

In summary, schema theory plays a very useful role in understanding human information processing. It provides a framework for understanding how information is encoded and stored in memory. It also permits an understanding of why recalled material can be accurate, or instead can be incomplete or distorted.

According to this theory only some highly selected subset of all the possible stimulus information is encoded; the selection is guided by the knowledge schema activated at the time of stimulus presentation. Memory traces are abstract representations of the original stimuli, and as such, make recall of the exact events/experiences impossible. Memory is an organized, hierarchical structure based on economy of storage, in which the semantic meaning of stimuli appears to have the highest priority for storage. Furthermore, memory is interpretive in that schema serve to add missing details or distort others so as to be schema-consistent. Memory is integrative in that the abstracted information is combined with prior knowledge, and any subsequent information, to create a single, unified memory trace of a detailed and complex event (Groves, 1989; Alba and Hasher, 1983, p.212).

Reconstruction of memory traces is the method by which subjects retrieve and recall event characteristics. The subjects add detail to the memory trace; they

will reconstruct the event using the probable details of the larger generic group of events of which the target event is a subset. The reconstruction process is a source of respondent error in the recall of event characteristics because there is a tendency to report event characteristics that are prototypical of those of the general class (Alba and Hasher, 1983; Yekovich and Thorndyke, 1981).

Recall accuracy depends upon many factors, including the elements of the original event selected for encoding and storage, the nature of the schematic connections established during encoding, respondent-interviewer biases in expressing information, differences in the number and timings of rehearsals (i.e., the number of times that the same information is requested and recalled by the respondent), the nature and number of cues/schema used to retrieve the requested information, the order in which elements are recalled, and chance matches between the reconstruction process and the original event/experience (Groves, 1989; Anderson and Pichert, 1978; Alba and Hasher, 1983).

For example, accurate recall is more likely for high probability events (i.e., high importance events/information) because these events develop more cognitive connections with other similar events, are referred to by others, and tend to be called into working memory more frequently. As a result of the rehearsal processes which occur in the working memory and the connections with other schema, important and highly probable events can more easily be retrieved by many diverse cues. The quality of recall can be good. In other words, these memory traces are more accessible in memory and are retrieved first. These events also serve to provide the probable detail in the reconstruction of the original event (Kintsch and Van Dijk, 1978; Black and Bower, 1979).

**Distortion** can be due to several factors including: the process by which the original event is reconstructed, the addition of information to the memory

representation of an event (constructive error), interpretations not actually intended, and the integration of memory traces over time (Alba and Hasher, 1983).

**Incompleteness** in recall is attributed to two encoding processes -- selection and abstraction: not all the elements of the original stimulus are selected for representation or become part of the abstracted meaning. Thus, incomplete details of the event are stored in memory. Because encoding involves the reduction of information for memory storage (economy of storage concept), the amount of information that can be recalled about the original event is reduced and incomplete.

### **2.3.3 Impact of the Properties of Memory on the Retrieval of Autobiographical Facts: A Review of the Literature**

Based on the preceding discussion, it will be recognized that responses to autobiographical questions could be accurate, but they could also be incomplete and distorted. Linton (1982) demonstrated that subjects may not be able to recall an event, even when they have numerous cues and the event is distinguishable from others. In this study of recall of personal events, 20% of the critical details -- selected at time of occurrence to be important and to be "certainly" remembered if the events were recognized -- were irretrievable after 12 months; 60% were irretrievable after 5 years.

In 1977 Cannell summarized the results of studies on interviewing methods which were designed to identify patterns of response bias, and to provide a basis for creating procedures to improve reporting. He recognized the inadequacies of the case-control method, and wrote that there was a need for improved data collection, as well as for research into the possibility of widespread

biases plaguing case-control design. One interpretation of reporting errors (over- and underreporting) is that such errors result from poor memory.

Cannell (1977) reviewed three theories. The "disuse" theory of Thorndike suggests that events from the more distant past are more likely to be forgotten than are more recent events, and that reporting errors arise from poor memory. Gestalt theory suggests that, generally, a respondent will forget events of low "salience" (or personal impact), especially with the passage of time; high salience events are better remembered. The "interferences" theory deals with the phenomenon of forgetting; forgetting does not occur absolutely. Information does not disappear completely from memory; rather, it may be more difficult to retrieve due to competing associations or interferences. Only the accessibility of information decreases, resulting in a lessening probability of recall from the storehouse of memory (Cannell, 1977).

This third theory suggests that underreporting is really a problem of retrieval, which can be alleviated by manipulating conditions which facilitate the recall of information. There are two critical steps for a respondent asked to report information from memory. First, she has to search for and retrieve it; secondly, she must transmit it to the researcher through a questionnaire or interview. Investigators have long been aware of the limited time span over which a person gives accurate reports. Some studies have showed a forgetting curve over time. However, the decrease in reporting may be due not so much to forgetting as to a tendency to misplace the event in time, and then recall it as being outside the reference period. Such misplacement may be only a minor factor, though. From the earliest memory studies, it has been recognized that the greater the impact of the event upon the person, the more readily the respondent recalls it. "Impact" generally refers to the personal importance of the event. Psychologically, this suggests that certain events occupy a greater part of one's psychic life.

Cannell (1977) also recognized that another factor affecting accuracy of reporting is the level of threat of embarrassment which the requested information holds for the respondent. Social psychological research has revealed the effectiveness of group norms in bringing about and maintaining approved behaviour. As well, one's self-image tends to censor communications. Research has shown a predictable and significant relationship between some characteristics of the information sought and the respondent's reporting behaviour.

Cannell (1977) concluded though that the characteristics of the respondent are not as consistent nor strong in their influence on underreporting as are the characteristics of the event. His general view is that research on improving reporting can best be devoted to the nature of the events and the factors underlying them, the most significant of which appear to be: elapsed time, impact, and the threat of embarrassment through revelations.

Cannell (1977) also concluded that research shows that the actual behaviour during an interview was the main variable that correlated with the index of reporting quality. This possibility is discussed in the later portions of this chapter dealing with "interviewer as a source of measurement error" and "task variables".

Coughlin, as well, has more recently reviewed survey research which has suggested that recall ability is related to the **salience** of an event; frequency, vividness, duration and meaningfulness of an event contribute to ease of recall (Coughlin, 1990). No consistent relationship has been found between accuracy of recall and demographic factors; this perhaps reflects the differences in study populations, the questions being asked, or the nature of the exposure under study. His overall thesis was that the factors which contribute to bias due to differential recall between cases and controls in retrospective studies have not yet been examined very thoroughly.

The National Centre for Health Statistics (1965) publication, 'Reporting of Hospitalization in the Health Interview Survey' (1965) focused on the underreporting of hospitalizations, and discussed hypotheses about the working of memory and about motivation as major variables possibly responsible for recall error. The survey accepted the validity of the two principles discussed above: memory is better for recent events than for those farther back in time; and, events having a greater impact on the person will be remembered better than those of minor impact. In general, there may be a "decaying" of experiences over time. However, to consider memories as fixed and "lifeless" is unrealistic. Memory is active, and dynamic, following patterns which can be predicted. Motivation is one of the most important forces in memory. Persons integrate events into their psychological life so that they fit most comfortably with past experience and with self-image. Numerous experiments testify to the selectivity and distortion which occur in the recollection of an event.

Decay of memory is modified by many factors, including the meaningfulness of the initial experience, the degree to which it was "learned", and the interference of other experiences. Motivation will affect how much effort a respondent will make to give an accurate report. To do this, a person must relive or review carefully his experience, constantly checking memory or using other aids such as reference to records. A more serious problem arises if the goals of the respondent are better served by inaccurate reporting, such as for embarrassing or socially undesirable behaviours.

Respondents may also have low motivation due to negative reactions to the study or its objectives. To participate in an interview requires the respondent to accept the goals of the survey, and to react positively to the interviewer. A negative reaction to either factor may be expected to result in inaccurate data.

The survey concluded that there is a strong relation between memory and motivation; they react dynamically. Findings included: the threat or embarrassment of a diagnosis starts a motivational pattern leading to suppression, and thus to underreporting of threatening episodes; the longer the hospital stay, and the more serious it is, the harder it is to forget; the elapsed time between the episode and the interview provides the opportunity for threat and decay factors to become effective. As time passes, perceptions are reshaped to fit one's total pattern of experiences. People remember selectively. It is found that the greatest underreporting is among episodes that provide the motive and opportunity for "forgetting". The survey found that re-interviewing elicited a sizable number of additional episodes remembered. The second visit provided additional stimulus to recall, and may also have increased the motivation to do so.

Wagenaar (1986) demonstrated that everyday personal events are forgotten very slowly, and that no event entirely disappears from memory. In addition, his work confirmed that the probability of recall depended on the number of retrieval cues used, as well as on the nature and the particular combination of these cues. A prompt about what occurred on a particular occasion (who was involved or where the event took place and/or the social occasion) improved recall. Asking for the date of an event is a poor cue to use if accuracy is the focus of the cueing strategy (Barsalou, 1987). Retention was significantly related to the perceived salience of the events, to their pleasantness, and to the degree of emotional involvement. The suppression of unpleasant memories was only significant for the shorter retention periods (p.249).

The work of Wagenaar (1986) and Bahrick (1983) demonstrated that the "forgetting function", the percentage of correct information recalled as a function of the retention period (time measured in years), depended on the nature of the material being queried.

Surveys requiring detailed and complex information -- such as number and duration of hospitalizations, personal expenses, drug consumption and cost, etc. -- often ask the subjects by advance letter to collect and review personal records, and then have them available during the interview. Other studies use recall aids (e.g., lists of products) at the time of the interview, thus enabling respondents to use recognition rather than recall as a strategy for reporting events and/or behavior. Aided recall appears to become more important as the length of the recall period increases.

These methods have opposite effects on memory errors. The use of records generally controls for overreporting due to telescoping errors, but has an insignificant effect on errors of omission because records were never meant to be 100% complete. Aided recall, on the other hand, tended to reduce the number of errors due to omission, but did not reduce (perhaps may even have increased) telescoping effects (Sudman and Bradburn, 1974, p.68). The use of records does not guarantee that respondents' reports will be accurate.

Williams and Hollan (1981) demonstrated that successive attempts at recalling a specific event/experience can result in additional recall of event characteristics. Furthermore, experiments on autobiographical memory have shown that respondents achieve better levels of recall if they are required to begin with the most recent event and then work backwards to the earliest occurring event. However, individual differences have been noted: some subjects prefer to recall in a forward direction; and often the direction of recall depends on the nature of the material being recalled (Whitten and Leonard, 1981; Williams and Hollan, 1981).

Groves and Kahn (1979) hypothesized that even though an event is never entirely forgotten, the effort to retrieve this information may be so onerous as to exceed the capacity of even the most attentive and motivated respondent. Reiser

et al. (1985) demonstrated that it takes on the order of several seconds to consider and retrieve specific information about even commonly occurring events (i.e., going to the barber, going for a walk or a visit to the dentist). This means that if a survey researcher asks too many questions within the limited time period during which respondents are motivated to participate and answer survey questions, the accuracy of the data will certainly diminish. Tourangeau et al. (1986) and Cannell et al. (1977) suggest design strategies which allocate more time per response and the use of longer questions. Accuracy will be increased because subjects have more time to use different retrieval strategies, more time to recall events, more time to consider their response, as well as more cues to stimulate recall.

Recall of autobiographical events is harder if memory contains many similar events. Initially distinguishable events can become confused; that is, the characteristics of the event become distorted, or the event itself becomes irretrievable due to interference from later events (Linton, 1982; Wagenaar, 1986).

Brown et al. (1987) note that events and personal experiences are organized temporally in discrete groupings of autobiographical sequences. For instance, a hospitalized patient may remember a visit to a doctor as part of an "extended causal sequence" beginning with the identification of a breast lump, making the initial doctor's appointment, then being referred for diagnostics, and ending up in hospital recuperating from a mastectomy. Evidence for 'autobiographical sequences' has been generated from studies on the effects of calendars on recall of personal events associated with work/school, as well as from 'free-recall' studies of personal events. The latter research demonstrates how subjects order events/experiences while reporting their recollections.

Taken together, these studies demonstrate the temporal organization of personal information. In addition, autobiographical sequences provide reference or anchoring points in time, that could be used as a design strategy for locating

other events in time. Autobiographical sequences provide a means by which subjects can organize their memories of personal events/experiences temporally. The survey researcher can use these sequences to counteract deficiencies in respondents' temporal inferences -- errors associated with estimations of the frequency or recency of an event.

#### **2.3.4 Judgment of the Appropriate Answer: Inferential Processing Strategies and Associated Errors**

Questions about the frequency or probability of occurrence of autobiographical events are difficult for a respondent to answer even when fully motivated to do so.

Cognitive research has shown that inference plays a vital role in what a respondent reports, and in the accuracy of response. In general, the respondent remembers a few facts relevant to a particular survey question, and then uses inductive inference to produce a reasonable answer. Inference, which adds detail to what the respondent can recall, can be inexact and misleading (Kahneman and Tversky, 1972). An understanding of the inferential processes used in answering survey questions and their cognitive limitations can aid survey researchers to implement design strategies which will improve the accuracy of recall.

As Groves states (1989, p.434), although there is no generally accepted theory on how humans make judgments, much cognitive research has been completed on how subjects form judgments about alternatives, and draw inferences from personal experiences (Kahneman and Tversky, 1972; Tversky and Kahneman, 1974; Nisbett and Ross, 1980). In surveys, subjects tend to put forth the minimal effort to meet the data requirements of the particular study. Krosnick and Alwin (1987) coined the term "cognitive misers" to refer to survey respondents

exhibiting this characteristic behavior in recall tasks. In general, respondents tend to avoid burdensome, intensive cognitive processing when forced to choose among alternatives. Rather, they use more readily accessible information about the alternatives to determine if sufficient discrimination can be made among them. Kahneman and Tversky (1972) note that in these situations, subjects are willing to accept the risk that they will be incorrect in return for the decreased effort and time needed to make their decisions.

Various "heuristics" (rules) have been proposed to explain the shortcuts taken by subjects to reduce cognitive processing (i.e., to reduce the required judgments to simpler ones) in decision-making, and in the evaluation of the frequency and probability of events. Although these heuristics are efficient, and in most cases yield judgments consistent with more intensive thought, they can also be sources of measurement error in survey research (Tversky and Kahneman, 1974).

The "availability heuristic" is often the first cited source of systematic bias in survey research. This inferential strategy refers to the tendency to choose as most important/recent/relevant that alternative which is most accessible in memory. It also influences the validity and consistency of judgments about frequency and probability. This judgment strategy relies on the quantity of information recalled, and the ease with which relative instances come to mind. The respondent is attracted to the accessible memory as an effort-saving strategy. Availability is considered a valid criterion for judging frequency because, in general, frequent events are easier to recall or imagine than infrequent ones (Tversky and Kahneman, 1973).

The most accessible memory is often the most 'vivid' (i.e., rich in detail and cognitive connections with other schema/events/ideas because it has been 'rehearsed' most often), the most recently accessed, and most connected to strong

emotions. Often the respondent takes this easy accessibility as a good indicator of relative importance and recency when a date must be put on an event.

Furthermore, a survey question, and the cognitive problem it poses, may resemble another one just performed (and thus "available"). Consequently the respondent may form a judgment using the procedural format just followed in the preceding task (Kahneman and Tversky, 1972; Tversky and Kahneman, 1974; Brown et al., 1985).

In many situations the availability heuristic works well. An individual may come to trust it in guiding his judgments. However, reliance on it can result in poor judgments and inaccurate responses when the most accessible, relevant memory is atypical of the respondent's experiences, and his answer is overly influenced by it. For instance, consider the survey question which asks respondents to determine the frequency of acute health problems in the past 12 months. If a respondent has recently experienced a bout of chest pain requiring emergency outpatient treatment, this easily recalled incident stimulates him to search his memory for other similar incidents in the requested reference period. To the extent that these episodes are seen as related events (i.e., similar in duration, diagnosis and effect), recall is improved for such a question of this kind -- one designed to measure the total frequency of occurrence in a specific reference period.

On the other hand, if the most easily accessible event is unrelated to an acute health problem, or if the episode remembered occurred early in the reference period, the answer to the frequency question may be incorrect. In the first instance, both over- and underreporting frequencies are possible; in the second case, there will be fewer reports of acute health episodes. Groves (1989) speculates that easily accessible memory is a poor indicator of an individual's

experience throughout a reference period, and one which can lead to errors with respect to frequency, recency and importance.

Brown et al. (1985) propose that an inferential process -- the "accessibility principle" -- is used by respondents for the estimation of the probability or frequency of an event, as well as for inferring event dates and recency. According to the accessibility principle, the subjective dating of events/experiences, and the judgment of frequency and probability, depend in part on the amount that can be recalled. Events with more facts accessible will appear to be more recent, more probable, and more frequent in occurrence.

These authors see a connection between their "accessibility hypothesis" and Tversky and Kahneman's "availability heuristic". In the former, the subjects tend to base estimates of frequency, probability and recency of an event on the amount known about an event; in the latter, the estimate is based on how easily the event can be recalled. The strength of the memory trace (accessibility) and the ease of information recall (availability) can lead to errors in recency, probability and frequency judgments. For instance, when two events occur at the same time, the one that is more retrospectively memorable will be estimated to have occurred more recently. Further, when two events occur at the same time, but recall of event generates more detail, it is that episode which will be mistakenly judged to have occurred more recently (Kahneman and Tversky, 1972; Brown et al., 1985).

Frequency and probability judgments are also affected by another judgment strategy called the "representativeness heuristic". Kahneman and Tversky (1972) describe it as a tendency to over generalize from incomplete information or small samples (Groves, 1989). An example of this inferential mechanism comes from the controlled experiments of Dawes (1988). He noted that when subjects were presented with evidence that the conditional probability of teenagers using marihuana was high given that they used any drugs, the

subjects inferred then that the conditional probability of teenagers using other drugs was also high given that they used marihuana.

Another inferential strategy to determine frequency of occurrence is "decomposition" (Armstrong et al., 1975). Respondents tend to break a problem into its subcomponents. For instance, if a subject were asked to determine the frequency with which members of his family ate out in a restaurant over the past 12 months, typically he would determine a general rate of occurrence, and then multiply the rate generated by the time period requested (i.e., the multiplicative approach). Bradburn et al. (1987) describe another decomposition approach -- additive decomposition. The respondent calculates values for mutually exclusive and exhaustive components of the desired quantity (in this case the number of breakfasts, lunches and dinners). Both of these techniques can be included in the survey design to increase response accuracy by means of guided decompositions controlled by the researcher.

Response errors can also be associated with an individual's explanation of the nature of memory. For instance, if a respondent has difficulty remembering an event (i.e., the memory trace is inaccessible or weak), he will infer that the event occurred infrequently, long ago, or not at all (Tversky and Kahneman, 1973; Lichtenstein et al., 1978). This incorrect reasoning strategy is identified as a factor responsible for "telescoping" -- the incorrect estimation of event frequency within a given reference period. Telescoping occurs when a respondent incorrectly includes into the reference period events which actually happened earlier. For instance, a respondent might erroneously include an episode of the flu experienced 16 months ago in answering a question about the frequency of acute health problems in the past year (Bradburn and Sudman, 1979).

Brown et al. (1985) suggest that one explanation for errors due to telescoping is that subjects recall episodes of the class of events requested (i.e.,

acute health problems), but cannot remember specific dates. If they recall an incident that actually occurred before the time period requested, but the memory is easily recalled and detailed (according to the accessibility principle and availability heuristic), subjects may incorrectly infer that the episode was recent enough to be included.

Brown et al. (1985) note that autobiographical sequences -- the temporal organization of personal events/experiences -- can be used as a recall strategy to diminish the bias generated by these faulty inferences. They argue that error is minimized because the sequences anchor events onto a 'personal time frame'. As such, they can provide additional information about dates. Judgment goes beyond inference based only on ease or detail of recall.

Neter and Waksberg (1964b) also show that "bounded recall" can reduce response errors due to telescoping. It is used frequently in health and consumer expenditure studies. This strategy uses data from a previous interview as recall cues in the next time period. Here, respondents report on events over an extended time interval (usually a year), but are interviewed periodically (every 3 months) about expenditures during that particular time period. The interviewer gives the respondent the data from the prior period and asks about expenditures since the last interview. The previous interview acts as an anchor as well as a retrieval cue to reduce faulty recall of events from an earlier period.

Finally, overreporting and underreporting of event frequency can be explained in reference to "anchor and adjustment heuristics" (Tversky and Kahneman, 1974). Anchoring and adjustment relates to questions requiring estimations. According to these heuristics, a question is answered by choosing a preliminary estimate or approximation (referred to as an 'anchor'); the respondent then adjusts it to specific differences implied by the question. For instance, if a woman is questioned about the frequency of pap smears in the last 5 years, she

may take as her anchor the normative expectation of 'once per year', and then adjust that by an awareness of deviations (i.e., consistency or inconsistency with the stated norm; if she goes more or less often for the checkups). Groves (1989) noted that anchoring and adjustment heuristics are most often used when respondents must estimate event frequency (total occurrence) over a long reference period, and also when the enumeration task is judged to be error-prone. For shorter reference periods, decomposition-enumerative techniques (additive or multiplicative) are used.

Inaccuracy of recall may result from incomplete memory, or from retrospective distortion of information after reflection on the issue, or from a combination thereof (Ross, 1980). "Faulty" recall is unintentional false reporting due to poor memory or changing perceptions of past events. However, a subject might also be biased toward the researcher, desiring to help the project or even desiring to conform to societal expectations about proper behaviour. "Falsified accounts" involve intentional false reporting. Explanations for such behaviour include the fear of being honest, or a desire to project a false image for ego enhancement.

Threats to accuracy of recall in survey reports arise from both cognitive and motivational factors. In order to report accurately, the respondent must understand and remember the information on the one hand, but must also be willing to report it, on the other (Rodgers and Herzog, 1987). The authors surmise that with the aging of society, there are greater concerns about the validity of surveys of elderly populations. Some experimental evidence suggests that increase in memory loss is a function of aging; this seems to happen for distant as well as recent events. Also, older respondents may resist reporting embarrassing information, but overreport desirable behaviours.

Although the accuracy with which people can recall past events is a crucial scientific issue in case-control research, it has received relatively little attention. Investigators seldom report the results of either large or small studies in which interviews were repeated at some suitable time after the original encounter, in order to determine the variability of responses to the same set of questions (Feinstein, 1985a, pp.501 et seq). As such studies are done after the outcome events have occurred, much time may have elapsed and the subjects may have great trouble remembering exactly what happened.

The research subject has two different tasks. First, she must try to recall, as accurately as possible, what actually happened; secondly, she must make the "anamnestic" effort with adequate vigour, regardless of personal status as a case or as a control. Additionally, because controls have not experienced an outcome event that might stimulate recall of exposure, they may not clearly remember what happened. A woman who has been diagnosed with breast cancer is much more likely to ruminate about lifetime exposures and to read about breast cancer etiology than is a woman with a normal mammogram.

Feinstein (1985a) suggests that an important first step in health surveys is a crude assessment of the subject's "sensorial" competence; this is the ability to understand questions, remember events, and respond accurately. The researcher can use various approaches to stimulate memory and improve recall: remind subjects of occasions on which exposures might have happened; provide lists of commercial names of possible agents to which subjects might have been exposed. However, such tactics may create bias rather than improve accuracy. Such multiple-choice questions may best be left until after the subject responds to more open-ended questions about exposure. Feinstein (1985a) suggests that a researcher can attack the problem of anamnestic bias by choosing a control group

who are likely to have reviewed their history with a vigour similar to that of the cases (p.508).

### Summary

In the previous sections are presented some findings of cognitive psychology and survey research regarding the memory and judgment factors which influence the reporting accuracy for autobiographical events/experiences. These factors can be used to understand and to explain why survey response may be accurate, incomplete or distorted.

However, these theoretical observations and hypotheses about factors relevant to survey response effects cannot be applied directly to survey research in general, or to the potential findings of my proposed research, should recall bias indeed be found to exist. The methodological limitations, and the resulting non-generalizability of the completed cognitive research discussed above, demand prudence in assessing the applicability of such findings to other kinds of research.

For instance, in evaluating different cognitive theories, researchers have often used a biased sample -- subjects who were for the most part university students. Due to the homogeneity of this study sample, the resulting data could not be used to generalize the results with confidence to other populations. These researchers often are reduced to generalizing their results solely on the assertion of their theory alone.

In the experiments on judgment heuristics, specifically the "availability heuristic", there appeared to be little or no concern with measuring the level of effort (i.e., subject motivation) which respondents were willing to invest in providing an answer. In several cases, respondents were allocated only brief time intervals to answer questions. Because of such shortcomings, the applicability of judgment errors to survey results might well be limited.

The irrelevance of many tasks used in these laboratory experiments also prevents the direct application of the results of cognitive psychological research to survey research, and in particular, the question-answering tasks given to survey respondents. Many of the retrieval tasks involved the recall of lists of words, nonsense syllables, visual images. These certainly cannot be equated to the retrieval of real-life, personal/autobiographical events or experiences.

Also, the time frame for retrieval tasks does not reflect the reference periods normally encountered in survey research. In the laboratory setting, the researcher designs retrieval tasks after relatively short time periods, often only minutes or hours between exposure to the material and the measurement of recall. Results based on such data may not reflect the nature or magnitude of the problems faced by survey researchers, who require recall of events from very-long-term memory.

Finally, such cognitive research does not consider the sociological factors which influence what is actually communicated once retrieved. Groves (1989) argues that cognitive research "often implicitly assumes homogeneity of cognitive and response behavior across persons given the same task". There is a need for cognitive psychologists to investigate the effects of various sociological variables, such as social desirability, task difficulty and the salience of the event, on the accurate recall and subsequent communication of autobiographical events, behaviors and attitudes (p.409).

It would be fruitful for the cognitive psychologist to work on interdisciplinary research projects with the social psychologist, whose interests in the impact of social context (desirability), the characteristics of the interview situation, and the effect of the characteristics and behaviors of the interviewer on respondent behavior would add a new dimension to the study of report accuracy.

### 2.3.5 Respondent Rule Effects

"Respondent rule" refers to the eligibility criteria of surveys: specifically, **who** may answer the survey questions (Groves, 1989, p.414). Here, the debate is focused on the relative accuracy of self reports versus proxy reports. Respondent rule effects are an issue only in non-attitudinal research, which seeks information about individual behavior or observable characteristics. (In attitudinal research only the individual is considered an acceptable respondent).

Although it is generally believed by researchers that self-reports are more accurate than reports obtained by proxy, this is not necessarily so. Cognitive and social psychology offer very different theoretical insights into the process of memory storage and the communication of responses, insights that may help to explain respondent rule effects in surveys. The first perspective draws upon inferences from schema theory about the nature of the encoding process for information about self, as opposed to information about others. Social psychology, on the other hand, focuses on the differential influences of social desirability upon responses which one gives about oneself and those which one gives about others. In addition to the effects of social desirability, social psychologists note that different 'roles' provide to their incumbents different information about different events.

From the perspective of the cognitive psychologist, the respondent rule controversy must be addressed by looking at the encoding and organization of memories. The purpose of any research is the generation of accurate data. Therefore, one necessary attribute of a good respondent is that she has encoded and retained in memory the information which is relevant to the survey questions.

Schema theory research has shown that there are differences both in how an individual perceives his own behavior as opposed to the behavior of others, and in how such information is encoded. Jones and Nisbett (1972), argue that self-

schemata differ from others-schemata because the images possible for the self are limited. Schema theorists argue that this perceptual difference affects the stimulus information encoded, and the general organization of memories about ourselves versus others.

Groves (1989), in discussing Lord's experiments (Lord, 1980), suggests that memories about oneself are organized around "central emotional states or other internalized attributes", whereas memories about others are organized about "observable traits and actions" (p.415). Furthermore, these findings may be generalized to naturally occurring events which are often the focus of survey measurement. If so, the implications are that proxy reports would be more accurate when the information requested covers characteristics of a person that involve physical action (e.g., episodes of acute health care, visits to the dentist). Self-reports would be better for information concerning more internalized states (e.g., frequency of chronic health problems, looking for work). Self-respondents may be inaccurate when an event is inconsistent with the internalized states dominating the self-schemata (e.g., underreporting of alcohol consumption when the individual believes that he has no drinking problem).

Mathiowetz and Groves (1985) reviewed the literature on respondent rule effects for health reporting and found that, contrary to the prevailing beliefs, self-respondents are not consistently found to provide more accurate health data than proxies. They delineate several reasons though why self-reporters might be more accurate than proxy reporters: (1) proxies may not possess the knowledge about the event or characteristic in question; (2) because events occurring to others are usually not as salient as events which occur to oneself, some events may not be reported, or only the most memorable (i.e., a serious health condition) is reported. Saliency may also affect a respondent's ability to date events accurately when reporting for others (Groves, 1989).

On the other hand, there are circumstances where proxy reports may be more accurate. Mathiowetz and Groves (1985), discuss role function within families as a factor affecting report accuracy. Health researchers have argued that knowledge about health-related events is more compatible with some self-schemata (roles) than others. The role of family 'health monitor' is therefore cited as an example where proxy reporting may be more accurate.

In general, different roles provide individuals with different information about different events. According to this hypothesis, it may be argued that the responsibility inherent in the 'health monitor role' may heighten for her the salience of events occurring to other household members, and lead her, as proxy, to provide more accurate reports. In other words, the health monitor may be better motivated, and also have the relevant knowledge, to answer the survey questions. Depending on the nature of material being requested in a survey, other role definitions may also be able to provide accurate proxy reports.

Social psychologists also identify social desirability effects as another factor influencing the accuracy of reporting about oneself and about others. They argue that when levels of information held by two persons are the same, if the trait being reported is perceived to be socially undesirable, it will be less often reported about oneself by oneself than by another. It has been demonstrated that respondents find it easier to report embarrassing or threatening information about someone else than about themselves.

In general, response effect studies suggest that self-reporting is not necessarily better. Estimates generated from survey research may differ depending on the respondent rule that is chosen (Mathiowetz and Groves, 1985, p.639). Furthermore, the choice of respondent rule may depend on the nature or purpose of the survey, and specifically, the type of information that is being sought.

The respondent rule chosen should reflect the best source of information while taking into consideration social desirability effects.

### **2.3.6 Nonattitudes and Acquiescent Response Behavior**

Another source of nonsampling respondent error is found in attitudinal research, and can be categorized as errors associated either with failure to comprehend the survey questions, or with the possession of nonattitudes.

Converse (1970), in studies of respondent opinions on various political issues, observed a group of subjects who provided inconsistent reports when questioned over repeated trials. It was argued that the inconsistency reflected a non-stable, non-permanent attitudinal state. Converse labeled these individuals as holders of "nonattitudes" because for them the survey measures concerned issues to which they had given little prior consideration. Within this group of respondents, some subjects provided a substantive response while others answered that they had no opinion on the issue in question. In trying to account for these inconsistencies, Converse proposed that either the persons did not understand the questions in a consistent manner, or they lacked all the information/knowledge to be able to take a consistent position on the issue. When responses were given, Converse noted that they were random, crossing different response categories. [This discussion is found in Groves, 1989, at p.417]

Groves (1989) went on to discuss Converse's use of metaphor in cognitive psychology to describe human information storage. Converse (1964) explained the nature of this nonsampling error by using cognitive terminology to describe the problems which occur in information storage when nonattitudes are held. He conceived respondent knowledge as an interlocking system consisting of pieces of semantic memories, concepts and arguments pertinent to the issue. A survey

question then acts as a stimulus to the retrieval of information from this system. The respondent will either report an opinion that has been well-rehearsed in the past, or she forms an opinion by weighing previously encoded arguments and counter arguments. However, when the respondent has little or no information encoded and stored about the issue, there is no strong network of concepts, arguments and counterarguments. The same survey question would in this case activate only weak ties between concepts and arguments of secondary relevance to the issue. Which of these ties is assessed by the respondent as critical in forming an opinion or stance will not be consistent over time, because few of these knowledge frames can be differentiated on their strength. This theoretical perspective can thus account for the random, inconsistent responses over replications of the same question.

Smith (1984), in a review of nonattitude research literature, delineated two design strategies that can be implemented to correct nonsampling errors related to nonattitudes. The first employs "don't know" filters; the researcher would, before beginning the actual interview, question the respondent about whether he has carefully considered the issue prior to the interview. This would encourage the subject to provide a "don't know" response if she had not formed an opinion on an item.

The second strategy builds on the work of Schuman and Presser (1981). Their research demonstrated that specific indicators of intensity of feeling about an issue could be used as a good predictor of the consistent reporting of opinions. As a consequence, Smith (1984) suggests the use of follow-up questions about the intensity of respondents' expressed opinions -- that is, how strongly they hold their positions on a particular issue. Abelson (1986) also suggests follow-up questions about respondents' experiences in defending their positions, and their attempts to convince others of the relative merit of their position.

However, the proposal for longer questionnaires, which ask for more detail about the issues of interest, has not been favored due to the increased costs and possibility of non-response caused by the very length of the questionnaire/interview, with the consequent loss of subject motivation (Groves, 1989, p.419).

Acquiescent respondent behavior (also referred to as yea/nay-saying behavior) has also been identified as a source of respondent error. Agreeing-response bias primarily refers to the tendency to agree with attitude statements presented to them, but has been extended to yes/no attitude questions (Groves, 1989; Schuman and Presser, 1981, Chapter 8)

Research in this general area has been generated from psychological measurement studies on closed questions which require "yes" or "no" answers, with statements to which the respondent is to "agree" or "disagree", and with questions using scales from which the respondent must choose a category from an ordered set (e.g., strongly agree... neutral...strongly disagree).

In personality studies, social psychologists such as Couch and Keniston (1960) found a tendency for some subjects to agree entirely with the statements of another person, thus apparently disregarding the content and context of those statements. They regarded acquiescence as a personality trait and studied it as such. Converse (1964) described it as the tendency for less-educated respondents to be "uncritical of sweeping statements and to be 'suggestible' where inadequate frames of reference are available".

Both interpretations relate education level to acquiescent behavior, but suggest different dynamics. The first focuses on interview-respondent interaction and regards education as one indicator of social status. The second implies deference to interviewers and to interview statements because of poor cognitive

abilities (i.e., poorer education, or simply lack of opinions (nonattitudes) on the issues.

Rorer (1965) and Nunally (1978) questioned the importance of agreeing-response bias. Their studies showed that the magnitude of the agreement response bias was not significant as a measure of personality nor as a source of systematic invalidity in measures of personality or attitudes.

However, what must be emphasized is that this questioning of the importance of acquiescence in psychological research is not incompatible with the assumption of survey researchers that acquiescence is a significant source of response bias in survey research. Researchers who take this position argue that the phenomenon plays a significant role when educationally heterogeneous populations are interviewed but it disappears entirely when student samples and self-administered questionnaires are used (Schuman and Presser, 1981).

Further, Schuman and Presser (1981) find that agreement due to acquiescent behavior can be remedied in the design of the study. The problem is minimized or eliminated when, instead of a statement, the subject is offered a forced choice between the statement and its opposite. This observation would imply that acquiescent behavior may be more a function of the questionnaire format than a characteristic of the respondent.

### **2.3.7 Sociodemographic Correlates of Respondent Error**

Four respondent attributes have been identified as potential sources of response error in retrospective survey research: education, sex, age and race. The impact of each respondent characteristic upon response effects will be discussed separately. Then this section will look at response effects due to the interaction of

respondent variables and "task" variables. Interaction between respondent attributes and interviewer characteristics will be discussed in Section 2.4, dealing with the interviewer as a source of measurement error.

Sudman and Bradburn (1974) comprehensively reviewed survey research studies which specifically investigated different sources of response errors. They then summarized all these studies and determined the distribution of response effects due to age, sex, education and race. They concluded that none of these variables is statistically significant when examined separately (p.98). The largest response effect among these variables is for number of years of education (which is measured as the number of years of formal education). However, as a measurement criterion, the number of years of formal education is criticized by researchers as a poor choice, because previous psychological research has shown it to be a poor indicator of crystallized intelligence (Groves, 1989; Schuman and Presser, 1981).

Much of the following discussion is taken from Groves (1989, pp.441-448).

As a source of respondent error, **education** is often considered as a 'proxy' variable for a respondent's cognitive abilities, i.e., the ability to comprehend survey questions and the ability to retrieve, reconstruct and communicate responses to survey questions (Krosnick and Alwin, 1987; Schuman and Presser, 1981). Because researchers believe that formal education is indeed indicative of an individual's cognitive abilities and general knowledge, they then make several assumptions: that less educated respondents are slower or unable to comprehend survey questions, lack the knowledge or opinions to answer them, will have difficulty communicating their responses, and might be influenced by the interviewer or be more apt to have their choice influenced by irrelevant cues (Schuman and Presser, 1981).

A review of the literature suggests that the education response effect was most pronounced for adult respondents with less than eight years of schooling. Bradburn and Sudman (1979) note that adult subjects with grade school education provide more missing data (i.e., a higher non-response rate and an increase in "don't know" responses) in surveys. In addition, these subjects are more susceptible to interviewer effects, such as being influenced by the perceived differences in status between interviewer and respondent.

Converse (1970), in a multivariate analysis of "no opinion" behavior in Gallup and Harris Surveys, included education, the length of the questions (>30 words), whether the question forced a choice between two or more response categories, and whether the topic concerned material related to foreign political affairs as predictors of "don't know" response rate. He found that education was the strongest predictor of "don't know" responses. But, as this study did not control for age, other researchers think that age differences may account for the observed differences over the education levels (Bradburn and Sudman, 1979; Groves 1989). The study was incomplete in that it did not identify the source of the problem -- whether due to lack of comprehension of the survey questions, lack of knowledge of the topic, inability to retrieve the information, the complexity of the language used in the questions, or interviewer effects. Groves did speculate that less educated subjects were more willing to answer "don't know", possibly due to less perceived pressure to appear informed relative to the more educated subjects. However, Schuman and Presser (1981) note that these results are not uniformly obtained when other studies are reviewed. In some studies asking for information on obscure topics, subjects with college education were found to provide a higher percentage of "don't know" responses.

Sudman and Bradburn (1974) also found that women tended to answer "don't know" more often than men, except on topics related to birth control and

morality (p.100). They suggest that this difference reflects the different roles played by men and women, hypothesizing that it is more acceptable in our society for women to admit that they "don't know" or have "no opinion", than it is for men. They also indicate that this observation could correlate with level of education.

As mentioned previously, when researchers discuss education effects, they often propose that less-educated respondents will be slower to understand the context and meaning of the question, and consequently will be more apt to have their choice influenced by irrelevant cues (Schuman and Presser, 1981). In studies on question structure and order, Schuman and Presser found very mixed support for the hypothesis that the less-educated are sensitive to question effects. What they did observe was that less-educated respondents gave different responses to open-ended survey questions than to equivalent close-ended questions. They "explained" this observation by noting some evidence that those with less education are either more influenced in their choice of answers by the very fact of being forced to make choices (closed responses); or they have difficulty communicating answers to open-ended questions.

Age is another respondent attribute associated with response error. It is an important source of measurement error and the only one related to memory. It is however difficult to integrate the results of research upon the effects of age upon response errors, due to the disparate and inconsistent definitions of the "elderly" age group. The term "elderly" spans the 55 - 70+ years range (Groves, 1989, p.441).

In general, it has been observed that with aging there is a decrease in response performance characterized by increasing failure to retrieve information from memory (Craik, 1977; Sudman and Bradburn, 1974; Groves, 1989).

From an extensive review of the literature on age effects and memory, Craik (1977) compared and summarized the results in performance for elderly versus younger subjects and concluded that:

(1) elderly subjects have larger deficits in recall from secondary or long-term memory than in recall from primary or short-term memory. This observation may have little if any impact on surveys designed to retrieve autobiographical information because the conclusions were based on recall of information from semantic memory only;

(2) elderly subjects' performance is poorer on recall tasks than on recognition tasks. In their studies on recall and recognition in the elderly, Herzog and Rodgers (1989) asked elderly respondents at the end of an interview session to name six physical functions that were the subject of the questions earlier in the interview. Afterwards, they presented 20 survey questions to the subjects (only 10 had been previously presented), and asked the respondents to identify which, if any, had been asked of them. The data from these studies suggest that both recall and recognition decrease over the age groups; but comparatively speaking, recognition tasks were performed better. Even when Herzog and Rodgers adjusted for educational differences, these effects remained.

Researchers such as Groves (1989, p.441) and Schuman and Presser (1981, pp.91-92) have also observed that it is often more difficult to focus and keep the attention of the elderly on the interview task; often, they stray off topic and fail to follow the interview protocol. Their answers are often only "tangentially relevant" because they do not respond to the particulars of the survey questions.

Sudman and Bradburn (1974) also note that the percentage of non-response rates, and "don't know" responses increase as a function of age. In explaining this, they argue that the elderly disengage from societal activities as they age.

Acceptance of this age-trend effect is not unanimous. Glenn (1969) argued against; his work suggested that when strict educational controls were introduced, differential effects due to age disappeared. However, the weight of the research would still favor Sudman and Bradburn's observation. Therefore, in summary, the percentage of "don't know " responses decreases with respondent education; and, the percentage of "don't know" responses increases with respondent age. Because of these two trends, it is assumed that there is a correlation between education and age in the occurrence of non-response rates.

A final observation in the literature is that elderly subjects tend to be more susceptible to interviewer bias effects in their responses to survey questions (Sudman and Bradburn, 1974; Schuman and Presser, 1981).

Groves observes further that different theoretical perspectives are offered to explain response effects related to age. Smith (1980) suggests that errors could be due to: (1) poorer organization of memories during the encoding stage; (2) decreased attention and cognitive processing at the acquisition and retrieval stages; (3) interference during recall because of the extensive, rich links between any particular retrieval cue and information in long-term memory. With older subjects, who have years of experiences, their memories are as plentiful as the connections between them. This may result in diminished, distorted recall.

Groves (1989), however, recognizes that there is really no causal model for the effect of aging upon encoding and retrieval processes.

Hulicka (1967) attributes memory response effects to physical changes occurring with chronological age and poor health. It is suggested that chronological age is a proxy measure for physical attributes such as loss of brain tissue and poor vascular circulation; these changes may affect brain functioning (i.e., encoding and retrieval processes); as well, poor health may account for diminished motivation and attention in elderly subjects.

Sudman and Bradburn (1974) found that sex differences interact with several task variables: the threat posed by the question, the structure of the questions (open versus closed format), the length and difficulty of the questions, the method of administration, and whether or not there is a preferred or socially desirable answer to a question.

The research on response effects due to sex differences indicates that response variance is larger for females with close-ended, threatening questions where a socially desirable answer is possible. The response effect for women is twice as high as that for men when the questions are threatening; the reverse situation exists when the questions are non-threatening.

The method of administration (face-to-face interview versus a self-administered questionnaire) and two respondent characteristics, sex and race, influence measurement error in surveys. Male respondents, both black and white, find face-to-face interviews more threatening, especially when some questions have preferred or socially desirable answers. Here, the response effects are the most pronounced. When the race of the respondent is considered separately, response effects are larger for black than white respondents on face-to-face interviews, and black subjects are influenced more by interviewer effects (i.e., deference is higher for black respondents). When a socially desirable answer is possible, the response effects are larger for white respondents.

Sudman and Bradburn (1974, pp.102-109) also note that five task variables interact with respondent attributes to produce measurement error in survey research: the degree of threat posed by a question, the method of data collection (face-to-face interview vs self-administered questionnaire), question structure (open-ended vs close-ended questions), question length, and whether or not a socially desirable answer is possible). They also found that women were influenced a little more than men by question structure. They postulated that the

difference may be related to different interpretations of the question. Close-ended questions with forced choices are a greater source of response error for women.

The length of the question may also affect response. Longer questions are often more difficult; respondents have difficulty understanding what is being asked of them. Also, the longer the question, the more likely it is that interviewer effects (i.e., intonation cues, different wording, etc.) will be activated. Furthermore, the length of the question often influences the length of responses, and can lead to incomplete or inaccurate reporting. Sudman and Bradburn (1974) found that two respondent characteristics (race and education) interact with question length. Racial effects are the largest source of error. When questions are short (<12 words), there are no differences between black and white respondents. The response effect is larger for blacks when the question is longer (>28 word). For the education variable, measurement error is greatest when the respondent is less educated (high school education or less), and when the sentences are long and complex (>18 words)(Sudman and Bradburn, 1974, p.110).

Two respondent characteristics, sex and race, interact with social desirability. The response effect is greatest for women (nearly twice as large as that for men), and for white respondents when a socially desirable answer is possible.

#### **2.4 Factors Affecting Exposure Reporting and Recall Accuracy: Interviewer as a Source of Measurement Error**

Because the survey interview is the means by which response measurements are obtained, the interviewer can play a significant positive or negative role in the process of recall, in the interpretation of what is recalled, and

in the actual recording of the response information. The interviewer herself can generate response measurement error at any one of these stages.

Despite the potential for interviewer bias, concern about it has decreased in the last 20 years because its overall impact on study variance has been minimized through such improvements in survey design protocols as the standardization of question wording and administration, and the use of non directive probing procedures. As well, there is more rigorous training of interviewers; training interviews teach one how to be objective, neutral and accepting of all responses when collecting and recording data.

Methodologists are more concerned now with survey measurement error arising from **task variables** (Sudman et al., 1977). Sudman and Bradburn (1974, pp.13-16) state that if a researcher wants to understand 'how' interviewers contribute to response errors by introducing "variable" measurement across interviewers, she must consider the problem from three perspectives: interviewer role demands, interviewer role behavior and the extra-role characteristics of the interviewer (i.e., race, sex, education, and social class).

First, interviewers can influence subject responses by the way they carry out the interview role demands -- how they read survey questions, how they clarify respondent misunderstanding, how they probe incomplete answers, and how they handle subjects' questions.

Training of interviewers is designed to produce complete uniformity of behavior among the interviewers, thereby removing any effects which variation in interviewer behavior might have on respondent answers. The greater the degree of structure in the interviewer's role, the lower the relative response effect will be (Sudman and Bradburn, 1974). However, training programs cannot always ensure that interviewer behavior will be consistent with the study protocol.

Secondly, interviewers may administer the questionnaire differently to different subjects. They have been known to reword or eliminate some questions; to record responses incompletely, inaccurately or falsely; and to use different probing strategies when the respondent does not comprehend a question or is having difficulty communicating a response to an open-ended question. Interviewer effects are therefore possible, even if the study protocol is adhered to and all questions are administered and read correctly. Furthermore, interviewers can unconsciously change their intonation, or use different, unplanned words. In these diverse ways, an interviewer's behavior is not always consistent with interviewer role demands.

The survey questions to the respondents will vary across interviewers, producing measurement error. A general finding is that the greater the degree to which the interviewer adheres to the role demands required by the interview/study protocol, the lower the relative response effect will be across interviews (Sudman and Bradburn, 1974).

Thirdly, a survey interview is a structured social interaction (Kahn and Cannell, 1957) conducted within the context of a complex set of social norms which guide interactions among individuals. Therefore, it is subject to the same social factors that influence other interactions, such as the interviewer's and respondent's sociodemographic profiles (i.e., their race, education, sex and socioeconomic status). These variables often act as cues which help respondents make decisions about their own behaviour while helping the interviewer to interpret responses (Groves, 1989, p.359).

In Cannell's 1977 summary of the results of studies on interviewing methods (discussed earlier with respect to the respondent as source of measurement error), the author concluded that research has shown that the actual

behaviour in an interview is the main variable which correlates with the index of reporting quality. The "cue-search model of interview interaction" posits that the respondent looks to the interviewer, or to some other source, for cues about expected behaviour; the interviewer is in a similar situation, searching for cues. Research has also shown that changing the characteristics of the process, including interviewer behaviour, can have marked effects on both the amount and accuracy of health data reported by respondents (Cannell 1977, p.37).

Cannell (1977) states that interviewer feedback can bias answers, or it can improve response validity. The effects of verbal reinforcement on respondents can be divided into three categories: cognitive effect, conditioning effect, and motivational effect. These categories overlap and interact. The first effect occurs when verbal reinforcement supplies cues about the interviewer's expectations and about how the respondent is meeting them. The second effect is important in many studies of the psychology of learning. In the simplified model of interview, the researcher's evaluation immediately follows the respondent's answer; this can reinforce the response or can also alter the frequency of the behaviour that preceded it. This process can thus either strengthen or weaken the probability of eliciting that behaviour in subsequent trials. The third possible effect is motivational -- the intensity or psychological effort which the respondent gives to the reporting task, and to other behaviours which may interfere with the adequacy of response.

Cannell (1977) concludes that reporting accuracy may be improved by manipulating the conditions under which retrieval occurs. The **conditions** of recall have a crucial impact on "what" is reported, and on accuracy. Different questioning strategies can improve reporting by changing the conditions under which the respondent is invited to search for past events. However, these studies were not concerned directly with the cognitive processes involved in recall.

Like other researchers, Cannell (1977) believes that an experienced event is not merely recorded in original form, as on a tape; rather, it is organized into a perceptual field. Its **meaning** depends upon how it is perceived, and with what other events it becomes associated in memory. What an interviewer might see as a simple item may, in fact, be organized in several frames of reference by the respondent; a single question about the event may not be the best stimulus to recall; several questions from different reference points may be necessary.

Cannell (1977) sets out a model of information processing (p. 53). This shows the respondent's cognitive organization and the researcher's questionnaire design as two diverging paths, which lead to two independent informational states -- memory trace, and stimulus question -- whose interaction in the interview is expected to produce the retrieval of the original information. This model suggests that the probability of accurate recall is a function of the ability of the stimuli questions to interact adequately with the respondent's cognitive organization. The appropriateness of the stimuli questions is itself a function of the researcher's ability to comprehend the nature of the respondent's cognitive path, and to use this knowledge in framing the questions. An event may be stored in memory under various states so distant from the original information state that a question stimulus merely traced from the original event, or from its straightforward conceptualization by the researcher, may not elicit the stored information. For instance, memory can process an "illness" in ways that transform and organize it around such varying concepts as pain, incapacity, costs, doctor's visits, hospitalization, medication, treatment, symptoms -- or even more generally around other causal, circumstantial, or consequential events.

Sudman and Bradburn (1974) refer to the variables which act as cues in interviews as "extra-role characteristics". Their study conclusions are that higher social status interviewers induce a larger response effect than do interviewers of

lower social class status -- if, and only if, the respondent is aware of the interviewer's socioeconomic status. The status of the interviewer is important in school studies, where status is recognized. The resulting error is associated with an incorrect interviewer perception of the respondent's answer: it is speculated that an interviewer may unintentionally "hear" and record answers more consistent with his own views. The response effect is most pronounced with questions dealing with social class. The nature of the question also determines the impact of interviewer characteristics on response effects. Results of numerous studies, including Katz (1942), show that the greater the saliency of the interviewer's extra-role characteristics for the subjects being investigated, the larger the relative response effect will be (Sudman and Bradburn, 1974, pp.15, 110-111).

The interviewer's sex and race have no significant impact on response errors. However, there is a trend that suggests a higher "don't know" response rate with female rather than male interviewers, and for inexperienced lower social class interviewers. Furthermore, the "don't know" rate diminishes as the education and experience of the interviewer increases, regardless of sex. The race of interviewer and respondent influence survey error when the questions deal with racial attitudes and issues. Stronger, more militant answers are given to black interviewers by black respondents when questions deal with race. Differences disappear with non-racial questions.

There is a paucity of data regarding response effect interactions between age of interviewer and respondent. One general observation is that older, female interviewers get lower response effects in face-to-face interviews than do younger interviewers, especially inexperienced undergraduate university students.

Sudman and Bradburn (1974) have also discussed the impact on measurement error of the joint effects of interviewer characteristics and a number of task variables. Their observations can be summarized as follows:

a. Method of Administration:

(1) The age of the interviewer is relevant when the interview is face-to-face. Response effects decrease as the age of the interviewer increases.

(2) In self-administered questionnaires, the education of the interviewer is the more important variable. The more educated this individual is, the larger the response effects, at least in school settings, where status is based on education. To extrapolate for example, a pregnant woman given a dietary questionnaire by a doctoral research student may indicate that she is consuming milk when her daily fluid intake consists of soda pop and coffee exclusively. Here, trait desirability and social approval factors play a role in what is reported on the self-administered questionnaire.

(3) The interviewer's sex is important only when survey data are collected by means of a personal interview. Somewhat larger response effects are attributed to male interviewers.

b. Structure of Interview Questions:

(1) Close-ended questions have a larger influence on respondents than do open-ended, but they minimize interviewer bias by providing more structure.

(2) Concerning the race of the interviewer, it was found that response effects obtained by white interviewers are higher than for black interviewers on open-ended questions. But the studies are small.

c. Possibility of a Socially Desirable Answer:

(1) When a socially desirable answer is very possible, response effects are larger for white interviewers. The reverse is true when no socially desirable answer is possible.

(2) Results for interviewers by sex were the opposite of those found for respondents. When there is a possibility of a socially desirable answer, the response effects are more than twice as large for male rather than for female interviewers. If there is little possibility of a socially desirable answer, there is no difference between male and female interviewers for response effects.

Several research studies have been completed about the influence of question type (factual versus attitudinal) on interviewer generated bias (discussed in Groves, 1989, p. 373). It was often assumed that factual questions with knowable and verifiable answers would be less influenced by such interviewer variations as differences in question wording, question administration, and delivery/intonation. The results of studies comparing interviewer effects on factual and attitudinal measures are in fact mixed, with only some of them corroborating the original assumption. For example, O'Muircheartaigh (1976), found larger response effects for attitudinal questions, and in particular, those stated as open-ended questions. Collins and Butcher (1982) found factual questions less susceptible to interviewer effects.

Fowler and Mangione (1970), used a regression model predicting Kish intraclass correlation coefficients (i.e., a unit-free measure expressing the ratio of variance between interviewers to the total variance). The predictor variables were defined as specific characteristics of the survey question and included the following: the difficulty of the question (degree of cognitive processing required), the vagueness of the terms in the questions, the threat or sensitivity of the topic, whether or not the question was factual or attitudinal, and whether the sentence

was open-ended or close-ended. The most important predictor of the Kish intraclass correlation coefficient was the task difficulty imposed by the question. Fowler and Mangione found no evidence that factual items are subject to lesser interviewer effects than are attitudinal questions. Open-ended questions are not susceptible to greater interviewer variance; however, the number of answers obtained to an open-ended question is quite sensitive to interviewer effects (i.e., variation due to different probing behavior). Probing by interviewers resulted in additional information.

In social psychology and sociology, there is literature about how interviewers' expectations influence response variation: they affect the manner in which questions are presented to the respondent, including word changes, variations in intonation, and other attributes of questionnaire administration that influence respondents in different ways. Interviewer expectations may influence both the answer given by the respondent and what is recorded by the interviewer.

Hyman (1954) was the first to investigate the role that an interviewer's prior expectations might play in invalidating survey data (Groves, 1989, p. 395). He identified three kinds of interviewer expectations: (1) role expectations (i.e., the interviewer expects certain responses from different groups of individuals such as women, blacks, clergymen, laborers); (2) attitude structure expectations, in which the interviewer expects respondents' views to be internally consistent; and, (3) probability expectations. Hyman suggested that prior to the commencement of the survey, interviewers have "probable expectations"; they expect a certain distribution of expected answers congruent with their own beliefs about the prevailing sentiments in the general population. Subsequently, their behavior during the interview may effect such a distribution. In other words, Hyman argues that the interviewer expects a certain distribution of responses and then unconsciously tries to fulfill that distribution expectation.

Work by Rosenthal (1966) and Rosenthal and Rosnow (1969) suggests possible interviewer effects related to both their opinions and to their expectations of the respondent. Interviewer's expectations might cause biased data collection in several ways. First, bias may occur because the interviewer's opinions are communicated to the respondent; the respondent then modifies her own responses accordingly to fulfill the expectations of the interviewer. Secondly, the interviewer might ask leading questions to probe the respondent's replies, or she may fail to probe unclear or inappropriate answers, or she might be biased in which replies are probed and how. Thirdly, the interviewer might be selective in which responses are recorded; she may even record what the respondent "meant" to say. Fourthly, the interviewer could also bias survey data during sampling through the choices of whom to interview.

It is evident that survey responses may be incomplete, inaccurate or totally false if interviewer effects are significant (Cannell et al., 1977; Sudman et al., 1977). Sudman et al. (1977) argue that prior expectations may also relate to anticipated difficulty in asking questions of the survey respondents, to subject uneasiness about answering threatening or sensitive questions, or to expectations about the levels of under- and overreporting/ percentage of "no opinion"/"don't know" responses. Here, the hypothesis would be that interviewers who anticipate difficulty with a study or high item non-response may not probe incomplete or ambiguous responses, or will communicate a lack of confidence to the respondent.

To test this assumption, Sudman et al. (1977) had all the interviewers complete a questionnaire prior to the survey. It measured the interviewers' perceptions of how difficult it would be to ask the survey questions, how inhibited the respondents would be in answering the questions, what the non-response rate would be for certain questions, and how large the underreporting of sensitive or threatening information might be. Their results do not support the hypothesis that

interviewer's prior expectations result in interviewer variance and response effects. Interviewer expectations concerning reporting errors and difficulty of administration were not predictive of respondent behavior. There was a slight, but insignificant tendency toward lower reporting of sensitive/threatening information to interviewers who anticipated difficulties with these questions. However, the results of the study must be considered only tentative, because of a serious design limitation: the interviewers were not assigned randomly to the respondents. This makes the interpretation of interviewer effects problematic.

If an interviewer is not "blind" to the outcome, he might pursue exposure information more vigorously for known cases than for controls who show no abnormality (Levin, 1983). However, information gathered by trained interviewers is more reliable than that collected by self-administered questionnaire because accurate interpretation of the questions is important. Levin also believes that design will allow for the demonstration of interviewer bias, if it exists.

Grichting and Caltabiano (1986) conclude, however, that no standard procedures are applied to measure and evaluate the amount and direction of bias resulting from the interviewing experience. From an extensive review of the literature on the dynamics of interviews, they had little doubt that interviewing does change a respondent's opinions, beliefs, and action tendencies.

Because Feinstein (1985a) believes that both an interviewer's attitude and mode of investigation can substantially affect the response of a subject to a question, for example about prior exposures, he says that where possible it is best for the researcher to be "blinded" from the research hypothesis and from the subject's status (case or control). To reduce interviewer bias, he also recommends that a rigorous and relatively rigid format be used for data acquisition. Phrasing of relevant questions -- those whose answers provide crucial research information

-- as well as the methods of recording answers should be applied uniformly in all interviews; the format should also allow for each pertinent positive answer to be followed by additional questions, which are also uniformly arranged and phrased (Feinstein, 1985b).

In conclusion, it is important to consider that interviewer bias is not as critical as the impact of various task variables on response effects. However, certain characteristics of the interviewer, either alone or in combination with the characteristics of the respondent and relevant task variables, might be sources of survey measurement error. Sudman and Bradburn (1974) suggest that interviewer bias must be addressed by considering: interviewer role demands, interviewer role behavior and the extra-role characteristics of the interviewer. They give three guidelines which provide assistance in understanding and preventing this source of error (p.15):

- (1) The greater the degree of structure in the interviewer's role, the lower the relative response effects. Interview protocols and interviewer training are important.
- (2) The greater the consistency between interviewer role demands and interviewer role behavior, the lower the response effects will be. This emphasizes the need for extensive interviewer training.
- (3) The greater the saliency of the interviewer extra-role characteristics for the questions being asked, the greater the relative response effects will be.

## **2.5 Factors Affecting Exposure Reporting and Recall Accuracy: Task Variables as a Source of Measurement Error**

Sudman and Bradburn (1974) found few studies about the effects of "task variables" on survey measurement error: these are the response effects due to the

conditions under which information is acquired by the interviewer and responses are generated by the respondent. The authors' synthesis of a large body of research on response bias suggested that the **task** -- retrieval and reporting of information -- and the conditions under which it is performed is the largest source of response effects.

From their analysis, they distinguished three categories of task variables which appear to influence the accuracy or the variance of responses: (1) task structure; (2) problems of self-presentation; and (3) the saliency of the requested information to the respondent. Within these categories, they identify the following factors as having the largest influence on survey measurement error: level of psychological threat, the possibility of a socially desirable answer, the saliency of the questions to the respondent, the method of administration (personal interview or self-administered questionnaire), the location of the interview, as well as the position and structure of the questions.

Self-reporting inventories are a cheap and efficient means of data collection (Furnham, 1986). They can be administered by research workers with clinical experience. However, like all self-response measures they are open to response bias, which must somehow be dealt with. The usual means have taken one of three forms. First, to provide a "lie scale" within the questionnaire itself to detect unreliable subjects, and expose differential reporting between cases and controls; this should be sensitive to both over- and underreporting. Secondly, to emphasize "honesty" in responses. Thirdly, to reduce the face validity of some questions, so that respondents are not as aware of what the assessors are trying to measure.

In 1990 Coughlin reviewed significant literature and concluded that little study has yet been done into the factors contributing to bias due to differential recall between cases and controls in retrospective studies. He too thought that interviewing techniques do influence recall. He noted that Schlesselman (1982)

and Rossi (1983) believe that the content and the form of questions both may affect recall accuracy. Supplementary devices such as introductions to sections of the questionnaire may improve responses, possibly due to the stimulus and time provided to the respondents.

Schuman and Presser (1981) noted that the greatest potential for non sampling error is how the questionnaire is constructed, that is, the choice of wording, the use of open-ended versus close-ended questions, and the characteristics of the interview situation in which the questions are delivered. The length of time since the event occurred (i.e., the total duration of the survey's reference period) and the referent person about whom the questions are asked (respondent rule effects) also influence response effects, but their impact on response variance is believed to be less significant.

Oksenberg and Cannell (1977) are concerned that the nature of the respondent's task -- first in comprehending the meaning of the question, and then in retrieving, reconstructing and reporting the required information -- may create demands which the subject is unable or unwilling to meet because they exceed, or seem to exceed, his memory or his ability to process and integrate information. Some respondents will not show the requisite motivation, and will subsequently perform these tasks with minimal effort; they approach the task demands as "cognitive misers".

When the respondent has not understood the question, or is not sufficiently motivated to retrieve and reconstruct the necessary information, extraneous cues (such as the status, behavior and appearance of the interviewer, the respondent's beliefs, values and goals, or his assumptions about the intended meaning of an ambiguous question) may drive the selection process, determining what is reported, and the degree to which it approximates the true response.

The normal complexity and demands of the information and response processing may be further increased when the respondent considers the response to be embarrassing, sensitive or personally threatening/uncomfortable. Here, the psychological implications of providing responses which accurately reflect the respondent's beliefs, attitudes, values or behavior may lead him to suppress the information (underreporting) or distort it into a more acceptable response for protection of his self-esteem (overreporting of desirable behavior/attitudes).

Raphael and Cloitre (1994) also discuss the impact of the respondent's affect on memory and memory retrieval. They suggest that a 'mood-congruence model of memory' may partially explain the occurrence of differential recall of prior exposures (i.e., recall bias). Commenting on the research of Blaney (1986) and Ucros (1989), these authors suggest that "negatively-toned prior exposures are recalled more easily by respondents in a negative mood state at the time of recall than those in a positive mood state" (p.556). They also note that this may be a significant problem in psychiatric epidemiology because "negative mood or demoralization is often an indicator of case status" (p.556), and the specific research requirements in this domain often involve the ascertainment of negative life experiences and events. Raphael and Cloitre (1994) also stated that "mood congruent patterns may not occur when recognition memory processes (i.e., respondents indicate that a proposed exposure did or did not occur) are invoked in the context of an epidemiological study" (p.556). Research suggests that mood congruent effects would be more likely to happen when the memory demands are low, that is, when the subject is required to remember only a small amount of information, and there is no time delay between the exposure and the request to recall the information (p.556). They further comment that mood congruent effects may not be relevant in epidemiological research due to the high memory demands -- the need to remember and to report relatively large amounts of

information about remote and low-salience life experiences and events. Lastly, they report that "mood may have an impact beyond memory retrieval: it may influence respondents' reconstruction or subjective evaluation of details about recalled experiences", specifically, the assessment of "how frequent, how important, or how positive/negative a prior exposure was" (p.556). In summary, the 'mood congruence model of memory' as an explanation for recall bias proposes that the particular mood state of the respondent affects recall of prior exposures, and that the mood state may often differ between the cases and the controls, unless the research design uses a 'mood equivalent' control group. Furthermore, a depressed mood predicts poorer recall of instances of prior exposure; however, subjective assessment of those exposures tends to be distorted in a "mood congruent manner".

Raphael and Cloitre (1994) also noted that a "mood related memory deficit may reduce effect sizes artifactually" (p.555). Consequently, they recommended that "the recall of event occurrence must be considered separately from subjective appraisal of event characteristics" (p.555).

These situational factors (task variables) can definitely lead to biased or distorted data. The most frequent distortion in survey data is the failure to report information (i.e., false negative reports). This can be due to a failure in the information retrieval process, a true memory lapse, or carelessness/unwillingness to make the effort necessary to retrieve the information. False negative reports are quite common in reports of past behaviour or experiences, especially when the time between event occurrence and interview is long.

Another common distortion involves making false positive reports -- that is, falsely reporting events, behavior or other factual information as having occurred. This distortion often appears when there is a reference to time; such **telescoping** (overreporting) errors results from compression of time -- the event is remembered

as having occurred more recently than it did. False positive reports may also reflect faulty recall or may be related to the need for social approval and the possibility of giving socially desirable answers (Cannell et al., 1977).

As Cannell noted earlier, a respondent's task in answering a broad question is enormous. She must create appropriate frames of reference to guide recall, create cues to reactivate traces of possibly low salience. One cannot expect the motivation to invest substantial effort to be high, especially if the questionnaire has no immediate benefit to the respondent (Cannell, 1977). Questioning only allows short periods of time to complete this process. Within this framework error is predictable. The broad question is not an adequate stimulus.

Instead of asking one standard question derived from a simple conceptualization of the event, several questions may be needed, from various hypothesized states of memory processing. That is, instead of requiring the respondent to build up her own cues, the researcher should try to create these recall aids and build them into the questionnaire. If the researcher can successfully predict and design the relevant cues and frames of reference, the respondent's recall process should be significantly assisted, with resulting improvement of recall.

Cannell (1977) suggests that an extensive questionnaire, containing a large number of questions providing multiple and overlapping cues, may assist retrieval. However, it also might inhibit participation due to time involvement and effort obviously required. Nevertheless, he concluded that the cue-giving approach, for instance using symptomatic manifestations of illness as a frame of reference, is more productive in eliciting the report of illness, than are standard general questions.

He thought that the involvement of the interviewer and the length of questionnaire might convey the message that the recall task was important,

heightening motivation. However, he concluded that very little was known about the asking of appropriate questions. It might be that **reporting** errors are often the result of **questioning** errors. Long questions might elicit both more information and a more accurate report, contrary to common assumptions. Question length may have a cueing effect.

Findings from a number of validity studies (Neter and Waksberg, 1964b; Sudman and Bradburn, 1974 (Chapter 3); and Cannell et al., 1977) on the effect of these task variables on response variance are summarized:

(1) Response accuracy/variance is influenced by where and how the interview is conducted. Regarding the method of administration, self-administered questionnaires are better than personal interviews (i.e., are associated with more accurate and complete reporting) when the questions to be asked are personally threatening, or when a socially desirable (preferred) answer is possible. Face-to-face interaction is an important factor in the generation of socially desirable responses; there is a tendency for respondents to present themselves favorably to the interviewer, or to report behavior and attitudes that conform to the socially acceptable norms. Socially desirable behavior is likely to be overreported (i.e., false positive reports). Behavior or attitudes which are sensitive, embarrassing or threatening, and which therefore conflict with a norm of "self-presentation", are likely to be underreported (Phillips and Clancy, 1972). Social desirability also works in conjunction with factors such as threat or saliency. Attitude questions rated as highly threatening and having a strong possibility of a socially desirable answer have much larger response effects than any other category of attitude items. Among **behavioral** items, the effects are largest for items with a strong possibility of a socially desirable answer and which are somewhat highly threatening (DeMaio, 1984; Sudman and Bradburn, 1974).

(2) Differential response effects occur for the different subjects studied regardless of the conditions of the interview. Sudman and Bradburn (1974) note that factors such as threat, social desirability, and memory factors are probably responsible for these effects. Their analysis found that threat and saliency work in opposite directions on response variance. The largest response effects are associated with threatening questions. Conversely, events important to the respondent are recalled more easily (in accordance with the availability heuristic and accessibility principle) and reported more completely and accurately than those of lesser psychological significance. Questions about salient events are more likely to motivate the respondent to follow the retrieval and memory reconstruction processes. When considering these two task variables in combination, the largest response effects would occur where saliency is low and threat is high.

(3) The age of the respondent and interviewer can each create response effects when survey questions deal with behavioral/attitudinal information perceived as threatening (i.e., illegal behavior, racist attitudes), sensitive or embarrassing (i.e., sexual practices). Here, the largest response effects are found with young respondents and interviewers, and in particular, college students. Self-administered questionnaires are the method of choice when highly threatening questions are to be asked and anonymity is required.

(4) For threatening questions or those with a socially desirable answer, the analysis of Sudman and Bradburn (1974) suggests that close-ended questions increase the threat by forcing the respondent to choose one of the response alternatives; the result can be large response effects. Underreporting or false negative reports are noted when a personal interview is used rather than a self-administered questionnaire, and also when the interview is conducted at home when others are present. Furthermore, short questions have a strong negative

(underreporting of behavior) effect on reports for threatening behavioral/attitudinal questions. Research has shown threatening questions should be asked towards the end of an interview when rapport is established between the interviewer and the respondent. It is believed that if the topic is threatening, sensitive or embarrassing, the greatest threat would occur at the beginning of the interview with the threat diminishing as the interview progresses and rapport is established.

(5) Respondents are more likely to report socially undesirable, sensitive and embarrassing attitudes/behaviors about others than about themselves. Therefore, self-reports are more likely to be less accurate than proxy reports under these circumstances.

(6) False negative reports (underreporting rates) are related to the time elapsed between the occurrence of the events to be recalled and reported and the interview, to the salience of the events for the respondents, and to the perceived social desirability of the events (Sudman and Bradburn, 1974).

Wicklegren (1970) reported that the majority of research in experimental psychology suggests that short-term and intermediate memory decays exponentially with time. Cannell et al., (1963) found that the failure to report visits to physicians over a two week period, increased from 15% after one week to 30% in interviews two weeks later.

There are no data available for long-term memory effects as a function of the reference time period. In general, as the time increases between the event and the interview, there is increased underreporting of information about that event (i.e., errors of omission). Because of the greater time lapse, the cognitive demands to define what information is relevant, to recall it and to reconstruct it are greater; extraneous cues (interviewer characteristics, respondent goals, etc.) may then erroneously affect the accuracy and completeness of the information reported.

Because remembering events in the distant past can be taxing to an individual's cognitive skills and capacities, the use of a personal interview and probing techniques could result in fewer omissions; however, personal interviews are associated with telescoping errors. It is interesting to note that errors due to omissions and telescoping can occur simultaneously during recall. For very long periods, there will be more errors of omission than overreporting due to telescoping. Errors of omission also depend on the saliency of the event: memory is better for highly salient items.

The cognitive approach to questionnaire design conceptualizes the response to a survey question as involving four distinct stages, each of which can involve erroneous reporting (Jobe and Mingay, 1989). The first is comprehension, interpreting the meaning of the question. Secondly, there is retrieval, in which the respondent searches long-term memory for relevant information. Thirdly, there is estimation or judgment, the evaluation of the retrieved information as to relevance; the respondent may then combine separate information items to form a response, or alternatively she may decide that the recalled information is inadequate, using that decision as a start point in forming an adequate response. The fourth stage is response. The subject weighs such factors as sensitivity of the question, social desirability of the answer, and probable accuracy.

The authors reviewed three reports from the US National Center for Health Statistics, which discussed ways to minimize reporting errors. Respondent comprehension rose when simpler terms were used, even though the original wording was considered to be comprehensible. Respondent recall was also improved by providing additional cues for hard-to-remember information. Researchers also used "decomposition" to lead subjects to break down generic memories so as to recall individual events such as health visits. Another technique is the creation of a personal timeline of accurately dated landmark events in the

subject's life, against which he can try to place, for instance, particular health events such as visits to doctors.

Kalton and Schuman (1982) studied the effect of the question on survey response. Such responses may be sensitive to the precise wording, format and placement of the question. Their conclusions are that questioning is not a precision tool; there is ample evidence that serious response errors can and do occur. Although much research has been done, we remain largely ignorant of the nature of question wording and or form effects.

Reviewing some of the authorities discussed above, such as Sudman and Bradburn (1974), and Cannell (1977), the authors discuss in particular the effective construction of "factual" questions for such surveys as case-control studies. The start point must be a precise definition of the fact/information to be collected. It has often been shown that apparently small changes in definition can have large effects on survey results. Of concern is that a precise definition may lead to an unwieldy question, which the respondent cannot or will not make the effort to absorb. A respondent needs to understand both what is being asked of him, and what is an appropriate response.

Such problems as telescoping and social desirability effects have already been canvassed. The authors also consider that the random response technique can protect a respondent's privacy, particularly when threatening or embarrassing questions are asked. The respondent chooses which of two or more questions he answers by a random device; he answers the chosen question, without the interviewer being aware which is being answered (Kalton and Schuman, 1982). Several studies have obtained higher rates of reports of sensitive information from random response techniques than from traditional questioning. However, any gain in bias reduction has to be set against a sizable increase in sampling error; the

technique also hampers analyses of the relationships between the responses to the threatening question and other variables.

The authors consider that the various approaches of Cannell (1977) to the problems of memory errors and of sensitive questions, have resulted in improved reporting. Although longer questions may sometimes yield fuller answers, they can be a cumbersome tool, however. Experiments on a carefully thought out mix of long and short questions show an increased yield of reports on health events. By essentially stating important questions twice, the questionnaire improves the respondent's understanding of what is required by giving more time to marshal one's thoughts and recall; as well, a respondent may interpret the length of the question as a sign of its importance and give it greater consideration.

Another technique involves the use of instructions to the respondent at the beginning of her task, to think carefully, search her memory, take her time to check records, and answer as completely as possible. Researchers may also use feedback, and deliberately secure the respondent's commitment to respond conscientiously. Evidence of experiments on the utility of these techniques suggests that each leads to improved reporting, with a combination of all three techniques giving the best results of less under- and overreporting (Kalton and Schuman, 1982).

For Feinstein (1985b) the format of the health interview is the most important scientific instrument of many case-control studies. In a well-conducted study, the investigator may take additional pains to check the consistency and accuracy of the interview process. This may involve data collection of the same information by the **same** method several months later; or it may require checking data with a family member who knows the subject, as well as any archival material (Feinstein, 1985b). He is also concerned that if the exposure is not well specified during data collection, inaccuracy may arise. For instance, the subject

may have used certain pharmaceutical substances, such as aspirin or food additives, without being aware. Unless the researcher has established a complete list of all the ways that exposure might have occurred, and unless the subject is asked about all those possibilities, the occurrence of an exposure might not be recognized. As well, if these inquiries are not then applied equally for all subjects, whether case or control, the results may be biased.

In conclusion, despite four decades of academic discussion about the nature, prevalence, characteristics, causes and indeed the very existence of recall bias, many conclusions still seem tentative. Perhaps this is understandable, as the basic problem is rooted in the nature of human memory and in the social motivations of humans, often when subjected to the additional stress of involvement in a disease process. Nevertheless, the literature discussed above does, at the least, provide both cognitive and sociological schemata within which to place the anecdotal findings of recall bias. It would appear that the most important lessons for an epidemiologist or designer of health research studies centre on the conclusion that both respondent and task variables are likely the largest source of response bias effects. There certainly are enough perceptive insights about the demonstrated shortcomings of case-control research to apply to future studies in an effort to eliminate or at least minimize recall bias effects, if they exist.

## **2.6 Review of the Literature: Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)**

In the previous section of this chapter, the cognitive, psychological and social sciences literature was canvassed, and the factors responsible for recall and recall accuracy were reviewed, along with a discussion of their possible influence on the reliability and validity of the data collected. The field of cognitive

psychology provides the most thorough understanding of the processes of human memory, how memory errors occur, and the means to control them through research design.

Part two of this literature review examines epidemiological and other health-related studies which specifically assess the reliability of the exposure data, and then determines the nature and impact of any resulting exposure misclassification (recall bias) on the estimates of effect. Harlow and Linet (1989) and Austin et al. (1994, pp.65-75) provide a fairly complete listing of the studies which have studied recall accuracy and recall bias, and have provided useful guidance in the compilation of this literature review in a table format. The evidence for the existence of differential exposure misclassification (recall bias), and the effects of misclassification bias on relative risk estimates in completed case-control studies are reviewed and summarized in Table 2 (pp.102-145) of this chapter.

Overall, none of these studies provides strong and consistent evidence for the existence of appreciable recall bias, and significant distortion of the relative risk estimates. Methodologists such as Coughlin (1990) recommend further studies of exposure misclassification in different research domains with respect to the effects of different exposures, length of recall, and other factors which may account for differential recall. In addition, Austin et al. (1994) concluded that future case-control studies must be evaluated for their ability to detect subtle and weak associations, and when possible, their methodology improved to be "more sensitive and specific to weak and moderate associations" (p.74). These authors also note the importance of considering non-differential exposure misclassification when they identify it as one of three "biggest threats to the validity in case-control studies"; one of the others is recall bias resulting from differential recollection of past events for cases and controls (p.75). Thus, there is a need to study both non-

differential and differential exposure misclassification in the context of case-control studies. Following on this, the next logical step for investigation is the feasibility of developing a validity scale for the measurement and control of non-differential and differential exposure misclassification. This dissertation is designed to address these areas of concern.

## **2.7 Raphael's Proposal for the Measurement and Control of Recall Bias: The Development and Implementation of an Exposure Data Validity Scale**

During the review of the literature, it became apparent that only a limited number of studies had directly addressed the problem of exposure misclassification and recall bias in case-control studies, and that the empirical evidence for the existence of recall bias was inconclusive (Lippman and Mackenzie, 1985; Mackenzie, 1986; Mackenzie and Lippman, 1989; Friedenreich, 1990). Furthermore, the findings of these studies did not provide strong evidence that differential exposure misclassification (recall bias) was as serious a case-control deficiency as it was conjectured to be. There was no significant group differences in the reporting of past exposures and the biasing of the estimates of effect. Those studies which provided findings in support of recall bias were themselves, often subject to methodological problems (e.g., insufficient study power, failure to assess the validity of both false positive and false negative reports of exposure for both the cases and controls, lack of (suitable) controls for case-control comparisons, the use of different data sources for both the collection and the comparison of prospective and retrospective exposure reports, the salience of the outcome event, the length of recall, etc.) and their evidence had to be called into question.

Nevertheless, the opponents of case-control studies have persisted in their strong criticisms, and in their challenges to the scientific structure and the credibility of retrospective, observational research to provide unbiased estimates of association, and the generation of valid study conclusions. At the same time, the case-control paradigm is acknowledged by these same critics, as the design of choice for the study of rare and chronic diseases such as cancer, where the latency period between exposure and disease occurrence was long, and logistical and ethical reasons precluded the implementation of randomized clinical trials or cohort studies.

At first, these contradictions and ambiguities seemed irreconcilable. However, Raphael (1987) provided the insight and the methodological guidance to resolve these issues. She proposed the development of a validity scale for the measurement and control of recall bias. If successful, this scale would improve case-control methodology overall, while increasing its acceptance as a valid research paradigm for the etiologic investigation of the determinants of health and disease. The development, implementation and evaluation of an 'exposure data validity scale' became the very motivation for, and a primary focus of this dissertation. If a 'validity scale' could be easily developed, and shown to be effective as a design standard for the estimation and statistical adjustment of exposure misclassification, then it could be used in every case-control study. Its routine inclusion would meet the requirements of the investigator to provide evidence regarding the reliability and validity of the exposure data collected in a case-control study, the existence, magnitude and direction of any existing exposure misclassification, and if significant, the means to statistically adjust the relative risks for any distortions (Raphael, 1987, p.168). Consequently, researchers and critics would be more confident and accepting of the ability of case-control studies to generate valid study conclusions.

Raphael's (1987) proposal was "adapted from the logic of the validity scales of the Minnesota Multiphasic Personality Inventory (MMPI) which attempted to adjust some of the other scales (in the inventory), based on a measure of each respondent's test-taking attitude or response set" (1987,p.168).

According to Raphael (1987), the exposure data 'validity scale' should consist of 'plausible but fake' risk factors for the disease under study. The exposure variables that are included in the scale must also meet **the following criteria**: 1) the exposures and events "should be of approximately equal plausibility when compared to the exposures which are the putative risk factors of the research study. Unless they are equally plausible, the validity scale will not appropriately measure 'search for cause' cognitive processes" (p.169); and, 2) the exposures and events cannot be related to the development of the study disease.

Once the scale items are selected, the respondents would then be questioned regarding their previous exposure to each item on the validity scale. Subsequently, their responses would be used to estimate the presence of exposure misclassification, particularly, differential recall (i.e., the tendency of cases and controls to either over- or underreport antecedent events and exposures).

If a specific exposure variable is not a risk factor for breast cancer, then the proportion of cases and controls exposed to this 'plausible but fake' risk factor should be approximately the same, and the resulting estimate of association (odds ratio) should be equal to 1.00, thus, indicating no risk for disease development.

According to Raphael (1987), differential exposure misclassification (recall bias) would be suggested, for example, when "case respondents positively endorse an excessively large number of validity scale items in comparison to control respondents" (p.169). She goes on to argue that "...the endorsement would likely be due to overreporting recall bias rather than actual higher rates of exposure"

(p.169). Here, the estimated odds ratios for these variables would be significantly different from 1.00.

Raphael (1987) suggests that "by comparing the **total** validity scores for cases versus controls", the researcher will be able to determine if recall bias exists, as well as its impact on the measures of association (odds ratio estimates) (p.169)[my emphasis]. Because "the validity scale score is a function of the extent of each respondent's recall bias" i.e., the subject's tendency to over/underreport previous exposures and events, the summary within groups validity scale score "may be entered into the final analysis as a statistical control for recall bias" (p.169).

In summary, Raphael's validity scale proposal was intuitively appealing because it offered case-control researchers the opportunity to assess and to control for the effects of differential exposure misclassification (recall bias) in **any** case-control study: the scale construction appeared to be straightforward. Section 3.11 of Chapter 3 describes the stages in the construction of the exposure data validity, as well as the statistical program used to assess the etiologic importance and specific weights for the exposure factors selected for inclusion in the scale.

## **2.8 Methodological Considerations for the Design of an Exposure Data Reliability and Validity Study to Assess Exposure Misclassification**

Mackenzie (1986) provided design clarification as to how the question of exposure misclassification (recall bias), could best be studied. She noted that the problem of reporting (recall) bias should be examined by collecting exposure information prospectively "when subjects are at risk of becoming cases", and then collecting the same information, by the **same** data collection method, retrospectively, once the subjects are cognizant of their disease status (Mackenzie, 1986, p.35). By using this design, the researcher is able to assess the impact of group membership on recall, and specifically, whether cases and controls recall

their past exposures similarly or differently. If group differences exist in the prospective and retrospective reports of exposure, the researcher can conclude with confidence and increased certainty, that any existing exposure misclassification is due in fact to systematic case-control differences in recall accuracy, and not to differences in the way the data were collected. Other aspects of study design are discussed in Section 3.1.

## 2.9 Summary

In this chapter, I have attempted to provide a review of the methodological limitations of case-control research, and its susceptibility to biases which could invalidate study conclusions. In addition, an extensive overview of the various subject response, task and interviewer variables which could affect respondents and their ability to recall past events and exposures both accurately and reliably was provided. This background information was included because of its importance for a complete understanding of 'how' and 'why' exposure misclassification occurs, 'why' the subjects in a case-control study may be predisposed to remember and report personal information differently, as well as the way these factors may contribute to exposure misclassification (i.e., the overreporting/overstating and/or the underreporting/understating exposure), and their impact on the biasing of the exposure-disease odds ratios (i.e., towards or away from the null value).

This chapter has also provided the background justification for and the significance of this study. As discussed in Chapter 2, very little research has been completed on the reliability and validity of exposure data, including non-differential and differential exposure misclassification. The research which has been done, has not provided strong and consistent evidence that exposure misclassification is as significant a problem in case-control research as it is

suspected to be. Furthermore, studies have not demonstrated that the exposure-disease odds ratios have been biased by either non-differential or differential exposure misclassification so that study conclusions were invalidated.

Given the susceptibility of case-control studies to inaccurate recall of past exposures and events, the possibility that odds ratios may possibly have been biased, and the relative lack of empirical research in this area, a research need was clearly identified in the area of case-control methodology. Therefore, this study was designed to determine the suitability of case-control research for studying disease etiology. Of particular concern was the requirement for an assessment of the reliability and validity of exposure data, the determination of the presence or absence of non-differential and differential exposure misclassification, and an evaluation of the impact of any resulting exposure misclassification of the estimates of effect. Raphael's proposal (1987) for the development of a validity scale to measure and control recall bias was also investigated within the context of this dissertation.

In the next chapter the specific research design and methods that were used in this study to address these questions will be outlined. As noted in Section 1.3, the discussion of the specific research methods will include such topics as: the choice of a nested case-control study design, the recruitment and selection of cases and controls, the use of multiple control groups and anamnestic controls, as well as the specific procedures used to collect and analyze the data. Chapter 3 also describes the specific steps in the development and construction of an exposure data validity scale which will be evaluated as a possible design strategy for the measurement and control of differential exposure misclassification (i.e., recall bias).

Table 2: Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
1. Klemetti and Saxén, 1967	<p>A case-control study of the association of non-chronic maternal disease and drug usage in early pregnancy with a deviant pregnancy outcome (i.e., neonatal death, abortion, stillbirth and congenital malformations).</p> <p>Prospective and retrospective exposure reports obtained by personal interview were compared with information recorded in the clinical records. The prospective data regarding antenatal drug usage and non-chronic maternal disease were collected during the fifth month of pregnancy.</p>	Yes	No	<p>Antenatal drug usage</p> <p>Non-chronic disease</p>	<p>1. It was concluded that both the prospective and retrospective exposure reports were unreliable. Overall, there were no significant group differences regarding recall accuracy.</p> <p>2. Only 25 % of the prospectively collected exposure information (i.e., drug usage and non-chronic diseases) was recalled and reported accurately in the retrospective postnatal interview (Klemetti and Saxén, 1967, p. 2075).</p> <p>3. Sixty six percent of the retrospective positive exposure reports "could not be confirmed from the prospective interview or information collected from other sources" (Klemetti and Saxén, 1967, p. 2075).</p>

Note: The section on 'Evidence of Exposure Misclassification/Recall Bias' refers to the researcher's interpretation of the study results, and whether or not the researcher's assessment of the data provided evidence that recall bias was present.

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
1. Klemetti and Saxén, 1967 (continued)	<p>The results of the prospective and retrospective interviews were compared to determine if recall bias was present.</p> <p>N = 406 (203 case mothers and 203 controls (matched to the cases for time of birth and clinic))</p>				<p>4. Klemetti and Saxén (1967) noted that "new" and incorrect exposure information was provided retrospectively by the mothers (p. 2074).</p> <p>5. There were no significant case-control differences in the percentage of identical replies. The pregnancy outcome (deviant vs normal) and the condition of the child did not affect recall accuracy (Klemetti and Saxén, 1967, p. 2074).</p> <p>6. There was no empirical evidence of recall bias.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
<p>1. Klemetti and Saxén, 1967 (continued)</p> <p><b>Remarks:</b></p> <p>The results of this study are inconclusive and cannot support the finding of no evidence of recall bias. Mackenzie (1986) noted several methodological limitations which prevented the proper assessment of these prospective and retrospective exposure reports for recall bias. These included:</p> <p>(1) Only the positive (+) exposure reports were considered. Properly conducted recall bias studies must evaluate the prevalence of both false (+) and false (-) reports of exposure by the cases and the controls, and then determine if groups differ systematically regarding overall level of recall accuracy.</p> <p>(2) Case-control distinction was not maintained in this study. "The units studied changed from the women (i.e., the cases and controls) who provided the reports, to the reports themselves" (Mackenzie 1986, p. 27). Here, the problem was the absence of case-control comparisons to determine if the prevalence of discrepant reports were sufficient to bias the estimates of association (exposure-disease odds ratios) for the various study factors.</p> <p>(3) Klemetti and Saxén (1967) also stated that approximately "two-thirds of the positive replies in the retrospective study could not be confirmed from the prospective interview or information collected from other sources" (p. 2075). Here, it must be emphasized that data reliability studies as well as studies of exposure misclassification (recall bias) must use the same source of data for the prospective and retrospective comparisons. Health and pharmacy records are not 'gold standards'; their accuracy and completeness may vary for the cases and the controls. Health records will only contain what the subject reports, or what is observed and reported by the health care provider. As such, case-control discrepancies could be related to the method used for data collection rather than real differences in recall accuracy (Mackenzie, 1986; Lippman and Mackenzie, 1985; Mackenzie and Lippman, 1989).</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
2. Hewitt et al., 1966	A case-control study of the relationship between antenatal x-rays (i.e., abdominal and chest), toxemia and anemia and the subsequent development of childhood cancers.  Personal interview data were compared with antenatal records.	Yes	No	Antenatal x-rays, toxemia and anemia	The authors concluded that there was no general tendency for mothers of live children to report fewer prenatal events when compared with mothers of dead children. These conclusions were based on a sensitivity and specificity analysis of "checked statements" (pp. 82-83).
<p><b>Remarks:</b></p> <p>The results of this study are inconclusive due to methodological problems. The use of antenatal records and the radiologists' reports weakened this study and its conclusions. The authors noted that the antenatal records did not contain any information about events which happened after admission to hospital, or before a woman seeks antenatal care. X-rays during labour or during the early weeks of pregnancy for non-obstetric reasons were missing. In fact, antenatal records were missing for 43% of the sample and were incomplete for the remaining 57%. When these records were used as the standards for comparing the maternal exposure reports, the relative risk estimates would have been upwardly biased for abdominal x-rays, toxemia and downwardly biased for anemia and chest x-rays. False conclusions would have been based on discrepancies related to inadequate data collection procedures, and specifically, missing documentation. The study emphasizes the requirement to use the same data source for exposure data reliability studies, and for those studies designed to assess exposure misclassification/recall bias.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
3. Hopwood and Guidotti, 1988	<p>A case series study. The authors assessed the recall of symptoms in 22 of 31 subjects. These workers were exposed to nitric acid fumes from drums ruptured during a hazardous waste site clean-up operation in 1983.</p> <p>Symptoms recalled at 6 months were compared to symptoms reported at the time of the incident.</p>	Yes	Yes	<p>Symptoms related to nitric acid exposure: dizziness, headaches, respiratory problems (shortness of breath, sore throat, cough, sputum production), lightheadedness, unusual taste, eye discomfort, fatigue, nausea, pruritis, abdominal discomfort, paresthesia, and anxiety.</p>	<p>The authors observed substantial disagreement which exceeded that expected on the basis of chance alone. This discordance was consistent, and in the direction of more prevalent reporting of symptoms with the passage of time.</p> <p>They concluded that a high level of recall bias was present.</p> <p>Six months post-outcome, the authors noted that symptoms were more likely to be recalled and reported retrospectively than forgotten. False positive reports were more prevalent than false negative reports.</p>
<p>3. Hopwood and Guidotti, 1988 (continued)</p> <p><b>Remarks:</b></p> <p>This was a small case series: Only 71% of those subjects that were originally exposed were found and re-interviewed at 6 months. This study was unable to assess for the presence of recall bias because there was no control subjects included for the required case-control comparison. Recall bias is defined as differential reporting of exposure status by cases and controls (i.e., a phenomenon of differential reporting accuracy which is dependent on group membership). As such, this study was an exposure data reliability study (i.e., a determination of the consistency of reports given by the 21 cases at the time of the incident versus 6 months later). They should have concluded that recalled symptoms at 6 months were unreliable and lacked precision.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
4. Tilley et al., 1985	<p>A case-control study of the effects of diethylstilbestrol (DES) exposure during fetal life. The authors compared prenatal records with obstetric histories. These histories were collected by means of a self-administered questionnaire which was completed by the women 10 to 30 years after the birth of their daughters.</p> <p>N=3650: (3078 cases mothers and 572 control mothers).</p> <p>The case mothers also included DES-exposed women who were walk-ins or referrals to the project centres.</p>	<p>Yes</p> <p>The authors considered recall accuracy which was defined as the level of agreement between prenatal records and the reports provided by a self-administered questionnaire.</p>	No	<p>Drug use during pregnancy, pregnancy history, parity, miscarriages, threatened abortion, hospitalization during pregnancy, trunk x-ray and birth weight.</p>	<p>With the exception of data on treatment (i.e., hospitalization during pregnancy and trunk x-ray) and drug use, there were no statistically significant differences in agreement (obstetric history vs antenatal record) between the group of DES-exposed mothers identified through review of their prenatal records and the unexposed mothers. Agreement was better for DES-exposed mothers regarding treatment and drug use.</p> <p>Recall accuracy (i.e., the level of agreement between the prenatal record and obstetric history) was slightly better for the walk-ins/ referrals when compared with the two groups identified by the review of prenatal records.</p> <p>According to the results of the Kappa analysis, this study found good to excellent agreement for all groups when the mother's</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
4. Tilley et al., 1985 (continued)					<p>recall of her reproductive history was compared with medical records.</p> <p>The agreement was poor for the following variables: medical treatment, x-rays and drug usage during pregnancy. 37% of the DES exposed mothers either could not remember (29%) or denied (8%) using DES although it was recorded in their antenatal record (p. 269).</p> <p>The accuracy of recall is dependent on the type of exposure to be reported, as well as the level of detail that is requested. Clinical records were more complete when compared with physicians' office charts.</p>
<p><b>Remarks:</b></p> <p>(1) The study population was homogeneous: predominantly Caucasian and middle class. Therefore, the results of this study cannot be generalized to other populations.</p> <p>(2) Sample size was insufficient to test the study's underlying hypotheses.</p> <p>(3) The impact of case-control differences in recall accuracy were not evaluated by means of odds ratio comparisons.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
5. Preston-Martin et al., 1985	<p>A case-control study of the association of dental radiation and the occurrence of parotid gland tumors.</p> <p>Telephone interview information was compared to dental records.</p> <p>N=163: (84 cases and 79 controls).</p> <p>Length of recall - up to 30+ years.</p>	Yes	No	Dental radiation	<p>The authors conclude from the comparisons of chart and interview information that exposure recall appears to be unbiased. The measures of agreement between the two data sources were similar for cases and controls.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
6. Mackenzie and Lippman, 1989	<p>A nested case-control study of the association of 39 potential risk factors and possible adverse reproductive outcomes.</p> <p>N=747 (85 case mothers (whose infant died, had malformations, or was admitted to the intensive care nursery for longer than 24 hours for serious complications); 217 mothers (intermediate group) with infants of intermediate health status; and, 445 controls (normal healthy infants)).</p> <p>Pregnant women provided reports of exposure prospectively and retrospectively for the 39 study factors by means of a self-administered questionnaire.</p>	Yes	No	39 potential risk factors for adverse reproductive outcomes: chronic illness; stress; coffee, wine, liquor consumption; smoking; poor nutrition; nausea; medications; contraception; reproductive history; acute illness, family history of malformations, etc.	<p>The data from this study did not provide any evidence for the existence of recall bias. The authors found that:</p> <ol style="list-style-type: none"> <li>1. Inconsistency in the reporting of the study variables was evident; however, these discrepancies were similar for the study groups. The retrospective reports were subject to more post-delivery deletion of exposure information rather than post-delivery addition.</li> <li>2. There were no statistically significant differences in the frequencies or prevalence changes for the 39 exposure variables for the 3 study groups. In other words, there was no significant case-control differences in the group's tendency to add or delete exposure information postnatally.</li> </ol>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
6. Mackenzie and Lippman, 1989 (continued)	<p>Prenatal and postnatal responses were compared (reliability study) for the three pregnancy outcome group.</p> <p>Changes in the odds ratio estimates were also evaluated: case vs normal control; intermediate case group vs control.</p>	Yes	No		<p>3. The changes in exposure reporting were not related to group status (i.e., pregnancy outcome, maternal concern about the baby, or maternal socio-demographic characteristics) (Mackenzie and Lippman, 1989, p. 65).</p> <p>4. A comparison of the odds ratios from the prospective and retrospective data did not show a tendency to increase or the decrease the estimates of association between the risk factors and pregnancy outcome.</p> <p>5. The estimates of association were not biased by the resulting changes in exposure reports.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
<p>6. Mackenzie and Lippman, 1989 (continued)</p> <p><b>Remarks:</b></p> <p>(1) Sample size was insufficient: the study lacked the statistical power to test the research hypothesis and to demonstrate significantly biased reporting.</p> <p>(2) Mackenzie and Lippman (1989) noted that if one assumes that biased reporting is dependent on the salience and emotional impact of the outcome event (i.e., severe infant malformations which would stimulate biased reporting), then the failure of this study to provide evidence of recall bias may be a function of the small case population and the low incidence of very sick/malformed infants (8.2%) (p. 74). Cases were originally defined as mothers experiencing stillbirth or having a child with severe medical complications or malformations. However, the majority of the cases were not consistent with the inclusion criteria. Consequently, there was a loss of statistical power due to insufficient sample size for case group definition.</p> <p>(3) The cases may not have been different from the controls because the case mothers' infants required only transitional NICU care. This homogeneity may account for the similarity of reporting among the cases and controls. The number of cases needed to be increased, and be comprised of stillbirths, abortions, severely ill or malformed infants only so that recall bias could be studied.</p> <p>(4) The length of recall was of shorter duration than that usually encountered in case-control studies of chronic and rare disease (a few months vs many years).</p> <p>(5) The study subjects were unrepresentative of the general population. Less educated women and immigrants were underrepresented or excluded because they were unable to complete the study questionnaire. The study population was predominantly Canadian born, highly educated and sought obstetrician-based prenatal care. This was significant because lower SES women and less-educated women were at an increased risk for deviant pregnancy outcome (Mackenzie, 1986).</p> <p>It was concluded that the lack of evidence to support the existence of recall bias "does not prove that the bias does not, or cannot, exist" (Mackenzie and Lippman, p. 74).</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
7. Werler et al., 1989	<p>A case-control study of malformations. Interview data collected during the postpartum were compared to exposure information collected during pregnancy and then recorded in the mother's obstetric record.</p> <p>N=270 (105 cases (mothers of malformed infants) and 165 controls (mothers of non-malformed infants)).</p> <p>The medical record information was considered the 'truth' or (gold standard) for the determination of case-control differences in the reporting of the eight exposure variables. The researchers assessed the proportion of case mothers who gave positive reports given</p>	Yes	Yes (For some of the factors and not for others)	Medications taken and illnesses during pregnancy.	<p>The cases compared to the controls recalled a greater proportion of documented exposure for two of the eight exposures: periconceptual birth control and urinary tract or yeast infection. For birth control after conception, case reports were 8x more complete. The proportion of agreement was equal in the two groups for over-the-counter drug usage and elective abortion, and less for cases for nausea and vomiting. The authors concluded that recall accuracy was better for the cases, and therefore suggesting the presence of recall bias.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
7. Werler et al., 1989 (continued)	<p>that the exposure was recorded in the chart.</p> <p>The researchers assessed for the presence of recall bias by estimating relative sensitivity (RS) (i.e., the ratio of reporting accuracy for mothers of malformed infants to that of mothers of normal, healthy infants). If the RS measure &gt; 1.0, recall accuracy is better for the case mothers (Werler et al., p. 415).</p>	Yes	Yes (For some of the factors and not for others)		
<p><b>Remarks:</b></p> <p>The results of this study have been criticized by Swan and Shaw (1990) and Berg (1990) for the following deficiencies:</p> <ol style="list-style-type: none"> <li>(1) The impact of case-control differences in exposure recall were not evaluated by comparing the odds ratios for the two data sources (i.e., medical records and personal interview data).</li> <li>(2) The study had a high rate of non-participation for both the cases and the controls. Therefore, sample distortion bias may be responsible for case-control differences in recall accuracy.</li> <li>(3) The failure to consider potential overreporting by the cases (a function of specificity), as well as underreporting by the controls.</li> </ol>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
7. Werler et al., 1989 (continued)					
<b>Remarks:</b>					
(4) Critics were opposed to the suggestion by Werler and coworkers to use malformed controls due to the possibility of selection bias. Swan and Shaw (1990) noted that etiologic agents could cause increased risk for different malformations, and that the exposure-disease odds ratio would be biased towards the null.					
(5) The use of obstetric records as the standard 'criterion' against which the maternal reports were compared was considered inappropriate. This record can only be used to estimate the prevalence of false (-) reports; medical records report exposures only (if complete) and fail to document non-exposures. Two different data sources were used to measure recall accuracy. Therefore, the method of data collection could explain the resulting reporting discrepancies. The same source of data (maternal reports) must be used to study the nature and impact of recall bias ' (Mackenzie, 1986).					
(6) The findings of this study suggest only small differences in recall accuracy. Therefore, only weak evidence exists for maternal recall bias. The methodological limitations of this study would negate even this minimal source of evidence.					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
8. Stolley et al., 1978	<p>A case-control study of thromboembolic disease and oral contraceptive use.</p> <p>Interview data were compared with physicians' records.</p> <p>N=276 (79 cases and 197 controls (women with a history of oral contraceptive use within the previous two years)).</p> <p>Length of recall was up to 10 years.</p>	Yes	Yes (for some of the exposure categories)	Oral contraceptives (OCs) (10 brands): total duration, brand name, start date, and stop date.	<p>Case-control differences existed regarding the subjects' recall of duration of use, and dates for the use of the drugs.</p> <p>For total duration of use of OCs, the cases showed a higher rate of agreement with prescriber records than the controls. Among the controls, a higher percentage reported a longer duration of use.</p> <p>Agreement rates between subject reports and physician records were poorer regarding the dates of usage.</p> <p>Cases tended to have a higher agreement rate with their prescriber on the starting date of use (60.6%) compared to (48.1%) for the controls.</p> <p>Recall accuracy depended on the types of information and level of detail requested.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
<p>8. Stolley et al., 1978 ( continued)</p> <p><b>Remarks:</b></p> <p>(1) Results are not generalizable to the general population because the study population was homogeneous regarding ethnicity. The cases and controls were primarily Caucasian (71%). Therefore, the sample may possibly have been unrepresentative of the general population with respect to ethnicity.</p> <p>(2) Sample size was too small to complete a full study of recall bias, and to test the relevant hypotheses (i.e., study power was insufficient). As well, only 52.4% of the study population had physician records which could be used for comparative purposes in this study.</p> <p>(3) The impact of case-control differences in exposure recall were not evaluated (i.e., no odds ratio estimates were calculated for comparison).</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
9. Rosenberg et al., 1983	<p>A case-control study of the association of oral contraceptive use and the diagnostic outcome of hepatocellular carcinoma (HCA).</p> <p>Study population consisted of N=130 (61 cases and 69 controls). Only 43% of the original study was available for this follow-up study.</p> <p>Length of recall: 4-16 years previously.</p> <p>Interview reports were compared to questionnaire information provided by the physician (i.e., the prescriber of the Oral contraceptives).</p>	Yes	Yes	Oral contraceptives	<p>Overall, the agreement between the 2 data sources for: (1) month-specific duration of use; (2) duration of use and brand; and (3) duration, brand, and dose was 90%, 62% and 54% respectively.</p> <p>Agreement was significantly better for the cases than for the controls in all 3 areas (i.e., duration, dose and brand). When analyzing agreement for all 3 variables combined, the difference in percent agreement for the cases versus the controls was 62% vs 47% respectively. "These differences in agreement did not change appreciably when adjusted for race, education, marital status, religion or age at index date" (p. 85).</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
<p>9. Rosenberg, et al., 1983 (continued)</p> <p><b>Remarks:</b></p> <p>(1) The impact of case-control differences in recall accuracy were not assessed by odds ratio comparisons calculated separately for the two data sources.</p> <p>(2) The study sample was too small. Thus, there was a loss of statistical power for hypothesis testing.</p> <p>(3) Only 66% of the original HCA study group were included in this analysis. Women who could not remember their prescribing physician were excluded. It is assumed that these women may not recall contraceptive use as well. Their exclusion would incorrectly inflate the percentage overall agreement.</p> <p>(4) Sample was predominantly Caucasian and educated (high school or better), and therefore, the sample may be possibly be unrepresentative of the general population. These subjects may have been more motivated to participate, and therefore, better prepared to remember and to recall what was being requested of them. Sample distortion (selection) bias may have been responsible for the high levels of agreement that were found in this study.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
10. Roht et al., 1985	A household health survey of residents living near 2 hazardous waste sites in Louisiana (1981-1982) compared with an unexposed community.	Yes	Yes	Eye, respiratory, upper and lower gastrointestinal symptoms.	<p>Results of the health survey indicated that residents living in the exposed communities reported more symptoms than residents of the comparison community.</p> <p>There was a statistically significant main effect for the respondents' opinion about waste site effects on health and the reporting of associated symptoms regardless of the location of residence (pp. 426-427). For those subjects who believed that waste disposal sites affect the environment, their reports of chronic illness were 2-3x more prevalent than those individuals who did not believe in this association (p. 428).</p> <p>Meteorological and hydrologic data demonstrated that residents near the waste sites were not directly exposed to the hazardous substances which were released from the sites.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
<p>10. Roht et al., 1985 (continued)</p> <p><b>Remarks:</b></p> <p>This study does not provide conclusive evidence about the existence of 'reporting bias'. There is no measure of reported symptoms and chronic illness for the comparison communities prior to the media coverage which focused the residents' attention on health problems and local environmental hazards.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
11. Hertzman, et al., 1987	<p>A prospective morbidity survey of workers and residents from the Upper Ottawa Landfill Site in Hamilton, Ontario. The objective of the study was to determine health effects associated with occupational exposure to a landfill site.</p> <p>Workers and unexposed controls completed a health questionnaire. To validate the cases and controls self-reported health problems, their medical records were reviewed for confirmatory documentation of reported exposures, health problems, and visits to the doctors.</p>	Yes	No	Possible health problems related to landfill site exposure and visits to the doctor for any resulting health problems.	<p>There were no statistically significant differences in the distribution of confirmed, possibly confirmed and not confirmed events in either time period (i.e., pre-publicity vs post-publicity where there was intense concern in the media re: exposure and health problems regarding the landfill site).</p> <p>For example, the percent of problems lacking chart confirmation was small and non-differential between the cases (7.9%) and the unexposed controls (7%) preceding publicity regarding the hazards of landfill sites. However, post-publicity, the proportion of unconfirmed events rose 9.9% in the exposed cases and 4.5% in the unexposed controls.</p> <p>No evidence of increased physician utilization by exposed cases.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
11. Hertzman, et al., 1987 (continued)					<p>None of the conditions of interest showed trends toward overreporting among the exposed cases. Overall overreporting rates were unbiased between the study groups.</p> <p>There was no evidence for recall bias.</p>
<p><b>Remarks:</b></p> <p>(1) Response rate for exposed workers was higher (84.5%) and significantly different from the response rate for controls (71.9%). Here, the possibility exists for selection bias.</p> <p>(2) Did not consider underreporting of exposures, health problems and visits to the family physician. Studies of recall bias must assess the case-control differences in false (+) and false (-) reports of exposure, which are a function of sensitivity and specificity. This problem may have threatened the study's validity in view of the fact that 36.5% of the medical records documented a visit to the doctor which was not reported on the questionnaire.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
12. Jain et al., 1980	<p>A case-control study of the association of diet and bowel cancer.</p> <p>N=52 (26 cases of bowel cancer and 26 matched neighbourhood controls).</p> <p>Personal interview data were compared with information recorded in a dietary history questionnaire.</p>	Yes	Yes (Modest)	Mean daily intake of 13 nutrients	<p>Jain et al. (1980) concluded that the cases were more likely to decrease intake after diagnosis. Therefore, cases had a tendency post-diagnosis to underreport intake for the various food items/nutrients. The authors concluded that current diet affects reporting of past dietary patterns.</p>
<p><b>Remarks:</b></p> <p>(1) Case-control differences regarding participation rate (i.e., the possibility of selection bias) may have adversely affected the study results/ conclusions. 80 % of the eligible cases vs 52 % of the eligible controls participated the original study. A low participation rate is a potentially serious threat to the study's validity because the controls differ from the cases by virtue of the fact that they are disease free. The salience of the outcome event is not sufficient to stimulate their motivation to recall and to report past diet.</p> <p>(2) No evaluation of case-control differences on the estimates of association (odds ratio) between the nutrients and outcome (i.e., bowel cancer).</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
13. Hislop, et al., 1990	<p>A nested case-control study of the relationship of diet and breast cancer among a cohort of women from an earlier case-control study in 1980-1982.</p> <p>This study was designed to evaluate dietary recall and the presence of differential misclassification.</p> <p>N=463 (263 cases and 200 controls (i.e., neighbors/acquaintances of the case women)).</p> <p>Self-reported dietary information from a food frequency questionnaire completed in 1980-1982 was compared with data re-reported in 1986 by means of a self-administered food frequency questionnaire.</p>	Yes	No	Dietary components reported in a food frequency questionnaire	<p>The authors found little difference in the responses for both cases and controls regarding dietary recall for the distant past.</p> <p>Systematic differences were noted for the recall of recent diet by the cases. Here it was suggested that recent dietary changes would be more frequent and likely to affect recall in the cases because they have had to alter their diets as a consequence of disease and its treatment.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
14. Baumgarten et al., 1983	A validation of self-reported work histories of cases and controls in a study of the relationship between various occupational variables and the diagnostic outcome of cancer.  N=297 (274 cases and 23 controls).	Yes	No	Occupational factors	<p>1. 82 % of the subject reports agreed with the records.</p> <p>2. The extent of agreement did not differ between the subgroups defined by age, education, and social class.</p> <p>3. There was no evidence of differential reporting of occupational factors by the cases and the controls. The data provides no evidence to support the finding of no recall bias.</p>
<p><b>Remarks:</b></p> <p>(1) Sample size was insufficient. There was inadequate study power to test the specific research hypothesis.</p> <p>(2) Work histories were validated for 274 cases but only 23 controls.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
15. Weinstock et al., 1991	<p>In a case-control study nested in the Nurses' Health Study Cohort, Weinstock and coworkers assessed for the presence of recall bias in the reporting of two risk factors for melanoma - (i.e., hair color and the ability to tan).</p> <p>Cases and controls provided risk factor data prospectively (before diagnosis) and retrospectively (after melanoma diagnosis) by means of a questionnaire.</p> <p>N=459 (143 cases and 316 age-matched controls) randomly sampled from the cohort.</p> <p>Response rate: 85% cases and 81% controls.</p>	Yes	Yes	Hair colour and ability to tan	<p>The authors concluded that recall bias was observed among female nurses with cutaneous melanoma regarding their assessment of tanning ability.</p> <p>Cases differentially reported a reduced ability to tan when questioned after the diagnosis of melanoma.</p> <p>Prospective OR = 0.7 (95% CI 0.3-1.5)</p> <p>Retrospective OR = 1.6 (95% CI 0.8-3.5)</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
15. Weinstock et al., 1991 (continued)					
<p><b>Remarks:</b></p> <p>(1) The format of the questionnaire (i.e., non-identical wording of the questions) may have been responsible for the biased reporting of ability to tan.</p> <p>(2) The participants were highly motivated as evidenced by their responding to multiple questionnaires. The findings are probably not generalizable to the general population: the participants may be more or less susceptible to recall bias.</p> <p>(3) Melanoma was diagnosed before the return of the first questionnaire in 104 of the cases. According to the authors, the diagnosis of melanoma may have affected the baseline exposure history.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
16. Lindsted and Kuzma, 1989	<p>A case-control study nested within the Adventist Mortality Study in California. This study examined the relationship between diet and cancer.</p> <p>In addition, the authors examined the recall reliability over 24 years and the differences in recall between the cases and the controls.</p> <p>Subjects who completed a 21-item food frequency questionnaire in 1960 were asked to recall this diet in 1984 using a subset of the original questionnaire.</p> <p>N=216 (117 incident cases and 99 controls).</p>	Yes	No	Usual frequency of consumption of 21 foods.	<p>1. Recall scores were similar for both the cases and the controls. The mean and median food frequencies did not show systematic group differences in recall ability after the researchers controlled for factors that were possibly related to recall ability (e.g., age, education, sex).</p> <p>2. Twenty-four year recall ability was dependent on two factors: vegetarian status and the stability of ones diet. The authors postulated that vegetarians had better recall because they were more health conscious, ate fewer of the foods listed on the diet questionnaire, and were more aware of their own dietary intake (pp. 145-146).</p> <p>3. Therefore, it was concluded that the data did not provide evidence for the existence of recall bias.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
17. Linsted and Kuzma, 1989	<p>This nested-case control study was also a part of the Adventist Mortality study. The same sample of cases and controls were used to examine the determinants of long-term diet recall with respect to the following variables: vegetarian status, diet stability and selected demographic characteristics.</p> <p>N=216 (101 vegetarians and 115 non-vegetarians). Length of recall: 8 years (short-term) vs 12 years (long-term).</p>	Yes	No	Mean frequency per week of 35 foods restricted to vegetarians.	<p>The authors investigated the determinants of long-term recall and observed the following relationships:</p> <p>(1) Better recall was noted for vegetarians who had stable diets, were educated, went to church and did not watch television regularly.</p> <p>(2) For length of recall (8 year vs 24 years), diet stability, vegetarian status and education were related to recall accuracy.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
18. Fenster et al., 1991	<p>The reliability of exposure data was examined in a case-control study of spontaneous abortion in Santa Clara County, California. Because of the concern about differential reporting of water consumption in regions with publicized water contamination, detailed information during pregnancy was collected and analyzed.</p> <p>Exposure data were collected prospectively and retrospectively by means of a telephone interview and then compared.</p> <p>N=300 (100 cases and 200 controls).</p>	Yes	No	Factors possibly related to pregnancy outcome: caffeine, tap and bottled water consumption, cigarette smoking, employment, pregnancy history, occupational exposures, and exposure to video display terminals.	The authors noted case-control differences in the prevalence of exposures reported on the two occasions. "However, the degree of differential reporting was not sufficient to appreciably alter the measures of association between water consumption during pregnancy and spontaneous abortion" (p. 477).

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
18. Fenster et al., 1991 (continued)					
<b>Remarks:</b> (1) Interviewers were not blind to case-control status. Case-control differences may be related to interview bias rather than differences in recall accuracy. (2) Cases were questioned about events that occurred during the entire pregnancy. Controls were questioned on the first 20 weeks of gestation. (3) Small sample size and insufficient study power to study recall bias properly.					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
19. Drews, et al., 1990	<p>A case-control study of sudden infant death syndrome (SIDS). N=452 (226 cases and 226 controls).</p> <p>This study examined case-control differences in the accuracy of maternal recall, and evaluated the impact of maternal reporting errors on the observed measures of association (odds ratios).</p> <p>Personal interview information was compared with medical record data.</p>	Yes	No	Events which had occurred during the mother's pregnancy, labour and delivery, as well as events or sickness happening to their infants within 5 weeks of the death of the matched case infant.	<p>The authors concluded overall that case-control differences in recall accuracy were not significant to create "spurious associations with SIDS, or to bias most associations away from the null hypothesis". There were large C-C differences in the estimated sensitivity of recall. However, overall, cases did not report events more completely than controls. Controls were more likely to report events documented in their records. Specificity of recall was at least 10% higher for the controls.</p> <p>These results would seem to indicate that enhanced recall among cases is not universal across factors. These results were opposite to the results reported by Werler, et al. (1986) who noted better recall among the case subjects.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
20. Lindefors-Harris et al., 1991	<p>An analysis of exposure data from 2 independent Swedish studies to determine if response bias could explain the tendency for an increased risk of breast cancer associated with induced abortion.</p> <p>Study 1 - case-control study in which data was collected via personal interview.</p> <p>Study 2 - cohort record linkage study using registry information on abortion.</p> <p>N=828 (317 cases (breast cancer) and 512 controls) randomly selected from the Swedish population register.</p>	Yes	Yes	History of spontaneous or induced abortions, reproductive histories and contraceptive drug usage.	The authors concluded that the results of this study suggested that there was a statistically significant bias in the underreporting of induced abortions among healthy controls compared with incident cases of breast cancer.

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
21. Spengler et al., 1981	<p>A case-control study of endometrial cancer and its relationship to exogenous estrogen use.</p> <p>Personal interview data compared with hospital and/or physician records.</p> <p>N=265 (88 cases and 177 age-matched neighbourhood controls).</p>	Yes	No	Exogenous estrogen use (any dose taken for a duration of at least one month or longer)	<p>Level of agreement between interview data and medical records was similar - 83% cases vs 81% controls. Interview vs hospital record (85% cases, 65% controls).</p> <p>False (-) rate was better for cases (21% vs 35%) showing slightly better recall among the cases.</p> <p>Two thirds of the disagreement between interview and hospital records was due to women reporting usage with no record documentation (false +).</p>
<p><b>Remarks:</b></p> <p>(1) Validation of estrogen use from medical and hospital records was completed for all cases, but only 50% of the controls.</p> <p>(2) Can't assume that medical and hospital records are equally complete and reliable for both the cases and the controls. Case-control differences in the reporting of estrogen use may be an artifact related to method of data collection rather than real differences in recall.</p> <p>(3) No estimation of the impact of case-control differences on the odds ratios calculated separately for interview and record data for the subject's exposure to conjugated estrogen.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
22. Horwitz et al., 1980	<p>These authors completed two case-control studies which investigated the etiological relationship between estrogen use and the development of endometrial cancer.</p> <p>Study 1: N=238 (119 cases 119 controls with only 50 controls being interviewed).</p> <p>Study 2: N=298 (149 cases 149 controls); but 104 cases, 87 controls were interviewed.</p> <p>Personal interview data compared with data in medical records.</p>	No	No	Use of oral estrogen $\geq 3$ mg for a minimum duration of at least 6 months.	<p>There was no evidence of recall bias. Disagreements between the interview and the medical records were similar for the cases and the controls.</p> <p>The authors concluded that "The results demonstrated that the odds ratio found in a case-control study may vary considerably according to the source of data used to define exposure . . . if substantial differences are noted in proportions of people from the basic groups who are available for interview, major variation can be expected in the odds ratio".</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
22. Horwitz et al., 1980 (continued)					
<p><b>Remarks:</b></p> <p>(1) Problems were encountered in this study regarding the selection of cases and controls. Fewer controls than cases were available for interview. For example, in the samples of patients from a tumor registry, more controls had died before the interview could be conducted; fewer control patients overall could be located; more controls refused to participate.</p> <p>(2) Significantly more estrogen users were interviewed than those who did not take this drug. Availability for interview was positively correlated both with estrogen exposure and a diagnosis of endometrial cancer. The authors suggested that the availability for interview plus the increased reporting of estrogen use by cases <u>may</u> lead to a falsely elevated odds ratio. However, this was not investigated.</p> <p>(3) The refusal rate was too high. Consequently, the sample size and study power were too low to test adequately the research hypotheses.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
23. Friedenreich, et al., 1993	<p>A nested case-control study conducted within the Canadian National Breast Screening Study. The overall objective of this study was to determine if there was evidence for recall bias in the reporting of past micronutrient intake.</p> <p>N=953 (325 cases (breast cancer) and 628 matched controls (i.e., matched by age, clinic and date of enrollment)).</p> <p>Dietary data collected prospectively during enrollment (1982-85) were compared with data collected retrospectively in 1988 after the diagnosis of breast cancer. Data were collected by questionnaire.</p>	Yes	No	Dietary factors (86 food items)	<p>The authors state that the data from this study do not provide evidence for recall bias in the reporting of previous food intake.</p> <p>The accuracy of recall of food intake patterns was comparable for the case and the control subjects.</p> <p>The odds ratios for the association of the various food groups/items and the occurrence of breast cancer for the prospective and retrospective dietary data were similar in magnitude.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
24. Giovannucci et al., 1993	<p>A nested case-control study conducted within the Nurses' Health Study cohort to determine the association of diet (dietary fats) and breast cancer.</p> <p>N=902 (300 cases and 602 controls).</p> <p>Participation rates: 77% for both the cases and the controls.</p> <p>Dietary data were collected prospectively and retrospectively by means of a food frequency questionnaire.</p>	Yes	Yes	Mean daily intake of 12 nutrients.	<p>Retrospective estimates of total fat and saturated fat showed positive and significant associations between intakes of total fat and saturated fat and breast cancer. Prospective assessments, on the contrary, showed no association. The authors stated that "apparently small biases of 2-5% in mean intakes of saturated fat and red meat resulted in biases of 50% or greater in odds ratio of breast cancer between extreme quintiles of intake" (p. 508).</p> <p>The authors concluded that several features of their study indicate that their estimate of bias may be representative, if not an underestimate of the degree of potential bias in a typical case-control study.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
25. Goodman et al., 1990	<p>A case-control study of the relationship between estrogen replacement therapy and breast cancer. This study was conducted in Hawaii between 1975 and 1980.</p> <p>Menopausal estrogen histories were obtained. Data from three personal interviews were compared with physician and/or clinical records.</p> <p>N=688 (344 cases, 344 hospital controls and 344 neighbourhood controls).</p>	Yes	No	Replacement or menopausal estrogen use for at least month or longer, and other prescribed drugs.	<p>No evidence of recall bias. Agreement between exposure reports tended to be better for Japanese vs White, younger subjects, non-smoker and those of higher SES.</p> <p>Women were able to recall estrogen use with a high level of accuracy and completeness.</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
26. Floderus et al., 1990	<p>A twin study was completed to assess for the presence of recall bias in subjective reports of familial cancer.</p> <p>Both twins in a pair in which one had suffered breast cancer were asked to report the occurrence of cancer among first- and second-degree relatives.</p> <p>N=230 (115 twin pairs).</p>	Yes	Yes	The occurrence of cancer among first-degree and second-degree relatives and site-specific cancer events.	<p>The data suggests differential reporting of familial breast cancer by the cases. The cases reported more first- and second-degree relatives having breast cancer than did the twin controls.</p> <p>"Preferential reporting of events by case twins gives evidence of differential bias; moreover, the number of discordant events in relation to the number of events agreed upon gives an indication of the reliability of subjective reports on cancer in relatives" (p. 319).</p>
<p><b>Remarks:</b></p> <p>(1) Differential reports may be attributed to interviewer bias vice differences in recall. Resulting bias may be related to the method of data collection. The interviewers were not blind to the case-control status of the research subjects.</p> <p>(2) Small study population and low study power to test adequately the specific research hypotheses.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
27. Feldman et al., 1989	<p>A nested case-control study conducted within the Motherisk Program in Toronto.</p> <p>The study investigated the determinants of recall and recall bias by studying the relationship between pregnancy outcome and reported drug and chemical exposures in pregnancy. Recall accuracy was assessed using prenatal and postnatal reports of exposure obtained by interview for comparisons.</p> <p>N= 145 (33 cases and 112 controls).</p>	Yes	Yes	Drug and chemical exposures in pregnancy.	<p>The authors concluded that the data supported the existence of recall bias.</p> <p>Cases and controls differentially reported alcohol use during pregnancy. During the postnatal period, the retrospective reports of mothers with an adverse pregnancy outcome demonstrated a tendency to report significantly less alcohol consumed than the amount initially reported during the prospective pre-natal interview.</p>
<p><b>Remarks:</b></p> <p>(1) Small sample size; insufficient study power.</p> <p>(2) The impact of case-control differences on measures of association were not evaluated using the exposure data from both the prospective and retrospective reports, assessed separately and then compared, regarding the validity of the study's conclusions regarding relative risk.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
28. Paganini-Hill and Ross, 1982	<p>An exposure data reliability study mounted as part of a breast cancer case-control study which was completed in 1977-78. The original study was designed to study the association of menopausal estrogen therapy and breast cancer occurrence. Interview reports of drug use and other health-related information were compared with data recorded in medical and/or pharmacy records.</p> <p>N=334 (117 cases and 217 controls).</p> <p>The subjects were predominantly white women living in two affluent retirement communities near Los Angeles. Age range: 57-79 years.</p>	Yes	No	<p><u>Medical history</u> (gallbladder disease, hypertension, diabetes, benign breast disease, hysterectomy, and oophorectomy);</p> <p><u>Drug history</u> (thyroid medication, antihypertensives, steroids, barbiturates, and estrogen use)</p>	<p>The authors concluded that the data did not provide evidence of recall bias. Cases did not differentially recall more drug use or past diseases than controls.</p> <p>Agreement between interview and medical record was at least 90% or better for all disease conditions, height, weight, and most menstrual and reproductive variables.</p> <p>Poor agreement was noted for age at last menstrual period.</p> <p>Agreement regarding drug usage varied according to the classification of drug studied. Poorest agreement was found for barbiturates (69%) and the best for antihypertensives (89%).</p>

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
<p>28. Paganini-Hill and Ross, 1982 (continued)</p> <p><b>Remarks:</b></p> <p>(1) Findings were not generalizable. The subjects were elderly, affluent, educated and very health conscious. It is postulated that recall accuracy may be dependent upon education, SES and personal motivation to participate, recall and report the required information.</p> <p>(2) Data comparisons were restricted to those individuals with exposure data available from both sources.</p>					

Table 2 (continued): Findings of Selected Studies of Recall Accuracy and Recall Bias (Differential Exposure Misclassification)

References	Study Design	Consideration of Exposure Misclassification /Recall Bias	Evidence of Exposure Misclassification /Recall Bias	Exposures/ Conditions	Results
29. Lindsted and Kuzma, 1990	<p>A case-control study nested within the Adventist mortality study in California (i.e., the relationship of diet and cancer). Responses to a food frequency diet questionnaire were compared.</p> <p>N=406 (181 incident cancer cases diagnosed between 1976-1984, and 225 controls randomly selected from the same population).</p> <p>Analysis was restricted to non-vegetarians.</p>	Yes	Yes	Mean frequency per week of 35 food items.	<p>For breast and gastrointestinal cancer subjects, the authors observed differential recall (recall bias) for some of the 35 food items. Dietary changes due to cancer diagnosis affects the accuracy of recall of past dietary patterns. Current diet may bias recall of distant diet.</p> <p>The authors also concluded that "certain cancer sites can lead to differential recall ability" (p. 400).</p> <p>Overall, there was no evidence for the existence of differential recall of past diet by the cases and the controls.</p>

## Chapter 3

### RESEARCH DESIGN AND METHODS

#### 3.1 Study Objectives

This research study was designed to investigate some of the methodological issues involved in the design, conduct and analysis of a case-control study. The overall objective was to determine the reliability and validity of retrospectively collected exposure information in a case-control study of breast cancer. The study was designed specifically to determine if the retrospective, post-diagnostic reports of exposure provided by the cases and the controls were systematically different, to assess the impact of any resulting exposure misclassification on the estimates of relative risk, and (most importantly) to develop and to evaluate a 'Validity Scale' as a possible design standard for the measurement and the statistical control of non-differential and differential exposure misclassification in future case-control studies. The specific objectives were:

1. To determine the reliability (reproducibility) of exposure data in a case-control study of breast cancer by comparing the prospective and retrospective case-control reports of exposure;
2. To confirm or refute the presence, direction and magnitude of exposure misclassification, and to determine whether the imprecision in exposure classification is relatively the same (non-differential) or systematically different (differential) for the cases and controls;
3. To determine whether the point estimates of the exposure-disease

odds ratios for the study variables are significantly different for the prospective and retrospective case-control reports of exposure;

4. To identify the factors that may be responsible for case-control differences in exposure recall, and the subsequent occurrence of exposure misclassification;
5. To determine if the type of control group chosen for the case-control comparisons results in differential exposure misclassification and the generation of biased odds ratios;
6. To evaluate the use of an anamnestic equivalent control group to minimize differential recall accuracy and to produce valid case-control comparisons;
7. To determine the relative magnitude and direction of misclassification for different study factors;
8. To determine if an 'Exposure Data Validity Scale' can be developed to estimate the presence, magnitude and direction of exposure misclassification, and then to adjust the relative risk (odds ratio) estimate for any resulting bias.

This research protocol was approved by the University of British Columbia Behavioural Sciences Screening Committee For Research and Other Studies Involving Human Subjects as well as the Academic Subcommittee of the Screening Mammography Program of British Columbia.

### **3.2 Study Design**

A nested case-control or case-control within a cohort design was used to address the research questions posed in Section 3.1 above. In a 'nested case-control study' the cases and controls are both selected from a defined cohort

(Last, 1995, p.111). All the cases are included in the study, and the controls are selected for the cases on the basis of inclusion criteria by means of incidence sampling (Checkoway et al., 1989, p.68). This sampling procedure involves considering each case separately, and then choosing controls from the cohorts who were at risk at the same time that the subject was selected as a case. Here, exposure histories are obtained on the cases within the study cohort, and the randomly selected controls (Pearce, 1989; Mantel, 1973).

### **3.3 Recruitment of Study Subjects**

The study groups were selected from the cohort of women residing in British Columbia who had enrolled voluntarily in the provincial Screening Mammography Program (SMP BC) between 1990 and 1992. The women who were eligible to participate in SMP BC were 40 years of age and over, did not have a previous history of breast cancer, benign breast disease or breast surgery, had not undergone mammography in the previous 12 months, and were not pregnant or breast feeding. Women under 40 years of age were also admitted to the screening program if their family history made them a high risk candidate for breast cancer development, and if they had been referred to the screening program by their physician.

To be eligible for inclusion in this research study, subjects required fluency in English, both verbal and written. Proxy respondents were not permitted; nor was reliance on other family members, coworkers or friends to explain the purpose of the study, or to translate questionnaire items or the subject's responses to these questions.

In order to select and recruit study subjects, the researcher obtained from SMP BC a monthly computer file listing all current program participants together with limited personal information, which included: the subject's unique

identification number (to protect anonymity), name, screening clinic, date of mammogram, birthdate, and mammogram results (i.e., abnormal or normal). SMP BC also provided the results of the diagnostic workups for the women with abnormal mammograms. The diagnostic results were received by the researcher on average 6 months after the date of the screening visit and the occurrence of an abnormal mammogram. This information was used to build the sampling frame and then to select the study groups.

The addresses and telephone numbers for potential subjects were obtained by the researcher by accessing the SMP BC Computer Data Base.

Following the selection of cases and eligible controls, a 'Letter of Information' (Appendix 4) was sent to the potential study participants. The letter was prepared on SMP BC letterhead and signed by the Executive Director SMP BC and the Senior Epidemiologist of the BC Cancer Agency. This letter was used for two purposes: 1) to explain the researcher's legitimate access to the SMP BC study population; and, 2) to increase response rates. Levy and Lemeshow (1980) state that endorsement by an agency whose field of interest includes the subject matter of the survey might increase response rates overall (p.262).

Women were told that the present investigation was being done in collaboration with the Screening Mammography Program of British Columbia. In addition, they were informed that the purpose of the study was to determine the best way of collecting medical information in order to better study factors associated with disease, and ultimately to achieve the goal of disease prevention in British Columbia.

The letter also described 'how' women were selected, exactly 'what' was required of them should they agree to participate, as well as the specific time involvement for each participant. The issues of confidentiality and voluntary participation were stressed. Women were informed that their refusal to

participate would in no way affect their ongoing involvement with the SMP BC. The 'Letter of Information' told the potential subject that the researcher would contact her by telephone to determine her willingness to participate, and to answer any questions that she may have regarding the study, or her participation.

When information letters were returned as 'undeliverable' because the individual had moved, every attempt was made to locate a new address. Another letter was sent when an alternate address could be successfully traced. If the second letter was unsuccessful, the potential subject was classified as a non-respondent, and further follow-up was impossible.

The women were then contacted by telephone 7 to 10 working days after the 'Letter of Information' was mailed. If the researcher was unable to contact the subject on the first call, a telephone contact sheet was started, and repeat telephone calls were made at different times of the day and evening, up to 10 callbacks per subject, over a two week follow-up period. If these attempts were unsuccessful, the subject was designated a non-respondent.

For those subjects with a telephone answering machine, a contact message was left reminding the women of the 'Information Letter', the name and telephone number of the researcher, and a request for them to contact the investigator at their convenience. Additional telephone calls were made if the initial call was not returned within 48 hours.

If another member of the household answered, and indicated that the subject was not available to take the telephone call, the researcher identified herself, explained the purpose of the call, determined when it would be best to call back, and left a message that the researcher had called, and that she would attempt to call back at the time suggested.

If the telephone number was incorrect or out of service, the researcher verified the telephone number on the SMP BC data base, and if that was

unsuccessful, the telephone directory, directory assistance, and the data base of the British Columbia Cancer Agency (for cases who may have received treatment post-diagnosis) were checked. Every attempt was made to contact the subjects selected for recruitment. Women who could not be contacted were designated as potential non-respondents (no current telephone contact number).

When contact was made with the potential subject, the researcher explained the purpose of the research protocol in detail, described the time required to complete the questionnaire and what was expected of each participant, addressed any questions or concerns raised, and then sought verbal consent to participate. A woman's verbal agreement to receive and to complete a second questionnaire was considered as implied consent to participate. Further written consent was not required. For those subjects who agreed to participate, a correct mailing address was obtained, and the research package was forwarded to them for completion.

For those subjects where no telephone contact was possible (initially classified as non-respondents), the research package was sent to the individual's last known address recorded in the SMP BC data base. Returned packages meant that no follow-up was possible for these subjects. These individuals were then coded as non-respondents.

### **3.4 Case Definition and Selection**

From the results of the initial mammography and subsequent diagnostic workup for those subjects with an abnormal mammogram, the researcher selected one case group and two control groups.

A **case** is defined in this study as any woman enrolled in the screening mammography program who was subsequently diagnosed with a histologically

confirmed invasive carcinoma of the breast. Every case subject was approached by the researcher for participation in this investigation.

As part of the initial study design, only incident cases (i.e., newly screened and diagnosed cases of invasive breast carcinoma) were to be recruited. However, in September 1991, it became necessary to recruit prevalent cases (i.e., cases who had been diagnosed prior to the beginning of this study) when SMP BC changed the content and format of the initial (first visit) screening questionnaire. This decision was based on three considerations, the study design (i.e., the data collection procedure used), power calculations, and the completeness of responses for the comparison of pre-diagnostic and post-diagnostic exposure reports for reliability (i.e., consistency/reproducibility).

In order to complete a valid exposure data reliability study, the procedure used to collect the prospective and retrospective exposure data must be the same. Otherwise, the observed case-control differences in recall accuracy, and the subsequent occurrence of exposure misclassification could also be attributed to the way in which the exposure information was collected, and not just to group (case-control) differences regarding their motivation and ability to remember and to report past exposures.

The recruitment of additional cases and matched controls ensured that the total sample size was sufficiently large to generate the statistical power necessary to answer the research questions. Without the addition of the 42 prevalent cases and the 84 matched controls, the study power was estimated just under 70% rather than 76% for the total sample size, including both the incident and prevalent cases and their matched controls. This will be discussed further in Section 5.4.

### 3.5 Control Groups: Definition and Selection

The most critical and controversial problem in the design of retrospective case-control studies is the selection of a suitable control group for valid case-control comparisons (Sartwell, 1974; Cole, 1979; Stavracky and Clarke, 1983; Spitzer, 1985; Linet and Brookmeyer, 1987; Smith et al., 1988; Kramer, 1988). Researchers have remarked that the use of inappropriate controls can pose serious methodological problems including the generation of study biases (i.e., information bias, response (recall) bias, surveillance bias, referral bias, etc.), and the possible invalidation of the study conclusions (Schlesselman, 1982; Kramer, 1988; Smith et al., 1988).

Therefore, researchers and methodologists, in discussing the relevant design issues for case-control studies, generally recommend the inclusion of multiple control groups (at least two) to minimize the potential for biased estimates of effect (odds ratios), as well as to demonstrate the consistency of the study findings (Mantel and Haenszel, 1959; Ibrahim and Spitzer, 1979; Smith et al., 1988). For example, the use of two control groups is analogous to doing two separate studies: should an association be found when different controls are used, the investigator can have more confidence in the study conclusions regarding the exposure-disease association (Greenland et al., 1981).

Based on the previous observations and recommendations, two control groups were recruited for this study, and were defined as follows:

- a. Control Group One - women participating in the SMP BC who initially had an abnormal mammogram, but after diagnostics were found to be negative for invasive carcinoma of the breast (i.e., abnormal mammogram, no breast cancer); and,

- b. Control Group Two - SMP BC participants with a normal mammogram and no breast cancer (healthy controls).

Jick and Vessey (1978) stress that "the series of control subjects should be chosen in such a way... that they are comparable with the cases... except that they do not have the illness under study" (p.3). In addition, they argue that "the method of obtaining information on exposure must be very closely similar in the cases and controls" (p.5). Control Group One (women with an abnormal mammogram, but no breast cancer) meet the requirements detailed by Jick and Vessey for the selection of a control group 'comparable' to the cases. Control Group One's comparability can be discussed with reference to the following criteria:

- a. anamnestic equivalence;
- b. equivalence re: the impact and salience of an event;
- c. diagnostic surveillance comparability; and,
- d. equivalency re: disease ascertainment.

The first criterion, anamnestic equivalence, is needed to avoid the bias that would occur as a result of differences in subjects' memories of past exposure. Horwitz and Feinstein (1979) state that 'anamnestic equivalence' is needed because it minimizes the inequalities among the cases and controls with respect to their ability or their motivation to scrutinize their previous exposure histories for the suspected risk factor(s). Anamnestic equivalent controls have the same incentives to participate, and to recall the previous exposures and events postulated to be associated with the outcome disease. It is also assumed that these

controls would be likely to recall and report more exposures than equally exposed normal population controls (Ellwood, 1988; Werler et al., 1989).

The subjects' motivation to review carefully their past exposure is strongly linked to the overwhelming need to understand the cause(s) of their disease (Sackett, 1979).

The second criterion (the salience and emotional impact of an event, such as an abnormal mammogram and the possibility of breast cancer) acts as a powerful stimulus to drive the cognitive processes of memory. Cochran (1953) states that control group selection is important and that "there is a need for a control group in whom the emotional impact on memory is roughly of the same order of magnitude as in the cases" (p.687). Feinstein (1979a) adds in his discussion of the prevention of anamnestic recall bias, "... that if the disease itself acts as a stimulus that makes the subject carefully review the possibility or degree of antecedent exposure, the control group should be stimulated to perform a similar review" (p.38).

In this investigation, it is presumed that the initial fear of the same diagnosis of breast cancer evoked a similar form and degree of selective recall in the women of Control Group One. Consequently, it was then postulated that the estimated relative risks that would be generated from case-control group one comparisons for the various study factors would not be distorted by the presence of recall bias (i.e., selective recall which would result in differential exposure misclassification). Therefore, an anamnestically equivalent control group was chosen to determine if it was capable of providing protection against differential exposure misclassification bias in a case-control study of breast cancer. In order to demonstrate that recall bias has been minimized, a second control group (normal mammogram) was included to indicate whether the estimates of effect (i.e., risk

estimates) differed for case-control group one and case-control group two comparisons.

The last two criteria -- equivalent diagnostic surveillance and complete disease ascertainment -- should control for biases such as selection (sample distortion), diagnostic referral and surveillance biases, while protecting against the possibility of disease misclassification. Because the controls have gone through the same diagnostic procedures as the cases and were found to be free of disease, there is increased confidence regarding the 'purity' of the control group with respect to the absence of disease (Cole, 1979; Horwitz and Feinstein, 1979; Feinstein et al., 1981). Once again, selective recall (recall bias) is minimized or prevented because the control group was examined using similar diagnostic procedures and was questioned to the same extent and intensity as were the cases. Both the cases and controls had a similar 'mind set' and motivation for remembering and reporting past exposures (Cole 1979, p.23).

In summary, multiple control groups were used in this research project to address research questions 5 and 6, to minimize the possibility of biased results, and to demonstrate the consistency of the study's findings.

### **3.6 Matching**

The controls were matched to the cases on four variables considered to be potential confounders for the study outcome (breast cancer): age (+/- 5 years), clinic, ethnicity (when known) and date of mammogram (year and month +/- 1). Both a woman's age and ethnicity have been strongly associated with the occurrence of breast cancer as well as other factors (exposures) considered to be relevant in the etiology of breast cancer (i.e., postmenopausal estrogen use, diet, standard of living, etc.).

Matching by clinic controlled for differences due to geography (urban versus rural), as well as for any clinic differences in breast cancer detection rates and/or clinic differences in the collection and data entry of the demographic, medical and exposure information provided by the screening mammography participants during their first clinic visit.

The matching of controls to cases on the date of mammogram controlled for the time interval between the administration of the pre- and the post-diagnostic questionnaires. Generally, this time period corresponded to the average amount of time required for post-mammography diagnosis, treatment and recuperation for the cases. Keeping the time between the administration of the two questionnaires comparable for the three study groups controls for the length of recall, minimizes any conditioning effects related to any response sets or memory of responses provided previously in the screening mammography enrollment questionnaire, while maximizing the likelihood that cases were well enough to participate fully in this study, and that their ability to remember was not negatively affected by poor health. The consideration of the time between the administration of the two questionnaires was important to the prevention of a high refusal rate.

For the incident and prevalent cases, the average length of time between the administration of the two questionnaires varied, and was approximately 6-10 months and 12-18 months respectively.

Four controls (i.e., two women per defined control group) were randomly selected for each case, without replacement, from all possible eligible controls in the sampling frame (Levy and Lemeshow, 1980).

### **3.7 Sample Size and Power Calculations**

When this study commenced in May 1990 there were three mammography centres in operation in British Columbia. At the rates of 20 and 4 per one

thousand women screened for an abnormal mammogram and breast cancer respectively, it was anticipated that approximately 120 eligible cases and 480 eligible controls (i.e., those subjects with an abnormal mammogram, but negative diagnosis for breast cancer) could be recruited in one year at the three centres. During the study period, three additional centres were opened, increasing the number of subjects in the sampling frame who were available to be recruited. An unbalanced design was chosen, and the selection ratio of controls in each control group to cases was established at 2:1. Cole (1979) recommended that when the number of cases available for study is limited, the selection ratio of controls to cases should be increased but ultimately stay within the bounds of 4:1. Increases beyond this selection ratio do not result in any further increases in statistical power (p.22).

With 234 cases and 468 controls per control group, assuming a two-sided test at  $\alpha=0.05$ , there was approximately 76% power to detect odds ratios of 1.6 or greater with a further assumption of a prevalence of responses in the controls of 0.25 (Schlesselman, 1982, pp. 150-152). The sample size and resulting study power was considered adequate to address the study's research questions. The actual power calculations for this study can be found in Appendix 9.

### **3.8 Data Collection Procedures**

#### **3.8.1 Questionnaire One**

As part of the SMP BC registration, women completed a standard, self-administered questionnaire (Appendix 3). Women routinely provided demographic information, reproductive and medical histories, and information on other lifestyle factors possibly related to breast cancer development. This information was coded by SMP BC clerks, and then entered into the SMP BC data

base. In this research protocol, the SMP BC screening questionnaire will be referred to subsequently as 'Questionnaire One' (Q1). The responses to Q1 items form the prospective, pre-diagnostic reports of exposure.

It must be noted that the researcher did not have access to the individually completed Q1 questionnaires, but only to the computer data base with the women's responses already coded. Therefore, it was impossible to assess the quality and completeness of the data entered. Quality control checks regarding coding errors also were not possible. Q1 information for all study participants was downloaded from the SMP BC data base to floppy disks by an SMP BC programmer. These data were reviewed and re-coded (when necessary) for use in this study by the researcher.

### **3.8.2 Questionnaire Two**

A second questionnaire (Q2), identical to the first (Q1) with the exception of the additional items comprising the 'Exposure Data Validity Scale', was sent to those cases and controls who had agreed to participate. Q2 (Appendix 6) was mailed to the study participant with a Study letter (Appendix 5) and a self-addressed, postage-paid envelope, together with instructions to complete Q2, and to return it at her earliest convenience. The date when the research package was mailed was recorded for follow-up of non-respondents. The subjects were provided with the name and telephone number of the researcher, and were encouraged to contact her if they required any further information about the study, or clarification of any of the questions included in Q2. The nature and frequency of written correspondence, and any telephone contacts with the study subjects will be discussed in Chapter 4 when considering the study findings.

A high response rate of at least 80% was expected, based on the experience of the BC Cancer Agency (BCCA) and SMP BC.

### **3.9 Procedures for Handling Non-Respondents**

If the questionnaire was not returned within 4 weeks of the date of mailing, a second research package and 'reminder letter' (Appendix 7) were sent to the non-respondent. The 'reminder letter' encouraged participation by emphasizing the importance of the study (i.e., disease prevention in British Columbia), and the need for a high response rate. A requested return date was included in the letter to encourage an immediate response. Two weeks after the second mailing, a telephone call was made to all persistent non-respondents. This was the final attempt by the researcher to address any additional concerns and to encourage study participants to complete and to return the questionnaires. If a subject indicated that she had decided not to participate in the study, the researcher determined the reason(s) when possible, and then recorded this information for later analysis. The reasons for non-response/non-participation are detailed in Table 4 (p.226).

### **3.10 Data Handling and Analysis**

All data were analyzed on the University of British Columbia mainframe computer (Unix System) using the Statistical Package for the Social Sciences (SPSS). The correspondence analysis program (CA) in BMDP was used for the selection of 'plausible' risk factors for inclusion in the 'Exposure Data Validity Scale', and for the estimation of the specific weighting factors for included scale items.

For all subsequent analyses, Q1 data (the prospective reports of exposure) were selected as the 'gold standard' for Q2 comparisons. Although Q1 data did not necessarily represent 'true' exposure status, they were considered the best estimate for two reasons: 1) exposure data were collected prospectively and "approximated the information collected in a cohort study" (Mackenzie, 1986,

p.74); and, 2) the exposure data recorded prospectively before diagnosis were assumed to be less susceptible to the effects of the knowledge of case-control status on (exposure) recall accuracy.

### **3.10.1 Data Coding and Entry**

For each study subject there were three response (data) files which included: 1) the prospective reports of exposure for the various study factors (Q1); 2) the retrospective exposure reports (Q2); and, 3) the responses to the items contained in the 'Exposure Data Validity Scale'. The data files were identified by the subject's unique SMP BC study number.

Coding policies were established by the researcher and are described in the codebook (Appendix 8). As noted previously, Q1 data had already been coded and entered into the SMP BC data base by the screening program clerks. Access to the original registration questionnaires was not possible. SMP BC provided Q1 data for all respondents and non-respondents on computer disks for use in this study. The Q1 responses were reviewed, recoded (when necessary) into the format required for this study, and entered on a coding sheet for data entry. The researcher, who was neither blind to the study hypotheses or to the case-control status of the subjects, prepared Q1 responses for data entry.

Q2 was coded by an independent coder. The coder was trained by the researcher and was blind to the identity of the participants, the subject's group status, the study objectives, as well as to the Q1 response set. The researcher worked closely with the coder, answering questions and clarifying any problems that occurred during the coding process. To ensure that Q2 response sets were being coded properly, the researcher randomly selected ten percent of the questionnaires from each study group and independently coded them to check for any errors. The resulting error rate was negligible.

Once coded, Q1 and Q2 data were entered into a mini-computer by one experienced data entry clerk, who identified eighteen incorrect numerical codings for the combined Q1 and Q2 response sets during the data entry process. The researcher reviewed and corrected all of these problem entries prior to data analysis.

### **3.10.2 Analysis of Non-Respondents**

The first stage of the data analysis involved the description of the study groups with respect to the various categories of data collected in Q1 and Q2.

Next, the respondents and non-respondents were compared on all the variables included in Q1 (i.e., demographic information, reproductive and medical histories, anthropometric variables and risk and lifestyle factors possibly related to breast cancer). An analysis of respondents versus non-respondents was completed to assess the comparability between the two groups, and to determine any group differences that might be attributable to group differences in recall accuracy and the subsequent occurrence of non-differential and differential exposure misclassification, and the possible invalidation of study conclusions.

Q1 responses were available on all potential study subjects (both respondents and non-respondents) because the SMP BC screening questionnaire (Q1) was completed by all participants as part of their SMP BC enrollment.

### **3.10.3 Test-Retest Reliability Analysis**

In order to determine the ('test-retest') reliability (reproducibility) of prospective and retrospective reports of exposure, and to assess if there were any significant case-control differences in recall accuracy, a correlation analysis was completed independently for Q1 and Q2 responses, for each study factor and

study group. The correlation analysis allowed the measurement of the overall agreement, and the strength of the relationship between the two data sources. Pearson product moment correlations and Spearman rank correlations were estimated for the continuous and categorical/ordinal study variables, respectively, that were included in the study. This analysis was completed only for those subjects who provided exposure information for a particular study factor, at **both** times -- prospective (Q1) and retrospective (Q2). The median percentage of missing responses for the calculation of the test-retest reliability at the item level was 6%, and ranged from 0% to 14%. In addition, there were no significant group differences in missing information ( $p > 0.80$ ).

The correlation study was conducted as a preliminary analysis to determine the magnitude of agreement between the prospective (pre-diagnostic) and retrospective (post-diagnostic) reports of exposure. It was noted in the literature reviewed that the resulting estimates of agreement might be subject to error (i.e., overestimation) because the estimated correlations did not consider the occurrence of chance agreement (Mackenzie, 1986, p.72).

The value of the correlation coefficient may range from +1.00 (a positive association) through 0.0 (indicating no relationship) to -1.00 (a negative association). The strength of the correlation coefficient, (i.e., the agreement between the prospective and retrospective exposure reports may be interpreted by using the following guidelines (Munro and Page, 1993; Colton, 1974):

0.00 - 0.25	little or no relationship
0.26 - 0.49	low agreement/correlation
0.50 - 0.69	moderate agreement
0.70 - 0.89	high level of agreement
0.90 - 1.00	very high agreement/correlation

### 3.10.4 Kappa Analysis

Next, the kappa statistic ( $k$ ) was calculated for each study factor to provide another index of overall agreement (reproducibility) between Q1 and Q2 responses. The advantage of the kappa statistic is its ability to correct for the extent of agreement that would be expected on the basis of chance alone (Cohen, 1960; Fleiss, 1981; Rosner, 1990). In other words, the kappa coefficient takes into consideration the effects of chance in the assessment of the reproducibility of Q1 (prospective) and Q2 (retrospective) reports of exposure, and it therefore avoids overestimating the degree of agreement between the two exposure assessments. Kramer and Feinstein (1981) note that  $kappa=0.0$  when the observed agreement equals that expected by chance (i.e., the responses measured on Q1 and Q2 are completely independent),  $+1.0$  if the two responses agree perfectly, and  $<0.0$  if the observed agreement is less than that expected by chance.

In order to evaluate the strength of the agreement, the classification system recommended by Landis and Koch (1977) was used with minor modifications (i.e., adjacent categories were combined to produce three rather than five levels of reproducibility). Specifically, kappa values  $< 0.4$  represent poor to marginal agreement (reproducibility);  $(0.4 - 0.75)$  moderate to substantial agreement (reproducibility); and,  $> 0.75$  excellent agreement (reproducibility). A kappa statistic can only indicate the reliability or reproducibility of Q1 and Q2 reports of exposure. When the kappa coefficient is low, and indicative of inconsistency between the prospective and retrospective exposure reports, nothing further can be said about the nature of the inconsistency, that is, whether or not there has been an addition or deletion of exposure information post-diagnosis (Q2) by the cases and/or the controls. Therefore, under these circumstances, a McNemar chi-square was calculated to reveal the presence of any resulting discordance, and to specifically delineate the nature and direction of the changes in exposure reported

after the diagnosis of breast cancer. As noted previously, Q1 responses were used as the criterion ('gold standard') against which Q2 exposure reports were compared.

### **3.10.5 McNemar Analysis**

According to Dixon (1992), McNemar's chi-square is a test of symmetry which is used in repeated measures or matched designs, to determine if the frequency of change in one direction (i.e., the addition of exposure information post-diagnosis) is equal to the frequency of change in the other direction (i.e., the deletion of exposure information). Fenster et al. (1991) describe the statistic as a measure of "directional discordance or systematic bias" (p.481). In this particular study, it could be used to detect differential misclassification bias (recall bias) which can be conceptualized as a **shift** in the self-reported exposure status for the study variables which is associated with the knowledge of the diagnosis of breast cancer.

In simple terms, the McNemar chi-square reduces to a test of equality (of the frequencies) of the two off-diagonal cells in a 2x2 contingency table. A significant McNemar would indicate a lack of symmetry, denoting a greater change in one direction than the other -- either the addition or deletion of exposure information, or change in adjacent or extreme categories for the specific study factor on Q2 (Dixon (BMDP), 1992).

### **3.10.6 Prospective and Retrospective Relative Risk Assessments**

In this phase of the data analysis, the odds ratios and 95% confidence intervals were estimated for Q1 (pre-diagnostic) and Q2 (post-diagnostic) exposure prevalences. All available data were used in the estimation of the (exposure-

disease) odds ratios; exclusion of respondents due to incomplete exposure information (Q1 versus Q2) did not occur.

The point estimates of the prospective and retrospective odds ratios and their respective confidence intervals were then compared to determine if there was any evidence of differential reporting of past exposures by the cases and controls. As noted by Friedenreich (1990), bias would be evident if there were systematic differences in the direction and magnitude of the odds ratios estimated for the retrospective, post-diagnostic reports of exposure.

By comparing Q1 and Q2 odds ratio estimates, and their respective confidence intervals, it was possible to determine if the prospective and retrospective exposure reports were equivalent or significantly different in their conclusions about the exposure-disease associations for the various study factors and the outcome event (the diagnosis of breast cancer). In other words, a determination was made as to whether or not the study conclusions remained the same when Q1 and Q2 odds ratios and 95% confidence intervals were compared.

To determine if the type of control group chosen for case-control comparisons results in differential exposure misclassification and the biasing of the odds ratio, this analysis was repeated for the following study group comparisons: 1) case versus control group one ('anamnestic equivalent controls' - abnormal mammogram, no breast cancer); and, 2) case versus control group two (normal, healthy controls).

### **3.11 Stages in the Development of an 'Exposure Data Validity Scale'**

As discussed previously in Section 2.8, Raphael (1987) proposed the construction and use of a validity scale to measure and to control for differential exposure misclassification (recall bias) in case-control studies. Composed of exposure variables that are plausible but unrelated to the disease under study, the validity scale is designed to estimate the propensity of cases and controls to differentially report (i.e., the tendency to overreport or underreport) past exposure. Raphael (1987) points out that the magnitude and direction of exposure misclassification (recall bias) can be estimated by comparing the "total validity scale scores for the cases versus controls" (p.169). The overall validity scale score which is "a function of each respondent's recall bias" (Raphael, 1987, p.169) is then used to correct main study estimates of association (odds ratios) for the risk factors under study.

In order to develop the validity scale suggested by Raphael (1987), for use in this nested case-control study of breast cancer, important decisions had to be made regarding which exposure variables should be included as the 'dummy' exposures, how to 'rate' the perceived etiologic importance of each factor (i.e., a measure of risk plausibility), and finally, how to 'aggregate' the individual responses of cases and controls to these variables into a single summary score to assess for the presence and magnitude of differential exposure misclassification (recall bias). The development and evaluation of the validity scale involved four steps: choosing the candidate (exposure) variables for inclusion in the validity scale, development of a questionnaire for the assessment of the 'plausibility' of the candidate variables, the administration of the validity scale to participant cases and controls, and the evaluation of the scale. These four steps are described in Sections 3.11.1 - 3.11.4.

### 3.11.1 Search for Candidate Variables

The most important consideration in the development of the exposure data validity scale was choosing the candidate (exposure) variables for inclusion in the validity scale. These variables must be perceived as 'plausible' (etiologically relevant to breast cancer development) by the group of subjects for whom the scale is intended (i.e., participants in the SMP BC program), but must also be 'unrelated' to the study disease (i.e., no association with breast cancer). Raphael (1987) stated that these variables must be 'sensitive'; that is, they must be capable of stimulating the 'search for cause cognitive processes' which are postulated to exist, and assumed to be responsible for differential reports of past exposure by the cases and controls.

This last criterion (unrelated to disease causation) may be very difficult to achieve in reality due to the contradictory evidence that exists in cohort and case-control studies regarding cancer etiology and the delineation of definitive factors related to its occurrence (Mayes et al., 1988). As such, this may be an overall limitation to scale development in this particular instance. Multiple items were necessary for scale construction because with several factors, the variability of response to any single item tends to average out over the array of items used.

Several options existed for the selection of validity scale (exposure) variables. These included: a review of previous empirical data (i.e., the results of etiologic case-control or cohort studies) which demonstrated that a particular candidate variable was 'ruled out' for breast cancer development, general clinical beliefs about the natural history of breast cancer, personal experience and beliefs, or the recommendations of 'experts.' In this study, two of these options were used to select candidate variables. First, a comprehensive review of the literature was completed. The source material for the review consisted of research studies (i.e., etiologic studies of breast cancer) completed since 1980, and published in refereed

journals in the fields of medicine, nursing, epidemiology, environmental and occupational health. Variables which were found to be unrelated to breast cancer ( $OR=1.00$ ), as well as those in which the etiologic relationship was evaluated as inconclusive because of methodological limitations in the study conducted, were selected for evaluation and for possible inclusion in the validity scale. A total of 45 candidate variables were extracted from this literature review. Second, the senior epidemiologist at the British Columbia Cancer Agency was consulted regarding variable selection: he reviewed the cogency of the variables suggested from the literature review, and suggested alternatives based on his own, and others' research experience at BCCA. The list of candidate variables was then reduced to the 33 variables.

### **3.11.2 Selection of Validity Scale Exposure Variables and Assignment of Weighting Factors**

The next phase of scale construction involved the development of a questionnaire for the assessment of the 'plausibility' of the candidate variables for breast cancer. The questionnaire (Appendix 2) consisted of two sets of variables: 'true' risk factors for breast cancer, and the 33 tentative validity scale exposures and conditions. Subjects were asked to rate the etiologic importance of the exposures, that is, to indicate "how important" they felt the factor was as a risk factor for the development of breast cancer in women. The participants indicated their choice by using a four-point scale which consisted of the following response statements: 1) very important (HIGH RISK); 2) moderately important (MODERATE RISK); 3) somewhat important (LOW RISK); and, 4) not important (NO RISK).

The assignment of weights to the candidate variables is an important step in scale construction. Here, the aim is to choose the most plausible (dummy)

variables to be included in the scale. Different respondents will probably have different opinions about the magnitude of risk posed by each item. The weightings will therefore account for the variability of responses regarding the etiologic importance of the validity scale items.

A pilot study was conducted to optimize the length (amount of time to complete the required task), readability and clarity of the questionnaire, as well as to determine any instructional or format problems that could affect the administration of the questionnaire. Most importantly, the pilot study determined whether or not the subjects understood 'what' was required of them (i.e., an evaluation of each exposure and condition as to its perceived importance in the development of breast cancer in women). A sample of students (graduate and undergraduate), faculty and staff at the University of British Columbia (N=30) was recruited to pilot the questionnaire. Overall, the questionnaire met the previously stated requirements. Feedback indicated that the questionnaire length was satisfactory, taking approximately 10 minutes or less to complete. A few subjects (N=6) expressed some concern about the stated purpose of the questionnaire; they believed that the real purpose of the questionnaire was to test their knowledge about the risk factors for breast cancer, not just their personal evaluation of the relative importance or probability that each exposure is a risk factor for breast cancer. They suggested that some participants would refuse to complete the questionnaire if they felt that their knowledge was being assessed, especially if they lacked such information. Consequently, the questionnaire instructions were reviewed and modified slightly. During the recruitment phase, the researcher conveyed to the potential study subjects that the purpose of the questionnaire was simply to obtain their personal assessment (perception/intuition) of relative risk, and definitely was not a measure of knowledge about 'risk versus no risk' for the included exposures and conditions.

A convenience sample (N=147) of women attending the Vancouver Screening Mammography Centre completed the revised questionnaire. These women were a subsample of those subjects for whom the eventual validity scale was intended. The responses of the 147 study participants were cross-classified in a two-way contingency table according to 5 levels of risk and 33 risk factors. Multivariate correspondence analysis was used both to select the variables considered by the study subjects to be responsible for breast cancer etiology, and to assign specific weightings.

These differentials (weights) reflect the ascribed etiologic significance or importance placed on the factors by the participants. It must be remembered that some of the variables are considered to be more significant than others; and this special emphasis, or weight, must be reflected in the final scale. In addition, it must be remembered that the assigned weightings were intended to be a measure of the exposure variable's sensitivity to stimulate the 'search for cause' motivation in the cases.

Therefore, a mathematical model was then selected to describe how the validity scale items and their specific weights were to be aggregated. In this study, an additive model was chosen for the estimation of a summary validity scale score (i.e., a summation of the individual items as function of their weighting factors). In this way, the summary score was a result of the etiologic strength of the different items which made up the aggregate score.

From the 33 candidate variables, 20 exposures and conditions were selected for the 'Exposure Data Validity Scale', and were included in Q2.

### 3.11.3 Administration and Analysis of the Validity Scale

The next step involved the administration of the validity scale to the participant cases and controls, and the calculation of the VSSCORE (overall validity scale summary score) for each study group (i.e., the cases, anamnestic equivalent control group and the normal (healthy) controls). The total validity score (VSSCORE) was defined as an aggregate score expressed as a sum of the product of the exposure rating (present/absent) and weighting factor for each scale item summed over all the individuals in the comparison group (cases versus controls). The research question to be addressed here is the statistical equivalence or difference of the estimated VSSCOREs for the three comparison groups. The estimated magnitude and direction of any statistically significant differences would subsequently be used to adjust the exposure-disease odds ratios for the study factors. Neugebauer and Ng (1990) and Kopec and Esdaile (1990) discussed possible limitations to the validity scale approach. These included:

1) the fact that the validity scale evidence for the non-differential recall of unrelated (innocent) exposures does not completely rule out the potential for differential reporting of true risk factors; and, 2) different exposures may behave differently regarding recall bias (Werler et al., 1989; Coughlin, 1990).

Prior to the comparison of the overall VSSCOREs for the cases and the two control groups, a one-way analysis of variance was completed on each individual item VSSCORE to determine if there were any significant group differences. The presence of group differences might interact (cancel out) when aggregated into the VSSCORE. Consequently, a comparison of the summary VSSCORE would fail to detect differential reporting by the cases and controls.

These analyses were completed for a group of variables identified by breast cancer cases in another research project as being etiologically relevant for the occurrence of their breast cancer. Only those variables which were included in Q2

could be used for this repeat analysis. There was significant overlap between the variables included in the exposure data validity scale and those identified by case subjects. The results of all analyses are detailed in Chapter 4 of this dissertation.

#### **3.11.4 Evaluation of the 'Exposure Data Validity Scale'**

The evaluation of the exposure data validity scale is the final and most important stage in scale construction. The evaluation phase was designed to determine the relative effectiveness and efficiency of the proposed scale as a strategic tool for the measurement and the control of differential exposure misclassification (recall bias). Consideration of the strengths and limitations of the scale will determine its suitability as a design standard for routine inclusion in future case-control studies.

As an integral part of the evaluation process, the scale (VSSCORE) results were compared to the findings of the exposure data reliability study for the 'true' risk factors for breast cancer (i.e., the comparison of the prospective and retrospective exposure prevalences and risk estimates for the study variables). If the scale were successful in its stated goals, the VSSCORE analysis would estimate the differential reporting patterns, or the absence of group differences which were to be disclosed in the main study results.

The proposed validity scale was also assessed for the ease with which it can be designed, implemented and analyzed. If excessive amounts of research time and monies are required to develop the scale or to collect the data, or if too much complexity is required to analyze the results, researchers would not consider the application beneficial or feasible for development and inclusion in their case-control research.

To be accepted as a design standard for the measurement and control of differential recall accuracy, the validity scale must be shown to be comparable to the results of an exposure reliability study, and be both easier and less costly to execute. Raphael (1987) suggests that the use of a validity scale may be more cost-efficient -- [lower cost per unit of information generated] -- as well as providing valid study conclusions. It is almost certainly more expensive to complete an exposure reliability study than to administer a validity scale questionnaire.

### 3.12 Summary

In summary, this chapter has described the materials and methods which were used to answer the study questions outlined in Section 3.1. These included: the study design, the recruitment and selection of the study groups, the rationale for inclusion of multiple control groups (including an anamnestic equivalent control group), sample size and power considerations and the procedures used to collect and to analyze the research data.

This study was designed to address two important and often controversial issues which significantly influence the interpretation and acceptance of case-control studies for etiologic investigations: 1) the reliability (reproducibility) of historically recalled exposure information which is reported post-diagnosis, once the cases are aware of their disease; 2) the presence and the impact of non-differential and differential exposure misclassification on the estimates of association. Here the objective was to determine if valid (unbiased) exposure-disease odds ratios can be generated in case-control research. To answer these questions, exposure information was collected prospectively and retrospectively by means of a self-administered questionnaire.

More importantly, this study was designed to go beyond the boundaries of previously conducted data reliability and recall studies. Based on Raphael's proposal (1987), an exposure data validity scale was designed and evaluated as a possible design strategy for the measurement and statistical control of case-control differences in recall accuracy (i.e., a validity substudy for use in future case-control studies). Willett (1989) noted that "procedures to correct estimates of association in epidemiological studies for the effects of exposure measurement error have rarely been developed or employed in practice" (p.1031).

The motivation and justification for this research study is summarized by Feinstein (1979b) who stated that the scientific validity of epidemiological research "depends on the quality of the basic, raw data", and that the "*sine qua non* of scientific research is the establishment of methods to demonstrate the accuracy and reliability of the basic data" (p.487). Consequently, more ongoing studies are required to improve the acquisition of comparable and complete exposure data from cases and controls, as well as to provide the means to investigate different sources of bias which may be influencing a particular case-control study. Studies specifically designed to address the stated methodological limitations of case-control methodology are urgently required. Feinstein (1985b) noted that systematic empirical research was required to improve the 'scientific credibility and stature' of case-control studies. Furthermore, he emphasized that there was no room for 'scientific complacency' when "a better set of scientific standards were required to guide the planning of case-control studies" (pp.127, 133).

## Chapter 4

### RESULTS

#### 4.1 Response Rate

As detailed in Chapter Three (Sections 3.4 and 3.5), the study population was composed of three comparison groups. These included cases (women with a histologically confirmed diagnosis of breast cancer), and two control groups -- anamnestic equivalents (women who were similar to the breast cancer cases in that they had an abnormal mammogram, had feared a positive diagnosis of breast cancer, had required the same diagnostic procedures, but were then found to be free of disease); and, healthy controls (women with a normal mammogram).

All three groups were equally receptive and willing to participate in this research study. Of the 1,394 eligible subjects (Table 3, p.232), 1,177 (84.4%) agreed to complete the study questionnaire. This overall response rate is further broken down into individual rates of response by study group as follows:

- 1) cases - 234 (participants)/280 (eligible subjects) -- 83.6% (participation rate);
- 2) control group one (anamnestic equivalents) - 464/556 -- 83.5%; and,
- 3) control group two (normal mammogram) - 479/558 -- 85.8%.

As can be seen from these data, there were no discernible group differences in response rate, and the achieved overall rate of response was considered to be excellent. This high level of participation was also significant in view of the methodology that was used to collect the exposure information (i.e., a mailed, self-administered questionnaire).

One of the stated disadvantages of a mailed survey questionnaire is the potential for a low response rate, and the occurrence of selection (sample

distortion) bias. Babbie (1973) noted that "mailed studies sometimes receive response rates as low as 10 percent, and that a response rate of 50 percent is considered 'adequate'" (p.165). In addition, Bailey (1987) commented that a low response rate is problematic because non-respondents who do not answer are "not a random selection of the sample, but have some biasing characteristics" which could affect the study results, and possibly invalidate study conclusions (p.149).

Consequently, the similarly high rate of response among the comparison groups in this study would suggest that the internal validity of the study has not been compromised by a biased selection of study subjects. A consideration of study validity as it pertains to this research will be presented in Chapter 5, and will include a further discussion of selection bias in Section 5.3.1.

Table 4 (p.233) summarizes the reasons given by the non-responders (N=217) for non-participation. Non-responders were by definition those subjects who did not consent to participate in this study, those who initially agreed to participate in the study but who failed to return the questionnaire even after receiving a reminder letter, a second study package, and a telephone reminder, as well as those eligible subjects who could not be located.

The reasons for non-response included: 1) Language - the inability of the subject to comprehend and to respond to the study questionnaire independently - (5.5%); 2) Age - the reported inability to remember the exposure information that was being asked in the study questionnaire - (7.8%); 3) Poor health of either self or an immediate family member - (3.7%); 4) Unqualified refusal - (no reason given) - (13.8%); 5) Qualified refusal - a statement that the subject did not wish to participate - (21.7%). Within this last group of non-participants were women who questioned the significance of research, and those women who stated that the results of this study could not help them. As well, there were seven subjects who complained about one of the following: the lack of medical care after treatment;

no health teaching by, or discussions with, health care providers regarding prognosis (i.e., "what is in store for them (cases) in the future"); and/or dissatisfaction with the provincial screening program. Two women said that their lawyers recommended that they not participate due to pending litigation resulting from the diagnosis and/or treatment of their breast cancer; 6) The potential subject could not be located, or had moved out of the province - (27.6%); 7) Initial agreement to participate, but then the subject was lost-to-follow-up - (15.7%); 8) Away on vacation/out of the country - (1.4%); and, 9) Women had died - (2.8%).

From the eligible study subjects (N=1177), 6% (15) of the cases, 4% (19) and 9% (43) of control groups one and two respectively contacted the researcher regarding Questionnaire Two (Q2) and its completion. The nature of the telephone contacts and written correspondence included: suggestions to the researcher to add other exposures to Q2 that they personally considered important for the development of breast cancer, personal theories on breast cancer etiology, clarification regarding the question on breast/chest trauma (i.e., whether or not trauma would include surgical procedures), drug exposures not covered in Q2, how to calculate either the frequency or duration of exposure for the included study factors. Case subjects also requested information regarding the latest research findings on breast cancer treatment and prognosis, specific counselling resources available to them, and what they should expect with respect to medical follow-up after treatment. Women also called the researcher to request a copy of study results once they were available.

#### **4.2 Description of the Study Population**

The next section of the analysis will provide a description of the study population regarding the various study factors (i.e., demographic variables, medical and reproductive histories, lifestyle factors (i.e., smoking and alcohol

consumption), family history of breast cancer, and the use of exogenous hormones such as oral contraceptives and postmenopausal estrogen). Population descriptive data were reported in Table 5 (pp.234-235), Table 6 (pp.236-240) and Table 7 (pp.241-244).

In describing the study population, four separate levels of analysis were considered: 1) the entire study population (respondents) - Table 5 (pp.234-235) and Table 6 (pp.236-240); 2) the study population versus all the Screening Mammography program (first-time and returning) participants - Table 5 (pp.234-235); 3) the study group comparisons involving the cases and controls (groups one and two) - Table 6 (pp.236-240); and lastly, 4) the respondents versus non-respondents - Table 7 (pp.241-244). In these analyses, it was important to know if the comparison groups differed on any of the study factors in order to assess the overall validity of the study (i.e., the study's internal and external validity), as well as the 'representativeness' of the study population to the target population (i.e., all women participating in the SMP BC program). Representativeness would indicate that the study results could be generalized to other subgroups of the target population.

The categories of information reported upon in the second analysis (i.e., the study population versus all first-time and returning SMP BC participants), as well as the specific vital statistics for the SMP BC target population were taken from the 1990-1991 Annual Report of the Screening Mammography Program of British Columbia (pp. 23-26).

1. **Age Distribution.** Study participants were between the ages of 38 and 94 years of age, with the mean age of the respondents being 62 years of age. There were no group differences (cases versus controls) in the specific **age distributions** because **age** (+/- 5 years) was used as one of the matching factors for the selection

of the eligible controls for the case subjects. The specific age of the study population - Table 5 (p.234) was distributed as follows: 15% were age 40-49 years, 23% were 50-59 years, 34% were 60-69 years, and, 28% were 70 years and older. Thus, the majority of the study population (62%) were older, aged 60 years and older. As well members of the study cohort tended to be older than the target population (62% versus 32% were over 60 years of age). The study population was older than the population from which they were drawn because of the following factors: 1) the fact that breast cancer occurrence increases with a woman's age (i.e., breast cancer cases will tend to be older); and, 2) the study design employed to investigate the research hypotheses (i.e., a nested case-control study). In this study design, every case was included, and the controls were then randomly selected for each case subject based on four matching (inclusion) criteria, including age  $\pm$  5 years. This age difference can also explain the other differences noted between the participants and non-participants (e.g., postmenopausal status, occupational classification (retired/unemployed), and their level of education (i.e., fewer subjects had completed post-secondary studies)).

2. **Ethnicity.** The study population was found to be quite homogeneous with respect to ethnicity. Once again, there were no individual group differences regarding ethnicity because this demographic variable was also used as a matching factor in subject recruitment. Table 5 (p.234) reports that 96% of the subjects were Caucasian; the rest of the study population was composed of Asian (Japanese, Chinese, Filipino) - 3%; and, other ethnic backgrounds (African, East Indian, Native Indian, Persian) - 2%.

3. **Marital Status.** In comparing the marital status of the three study groups - Table 6 (p.234), approximately 95% were married (i.e., presently married, living

common-law, divorced, separated, or widowed) while 4% of the cases, 5% of control group one, and 5% of control group two were never married. These data indicate that group differences by marital status were not evident in the study population. The data in Table 5 (p.234) demonstrated that the study population and the SMP target population were comparable regarding the distribution of subjects by marital status.

4.     **Education.** Analysis of the population by **education** indicates that study participants were well-educated - Table 5 (p.234) and Table 6 (p.236): 51% of the subjects had completed a secondary school education, and 44% were high school graduates with post-secondary education/training. Only 6% had received just an elementary education. No differences in educational achievement by study group were noted; the educational profiles of cases and controls were similar.

5.     **Occupation** - Table 7 (p.241). As noted previously, 63% of the population were over 60 years of age. Therefore, it was not surprising to find that 39% of the women were classified as either retired or unemployed. Less than 1% of the study group were classified as students. Of the group of women who were employed (60%), 32% worked at home (housewives), 14% in the service industries including clerical, sales and manufacturing jobs, 1% in trades-related jobs, and 13% were classified as professionals. As seen in Table 6 (p.236), there was a higher percentage of retired and unemployed women in the case group (45%) than among the two control groups (33% and 40% respectively).

6.     **Location of Residence** (urban versus rural). Here, the majority of the study population resided in an urban setting (79%); there were no differences in the cases and controls according to the distribution by location of residence.

The preceding analyses would indicate that the study population was primarily composed of urban, Caucasian women, aged 60 years and over, who were married, fairly well educated, and were either retired or unemployed, or working within the home (housewives).

In summary, when the study population was compared to all screening mammography subjects - Table 5 (pp.234-235), a few group differences were noted. The respondents in this study tended to be older (over 60 years of age) and postmenopausal; fewer subjects had completed post-secondary studies. For the remaining demographic variables, there were no group differences.

7. **Ovulatory History** - Table 6 (p.237). Cases and controls did not differ ( $p>0.05$ ) regarding the mean ages at menarche (13 years) and menopause (54 years). At the time of the study, 82% of the study population were menopausal - Table 7 (p.242). Among the women who were menopausal, 59% reported a naturally occurring menopause, 28% reported having a hysterectomy, and 13% a hysterectomy with a bilateral oophorectomy. As reported in Table 6 (p.237), the cases experienced fewer hysterectomies (19%) than the controls (31% and 33%). However, among those who had a hysterectomy, cases reported a higher frequency of having a hysterectomy with bilateral oophorectomy (18%) than the controls (11% and 11%). From Table 5 (p.234), it is noted that the study population differed from all the SMP BC participants in terms of menopausal status. More study subjects were postmenopausal (82% vs 55%) than premenopausal (18% vs 35%) when compared to the target population.

8. **Reproductive History** - Table 6 (p.238) and Table 7 (p.243). 84% of the study participants reported ever being pregnant (parous), and 16% were nulliparous (never pregnant). The cases and controls did not differ ( $p>0.05$ ) with

respect to the mean number of pregnancies (2.2, 2.4, 2.3), and the mean age at first full-term birth (24.7 yrs, 24.3 yrs, 24.7 yrs). In addition, the study population was comparable to all the SMP BC target population with respect to parity, and age at first birth - Table 5 (p.235).

9. **Use of Hormones** - Table 6 (p.239) and Table 7 (p.243). 44% and 34% of the respondents reported having ever used oral contraceptives (OCs) and postmenopausal estrogen replacement therapy (ERT) respectively. Cases were more likely than the control groups one and two to report never having used OCs (61%, 51% and 56%). Among 'ever' users of OCs and ERT, the mean duration of use (years) for the cases and controls for OCs (5.2, 5.4, and 5.7 years) and ERT (8.2, 7.1 and 7.9 years) did not differ significantly ( $p > 0.05$ ).

10. **Family History of Breast Cancer** - Table 6 (p.239) and Table 7 (p.244). Among the study population (respondents) 17% of the women reported a first degree relative (i.e., mother and/or sister(s)) who had been diagnosed with breast cancer. Cases and control group two respondents had a higher prevalence of breast cancer in their families when compared to control group one women (20% and 19% vs 14%). Respondents reported more cases of breast cancer in sisters (10%) than mothers (8%) - Table 7 (p.244). Here, the rate of breast cancer in sisters was greatest for the cases and controls (group one) when compared to control group two (11%, 11% and 7%). History of maternal breast cancer in the three study groups was similar with a slightly higher frequency reported among the cases (9% vs 7% and 7%).

Among the reported cases of breast cancer in any first degree relative (17.3%), 1.7% of the cancers were classified as premenopausal, 7% were unilateral and less than 1% were bilateral breast cancers. There were no significant group

differences regarding the occurrence of premenopausal breast cancer, and bilateral breast cancer. However, group differences were noted regarding the frequency of unilateral breast cancer in family members. Here, the history of unilateral breast cancer in either mothers or sisters was more prevalent in the cases than in either of the control groups, whose reported rates of occurrence were comparable (i.e., 9% vs 6% and 6%).

When comparing respondents to all the SMP BC participants (i.e., the target population), family history of breast cancer in any first degree relatives and type of breast cancer (premenopausal vs postmenopausal; unilateral vs lateral) were combined to produce a family history breast cancer risk profile - Table 5 (p.235). The levels of risk were: 1) No risk (nil history of breast cancer in first degree relatives); 2) Yes, lower risk (family history present; breast cancers reported are unilateral and postmenopausal); 3) Yes, higher risk (family history present; breast cancers reported are bilateral or premenopausal). Given these classifications of risk pertaining to family history, - Table 5 (p.235) indicates that more of the study population than the target population were classified as higher risk because of a higher prevalence of first degree relatives with reported bilateral or premenopausal breast cancer (i.e., 3% vs 5%). However, these differences were not statistically significant ( $p > 0.05$ ).

11. **Smoking History.** Among the study population (Table 7, p.244), 58% were classified as current or former smokers (i.e., 'ever' smokers), and 42% reported that they had never smoked. Group differences were noted between the three study groups - Table 6 (p.240), and between the respondents and the SMP participants - Table 5 (p.235) regarding their smoking status (ever vs never). More cases and control group one subjects were 'ever' smokers (65% and 59% respectively vs 49% for the second control group). When compared to the target population (Table 7,

p.244), respondents also had a higher prevalence of 'ever' smokers (58% vs 48%). There were no group differences between respondents and non-respondents regarding the mean number of cigarettes smoked per day and the mean duration or total number of years that the particular group had smoked.

Among current and former smokers, the reported mean number of cigarettes smoked per day by the cases and control groups one and two was comparable (i.e., 15.6, 16.2 and 14.2 respectively). The cases and control group one subjects reported smoking the most number of cigarettes per day (16, 16 and 14 cigarettes per day respectively). There were no statistically important differences by group ( $p > 0.05$ ) - Table 6 (p.240). Similarly, these two groups differed significantly ( $p < 0.01$ ) from control group two with respect to the average number of years (duration) that a subject had reported smoking. On average, the cases, control groups one and two had reported that they had smoked 16.3 years, 15.9 years and 12.2 years.

12. **Alcohol Consumption** - Table 6 (p.240). More cases than controls indicated a history of 'ever' drinking (70% vs 63% and 66%). Control group one reported the highest level of nondrinkers (37%) among the three study comparison groups. Among 'ever' drinkers, the mean number of drinks consumed per month for the three study groups (13.5, 13.8 and 11.7 drinks) were comparable; group differences were not found ( $p > 0.05$ ). However, there were group differences regarding the duration (i.e., the mean number of years of alcohol consumption). When compared to control group one, both the cases and the control group two respondents reported the highest duration of alcohol consumption (i.e., 23.2 and 23.3 years vs 20.2 years).

From Table 7 (p.244), it was observed that the prevalence of drinking was higher in the respondents (i.e., 66% vs 56%). Although the mean number of

drinks consumed per month was comparable for the respondents and non-respondents (i.e., 13 vs 14), the respondents reported the highest duration of alcohol consumption (i.e., 22 vs 17.6 years).

**13. Health Practices and Diagnostic Procedures** - Table 6 (p.237) and Table 7 (p.242). Of the study population, 75% (respondents) reported the regular practice of breast self-examination (BSE). The lowest frequency of breast self-examination was reported by the cases (i.e., 73% vs 74% and 77% for the controls). The frequency of breast self-examination per year varied with the mean frequency of BSE being reported as 11 times per year for the cases, 13 times per year for control group one, and 9 times per year for control group two.

More controls than cases reported having had a previous mammogram (i.e., 48%, 55% vs 40%).

More cases and control group one subjects reported having had at least one breast needle aspiration (i.e., 9.2%, 11.4% vs 6/9%).

**14. Anthropometric Characteristics** - Table 6 (p.239). Comparison of the study groups with respect to Quetelet's Index (i.e., body mass to height -  $[\text{kg}/\text{m}^2]$ ) shows that the majority of the study subjects in all three groups have excess body mass (i.e., an index score  $\geq 25$ ). More cases than controls had an Quetelet's Index score  $\geq 25$  (i.e., 66% vs 53% and 53%). Likewise, the cases had fewer subjects with normal (22-24) and low ( $\leq 21$ ) body to mass index scores. In summary, all study subjects tended to be overweight (a high body mass ratio), with obesity being the greatest in cases.

The respondents and non-respondents were similar regarding body mass index as measured by Quetelet's Index Table 7 (p.243).

### 4.3 Analysis of Non-Respondents

As noted in Section 3.10.2, an analysis of respondents and non-respondents was required to determine the comparability of the two groups with respect to the various study factors, as well as to ascertain any group variances that might be associated with group differences in recall accuracy, and the subsequent occurrence of either differential (recall bias) and/or non-differential exposure misclassification. Such differences, if found to exist, could possibly invalidate study conclusions.

The analysis of Table 7 (pp.241-244) indicated that the non-respondents and respondents were comparable regarding most of the study factors, including the risk factors associated with breast cancer development - with the exception of a family history of breast cancer and the use of postmenopausal estrogen.

Respondent and non-respondent group differences were noted in the following areas: education, location of residence, whether or not the subject practises regular breast self-examination, a reported history of previous mammograms, the existence of somatic changes possibly related to ovulation, the use of postmenopausal estrogen, history of breast cancer in any first degree relative (i.e., mother or sister(s)), smoking status (ever smoker vs never), and alcoholic consumption (ever drinker vs never).

The relationship between education and response rates has been well documented by researchers in the social sciences (Groves, 1985, p.205). In general, it was noted that the more educated subjects will be more likely to respond to survey requests, and to return completed questionnaires (Dillman, 1978; O'Neil, 1979). These observations were possibly relevant in this study.

Here, it was observed that a higher percentage of the non-respondents had only received an elementary school education (14% vs 6%), and fewer of them had gone on to complete post-secondary training (33% vs 44%). Thus, respondents

tended to be better educated, and perhaps were more motivated to participate in this study. Their education also better prepared them to comprehend what was being required of them, and to complete the self-administered exposure history questionnaire with minimal difficulty.

The accuracy of recall and subsequently the reliability and validity of retrospective reports of exposure, may be due in part possibly to the educational and related motivational differences of the study population -- respondents versus non-respondents. However, the rate of non-response is not frequent enough (16%) to make self-selection on the basis of education a likely occurrence with concomitant adverse affects on study conclusions.

In analyzing health promotion behaviour, such as the practice of regular breast self-examination (BSE), and prior mammography, it was observed that more respondents than non-respondents reported practising BSE (75% vs 67%). Likewise a higher percentage of respondents had a previous mammogram (48% vs 38%). These factors may be considered as a proxy measurement for a subject's level of motivation to participate in a research study, her willingness to invest the time and the effort to remember and to report past exposure information, and the signs of an informed health care consumer who has the knowledge and awareness of breast cancer risk factors, as well as the understanding of prevention and screening. The reliability and validity of the reported information may be related to the increased motivation of the subjects to participate and to provide complete and accurate medical and exposure histories.

Non-respondents also differed from respondents regarding the recall and reporting of somatic changes possibly related to ovulation, as well as smoking history and the consumption of alcohol. Here, a higher percentage of respondents reported a positive history of breast tenderness (57% vs 46%); however, among those subjects who reported a positive history of breast tenderness, the non-

respondents reported a higher frequency of period-related breast pain and/or tenderness (59% vs 45%).

In addition, a higher proportion of respondents were classified as drinkers (66% vs 56%) and current or former smokers (58% vs 48%). Respondents and non-respondents were similar regarding the mean number of drinks consumed per month (13 vs 14); however, there were group differences in the mean duration (years) of alcohol consumption. The highest duration of alcohol consumption was observed in respondents (i.e., 22 years vs 18 years). These reported group differences should have no bearing on the study conclusions because somatic changes related to menses, smoking history and alcohol consumption have not been associated with breast cancer risk; nor do they influence participation rates, or an individual's motivation to recall exposure information both accurately and completely. Furthermore, no increased risk was found to be associated with alcohol consumption when exposure-disease odds ratios were estimated for this study variable - Table 11 (pp.266-273). This last observation does not hold for smoking history because significant odds ratios ( $p < 0.05$ ) were estimated for the case-control group two comparisons.

Group differences between respondents and non-respondents were also noted for two variables which have been investigated as risk factors for breast cancer development. These included: a positive history of breast cancer in first degree relatives and the use of estrogen replacement therapy (ERT) postmenopausally. Respondents reported higher percentages of first degree relatives with a positive diagnosis of breast cancer (17% vs 11%), and more frequent use of postmenopausal ERT (34% vs 28%) when compared to the non-respondents. In view of the low overall non-response rate (16%), and the results of the analysis of breast cancer risk that is associated with these variables (i.e., the

exposure-disease odds ratio estimates), the differences between the respondents and non-respondents should not adversely influence the study results.

In summary, no significant differences were noted between the respondents and non-respondents that would adversely impact on the analyses, thereby invalidating any generated study conclusions.

#### **4.4 Test-Retest Reliability: Agreement Between the Prospective and Retrospective Exposure Reports**

In this part of the analysis, the 'test-retest' reliability of Q1 and Q2 (prospective versus retrospective) reports of exposure was assessed by means of a correlation analysis. Pearson and Spearman Rank correlation coefficients were estimated independently for each study factor for the cases and the two control groups separately. The correlation coefficients were then compared to determine the level of agreement (reliability) between the prospective and retrospective exposure reports, and the occurrence of any group differences in recall accuracy. The results of the correlation analysis (test-retest reliability) are found in Table 8 (pp.245-249).

The correlations estimated for the study variables suggested a high degree of correlation (reliability) between the prospective and retrospective reports of exposure. Knowledge about diagnosis (i.e., group membership) did not seem to influence the consistency with which the retrospective, post-diagnostic exposure assessments had been reported by the study participants. For the majority of the variables examined, the magnitude of the Pearson and Spearman Rank correlation coefficients suggested moderate to high levels of agreement (or correlation) between Q1 and Q2 reports of exposure (i.e., the correlation coefficients from 0.50 to 1.00).

In general, the correlation coefficients ranged in value from (.74 to .94) for the demographic variables which included marital status, ethnicity, and education; (.83 to .92) for subject information on menarche and menopause (i.e., menopausal status, age when periods started and stopped, as well as type of menopause); (.68 to .97) for reproductive history variables including parity and age at first full-term birth; (.71 to .92) for information on the use of exogenous hormones (i.e., oral contraceptives and postmenopausal estrogen); (.84 to .97) for reports of a positive history of breast cancer in first degree relatives and the type of familial breast cancer (i.e., premenopausal or postmenopausal, unilateral or bilateral); (.63 to .97) for lifestyle factors such as smoking, the consumption of alcohol and level of dietary fat; and lastly, (.29 to .65) for reports about breast swelling, pain and tenderness possibly related to ovulation.

Low correlations (.00 to .49) were observed for the subjects' reports regarding the occurrence of somatic (bodily) changes preceding menses, the identification of the symptoms associated with breast pain and tenderness (excluding periods during pregnancy), and among those who reported having had a previous mammogram, the date of the last mammogram. Here, the nature, perceived salience, as well as the detail of the information requested may be responsible for the factors' susceptibility to measurement error, and the resulting low correlations.

The estimated correlation coefficients were similar for all three study groups, and were statistically significant ( $p < .001$ ). There was no evidence of differential exposure misclassification bias (i.e., the observation of systematic group differences in the level of consistency (reliability) between the prospective and retrospective reports of exposure as measured by the correlation coefficients).

Table 8 (pp.245-249) identified group differences for only three factors:

1) the symptoms associated with a history of breast pain and tenderness (i.e., pain, tenderness, or both pain and tenderness); 2) pregnancy history (i.e., ever versus never pregnant); and, 3) the frequency of having a breast needle aspiration and the specific age (years) when the first aspiration procedure was completed. For the first variable, the symptoms of breast pain and tenderness, the reliability of the case reports was better than those provided by the two control groups (.55 vs .43 and .47). Conversely, there was better overall agreement (correlation) for the reports of the two control groups regarding breast needle aspirations (i.e., ever versus never) - (.96, .97 vs .51), and the subject's age (years) when the first procedure (aspiration) was performed. Group differences for these two variables were not considered to be important because they are not classified as potential risk factors for breast cancer. The final study factor -- reproductive history (ever pregnant versus never) - was most reliably reported by the cases and control group one subjects (i.e., .83, .86 vs .68). No clear explanation can be cited to explain the group differences noted for this variable.

The results of the correlation analysis would suggest that the Q2 retrospective (post-diagnostic) exposure reports were reliable (consistent) when compared to the respective Q1 prospective (pre-diagnostic) assessments. However, as Raphael observes (personal comments on this doctoral thesis, December 1995), "the Q1 responses may have served as a rehearsal for the Q2 responses. That is, what was recalled at Q2 was probably not lifetime history of exposure (although this is what the researcher wants to assess in a case-control study); but instead, whatever was reported just several (approximately 6 months) earlier at Q1. Thus, the design inherently overestimates the overall accuracy of recall considerably. The Q1 gold standard is not 'tarnished' simply by occasional coding errors; the standard is more seriously flawed. For example, report of date of onset of menarche is likely to be the same 50 years later at Q1 and 50.5 years later at Q2.

The six months of elapsed time between Q1 and Q2 really does not much matter, given that the respondents are asked to recall exposures happening decades earlier. In this situation reliability does not imply validity". In addition, the existence of moderate correlation coefficients (0.50 - 0.69) for several study factors was evidence that there were inconsistencies in exposure reporting. However, these inconsistencies were similar for the three study groups -- systematic group differences were not observed. In general, there were no discernible group differences in recall accuracy (i.e., level of agreement between the prospective and retrospective reports of exposure).

#### **4.5 Kappa Analysis: Agreement Between Prospective and Retrospective Exposure Assessments**

To avoid overestimating the level of agreement, concordance or reproducibility between the two exposure assessments (i.e., Q1 - (pre-diagnostic) and Q2 - (post-diagnostic)), Cohen's Kappa coefficient ( $k$ ) was calculated for each study factor. The distinct advantage of using the Kappa statistic was its ability to control for the proportion of cases in which agreement was due to chance. In other words, the Kappa coefficient was able to estimate the level or degree of agreement between Q1 and Q2 reports of exposure that went beyond that expected on the basis of chance alone. The results of the Kappa analysis for the various study factors are reported in Table 9 (pp.250-253).

The estimated kappa statistics for all study variables were grouped by level of agreement (reproducibility) according to the classification system suggested by Landis and Koch (1977, p.165) and then modified by Fleiss (1981, p.218).

The data suggested that the post-diagnostic reports of exposure are reasonably consistent with the exposure information provided by respondents

before they were aware of their case-control (group) status. In other words, the degree of agreement (reliability) between Q1 and Q2 reports of exposure was better than would be expected on the basis of chance. The knowledge of group status did not appear to have had any significant influence on the level of recall accuracy. The specific information provided on Q1 had been provided by the same respondents on Q2: the majority of the study risk factors had a Kappa value greater than or equal to .40, which indicated moderate to substantial reproducibility: several variables had kappa values  $> 0.70$  indicating substantial reproducibility.

Table 9 (pp.250-253) data also suggested that the level of reliability or reproducibility of exposure information (i.e., the respondent's reports of exposure) was dependent on the particular variable being reported upon. Q1 and Q2 reports of exposure for several study variables had low Kappa values (0.0 to 0.4), and were assessed to be unreliable, and inconsistently reported by study subjects regardless of group membership. These factors included: mammography history (i.e., the total number of previous mammograms and the date of the last mammogram), reported history of ever having a breast needle aspiration, the total number of aspirations, and the subject's age at the time of the first aspiration. Memory variables such as salience, detail of information requested, respondent motivation to participate and to recall the requested material, age, the absence of an available heuristic, time, judgment factors, etc., may be relevant in explaining the lack of agreement between the two exposure reports.

Group differences in the level of consistency for reported exposure information (Q1 vs Q2) were also noted for three study factors: 1) among the postmenopausal women, the age (years) when their menstrual periods stopped; 2) the amount of alcohol consumed on average per month; and, 3) the duration of use of postmenopausal estrogen supplements. Once again, there was no evidence

for systematic group differences in the reporting of exposure information (i.e., reliability/consistency); however, measurement error in Q2 exposure reports was observed.

In summary, the analysis of the overall agreement of Q1 and Q2 reports of exposure indicated moderate to high reproducibility; the level of consistency between the two exposure reports was better than would be expected on the basis of chance alone. As noted previously, the Kappa coefficients have probably been overestimated due to the correlation existing between Q1 and Q2 reports of exposure. Consequently, the level of agreement would not be as large as suggested by mere observation of the magnitudes of the Kappa coefficients. The nature of any differences (i.e., inconsistency which was reflected in Q2 changes of exposure classification) was then explored by means of the McNemar Test.

#### **4.6 McNemar Analysis**

Tables 10.1 - 10.3 (pp.254-265) report the results of the McNemar analyses for the three study comparison groups (i.e., the cases, and control groups one and two). As discussed in Section 3.10.5, McNemar's test for correlated proportions is a 'test of symmetry'. This analysis will specifically evaluate the number of discordant pairs which occur when Q1 and Q2 exposure assessments are compared for the various study factors. In this study, McNemar's testing complemented and clarified the results of the Kappa analysis, which was designed only to measure crude overall agreement between the prospective and retrospective reports of exposure. The primary objective of this analysis was to determine if the changes in the reports of exposure in one direction -- an overreporting (overstatement) or underreporting (understatement) of exposure level (Marshall, 1981) were similar to the frequency of the changes in the other

direction. Other objectives included the comparison of the study groups to determine if they differed significantly with respect to the frequency of discordant pairs, and subsequently, whether or not there was a tendency by any of the groups to systematically alter their reports of exposure in a particular direction, thereby biasing the risk estimate and invalidating the study conclusions. Here, one is attempting to assess the presence of differential exposure misclassification, which is characterized by a systematic shift in self-reported exposure status (i.e., patterned misclassification) which is associated with the diagnosis of breast cancer.

The factors which are reported on in Tables 10.1 - 10.3 (pp.254-265) had at least one group McNemar test which was statistically significant (i.e.,  $p < 0.05$ ). As part of this analysis, the prevalence of both concordant and discordant pairs was determined, and then used to estimate the percent discordance and the percent concordance by study group for that variable.

The comparison of the resulting discordance rate to the overall rate of concordance for that factor permitted an assessment of the relative importance of the statistically significant McNemar test -- whether or not exposure misclassification was present, and its impact on the estimates of effect. Armitage and Berry (1971, pp.122-123) noted that the McNemar's (significance) test is based entirely on the frequencies of discordant exposure reports (Q1 vs Q2). That is, the evidence for the existence of a difference or change in the reporting of individual exposure is provided entirely by the discordant pairs. These authors caution that "an assessment of the **magnitude** of that difference must refer to the remainder of the data. The distinction between statistical significance and clinical significance must also be considered in the overall analysis" (p.123). Consequently, this analysis required that the discordance rates be compared against the concordance rates to determine practical significance in view of the stated research questions.

The percent discordance for the 17 included factors ranged from (4-18%) for the cases, (3-20%) for control group one, and (2-17%) for control group two. Of the 17 study factors reported on in Tables 10.1-10.3 (pp.254-265), 65% of them had discordance rates less than 10% and concordance rates greater than 90%. The problem of discordance resulting from overstatement or understatement of exposure on Q2 was not very frequent when compared to the overall concordance -- a measure of the agreement between pre-diagnostic (Q1) and post-diagnostic (Q2) reports of exposure. Therefore, the prevalence of discordance for the 17 reported risk factors (Tables 10.1 - 10.3, pp.254-265) can be considered 'clinically' unimportant, even though statistically significant. It can be concluded that Q2 - retrospective exposure assessments provided reliable exposure information when compared against the 'criterion' measure (Q1).

As can be seen from the results reported in Tables 10.1 - 10.3 (pp.254-265), if the risk factor was dichotomous, wherein a subject's exposure was classified as either 'Yes vs No', or 'Ever vs Never', the prevalent direction of change (for all study groups) was from 'not exposed to exposed', and from 'never to ever'. Therefore, the resulting discordance would be classified as an overstatement of exposure, when Q1 (pre-diagnostic) exposure reports were used as the 'gold standard' or criterion against which the validity of the retrospectively collected exposure data (Q2) were compared. When the risk factor was categorical and ordered, and consisted of three or more levels of exposure, it was noted that the exposure misclassification occurred between adjacent categories. The data reported in Tables 10.1 - 10.3 (pp.254-265) indicated that there was no tendency for any of the comparison groups to overstate or understate exposure systematically. The direction of the discordance cannot be predicted, and varied between risk factors, and study groups. In general, all three groups tended to overstate their exposure level. As well, the control groups had more significant discordant

reports as measured by the McNemar test, than did the cases. However, in view of the low discordance rate, these observations were not considered to indicate a trend, and should therefore be discounted.

In discussions about recall bias (i.e., differential exposure misclassification), the authors often assume that due to the salience of the outcome event, the cases rather than controls would be more motivated to remember and report past exposures in their attempt to understand the possible etiology of their disease. As a result of the 'search for cause' cognitive processes, it has been postulated that the cases may "recall more information about exposures than equally exposed but healthy referent subjects" (Mackenzie, 1986, p.11). The fact that our cases have fewer discordant reports overall does not support the view that cases have a tendency to overstate their past exposure in their attempts to understand their illness, and to establish a link between cause and effect. The results of this study would suggest that, for both the cases and controls, there was adequate reliability of the reports of exposure, accompanied mostly by random exposure (misclassification) measurement error.

As noted previously, the overall discordance rate was minimal for all groups. Thus, the overall high concordance of the exposure data implies that the retrospectively collected exposure data were consistent (reliable) and were not distorted by knowledge of disease outcome - the diagnosis of breast cancer. Consequently, the exposure-disease odds ratios were not biased by the resulting overstatement or understatement of exposure that is present in Q2 responses.

In summary, when the Q1 and Q2 reports of exposure were compared, differences in the prevalence of exposure categories were noted. However, the proportion of discordant to concordant reports of exposure was not sufficiently frequent to bias the odds ratio estimates either toward or away from the null value. The data do not provide evidence for differential reporting of exposure;

however, inconsistency (exposure misclassification) existed. That is, the data suggested that random measurement error occurred in the retrospective reporting of selected exposure information. The resulting moderate to high Kappa scores did not suggest that there was no distortion in the measures of association, nor did they imply an absence of bias.

#### **4.7 Retrospective Versus Prospective Reports of Exposure: A Comparison of the Exposure-Disease Odds Ratio Estimates**

The next phase of the data analysis consisted of the calculation of odds ratios for each study factor estimated separately for the prospective (Q1) and the retrospective (Q2) reports of exposure. Here, we were interested in assessing the impact of changes in the exposure reports (i.e., any overstatement/overreporting or understatement/underreporting of exposure) on the measures of association between the risk factors and disease (i.e., the occurrence of breast cancer). Any resulting changes in the risk estimates could then be classified and attributed either to non-differential exposure misclassification (i.e., exposure misclassification errors that are independent of the disease outcome), or to differential exposure misclassification (i.e., exposure classification errors that are systematic and dependent on disease outcome -- group status in a case-control study). The prospective and retrospective exposure-disease odds ratios were calculated for the two case-control comparison groups (i.e., cases vs control group one and cases vs control group two). In comparing Q1 and Q2 odds ratios, one should account for the correlation in the data which occurs because the Q1 and Q2 exposure reports were obtained from the same subjects. However, in this study, the odds ratio analyses were based on the assumption of the independence between Q1 and Q2 response sets for the various study factors. This will result in an extremely

conservative test of the comparability of the prospective and retrospective odds ratio estimates.

Exposure misclassification and its impact on the estimates of risk have been investigated and reported upon by several researchers and methodologists including Bross (1954), Gullen et al. (1968), Copeland et al. (1977), and Goldberg (1975). Rothman (1986) noted that "differential misclassification (DEM) can result in an information bias that exaggerates or underestimates an effect" (p.85). The pattern of exposure misclassification for the cases and the controls would be systematically different; that is, there would be a systematic shift in the reporting of past exposures because the subjects are aware of their group status (disease outcome). This would imply that the Q2 (retrospective) odds ratio when compared to the Q1 estimate of effect would be either: 1)  $> 1.0$  (indicating a change from no association to a positive association between the risk factor and breast cancer, or 2)  $< 1.0$  representing a change in risk from no association to a negative (protective) association. In addition, an assessment of differential exposure misclassification (DEM) would require that, overall, the retrospective (Q2) estimates for the association between the various study factors and breast cancer would have to be systematically larger or smaller than the prospective (Q1) estimates for a particular study group.

Conversely, if non-differential exposure misclassification (NDEM) were present, the risk estimate would always be biased in a predictable way toward the null condition, that is, the risk estimate would equal 1.0 (Bross, 1954; Copeland et al., 1977; Rothman, 1986). Rothman (1986) noted that with NDEM, the resulting exposure misclassification is "incorrect for equal proportions of subjects in the groups compared" (p.86). Here, depending on the prevalence of the misclassification and the nature of the exposure misclassification errors, a positive or negative association may not be detected, and the investigator may incorrectly

conclude that there is no association between the risk factor and disease being assessed (i.e.,  $OR=1.0$ ). Rothman (1986) also stated that NDEM "has generally been considered a lesser threat to validity" because of the predictability of its biasing effect on measures of association (p.86).

In this study, if the Q2 risk estimates change in ways that are suggestive of differential and/or non-differential exposure misclassification, the resulting exposure misclassification can be attributed to case-control differences in recall accuracy because the prospective and retrospective exposure data had been collected in an equivalent manner (Mackenzie, 1986; Friedenreich, 1990).

The estimated odds ratio and their corresponding 95% confidence intervals for Q1 and Q2 reports of exposure for the two case-control comparison groups are reported in Table 11 (pp.266-273). The exposure categories and levels of exposure were adapted from Friedenreich (1990, pp.154-160). Table 12 (pp.274-277) details the study factors for which the prospective and retrospective odds ratios differed significantly, as well as the direction of change. This table was used to determine the prevalence of DEM and NDEM, as well as the comparability between the two case-control comparison groups regarding recall accuracy and the occurrence of exposure misclassification.

Analysis of the data in Table 11 (pp.266-273) revealed only 13 study factors which had a significant exposure-disease odds ratio based on either the prospective (Q1) or retrospective (Q2) exposure assessments. Two of these factors were disregarded from this part of the data analysis (i.e., breast swelling and breast pain and tenderness) because they had not been considered in the epidemiological literature reviewed as being risk factors for breast cancer development.

Of the remaining 11 study factors with a significant odds ratio, only four variables - pregnancy history, smoking history, and the use of oral contraceptives

or postmenopausal estrogens demonstrated changes in the odds ratio (OR) estimate that were suggestive of the existence of either differential (DEM) or non-differential exposure misclassification (NDEM). However, the retrospective (Q2) estimates for the association between the various study factors and breast cancer were not systematically larger or smaller than the prospective (Q1) estimates. That is, there was no systematic shift in the reporting of exposures once case-control status was determined. Therefore, strong evidence for DEM was absent.

For case-control group one comparisons, the risks associated with oral contraceptive (OC) use and smoking history changed from no association to a negative association (i.e., never used OCs vs ever used: 1-5 years) and from no association to a positive association (i.e., never smoked vs smoked: 11-19 years), respectively. These changes were observed to be in the direction expected in the presence of differential exposure misclassification (DEM). For another category of risk associated with oral contraceptive use (i.e., never used OCs vs ever used: >5 years) the change in the direction of risk was from a negative association to no association. This change in risk was in the direction suggestive of non-differential exposure misclassification (NDEM).

For case-control group two comparisons, the risk associated with postmenopausal estrogen use (i.e., never used vs ever used: 6-10 years) changed from no association to a negative association (suggestive of DEM). The changes in risk associated with pregnancy history (i.e., ever vs nulliparous) and smoking history (i.e., never smoked vs smoked: 20-29 years) went from a negative and positive association respectively to no association, and were suggestive of NDEM.

Overall, there was no systematic overreporting/overstating or underreporting/understating in the subjects' reports of exposure: there was no discernible pattern or propensity for changes in risk estimates in one particular

direction for several factors when Q1 and Q2 exposure-disease odds ratio (OR) estimates were compared.

Case-control group one and case-control group two OR comparisons were similar, both in the number of factors with significant OR changes, as well as the prevalence of the types of misclassification observed (DEM vs NDEM). Overall, neither the frequency of occurrence, nor the specific changes in direction in the risk estimates provided strong evidence for the presence of significant differential and non-differential exposure misclassification in this case-control study of breast cancer.

In addition, the use of anamnestic controls (control group one) did not appear to have any advantage over the use of normal, healthy controls (group two) with respect either to the overall consistency of reporting of exposures retrospectively, or subsequently to the minimizing or prevention of exposure misclassification.

Generally, the data suggested that: 1) retrospective OR estimates have adequate reliability and validity, subject only to measurement error; 2) the type of control group selected had minimal effect on either minimizing or preventing exposure misclassification; and, 3) differential and non-differential exposure misclassification were not significant problems in this study. Neither the frequency nor the types of changes in the OR estimates provided strong and conclusive evidence of differential exposure misclassification. There were no systematic group differences in the occurrence of reporting inconsistencies. The random errors depended on the factors being reported on; the factors subject to error could not be predicted in any way.

In all of the analyses completed to this point, including test-retest reliability (correlation analysis), Kappa, McNemar, and OR comparisons (Q2 vs Q1), there was little evidence of DEM or NDEM. Retrospectively collected exposure data

appear to be relatively reliable, subject only to measurement error. However, these apparently favourable results must be tempered by a concern that the agreement observed in sequential administrations of the exposure questionnaire reflects a possible shared error, that is, either an overreporting or underreporting of past exposure by the cases and the controls. In this nested case-control study of breast cancer, the problems of DEM or NDEM, and the invalidation of study conclusions did not occur.

#### **4.8 The 'Exposure Data Validity Scale': Development and Analysis**

The specific steps in the development of an exposure data validity scale were outlined in Section 3.11. These included: 1) the search for the candidate variables; 2) the selection of exposure variables for inclusion in the validity scale and the assignment of their specific weighting factors; 3) the administration of the validity scale to the target population and its analysis; and, 4) the evaluation of the effectiveness and efficiency of the validity scale as a design strategy to measure and to control exposure misclassification in a case-control study.

In order to select the 'plausible but fake' exposure variables and their specific factor weightings (i.e., a measure of the relative importance of the factor for breast cancer risk), the questionnaire found in Appendix 2 was administered to 147 screening mammography subjects. The data matrix, in Table 15 (pp.287-288), was then analyzed using the correspondence analysis program (CA) in the BMDP statistical package (BMDP (Dixon), 1992). Several sources (Lebart et al., 1984; Greenacre, 1984; Hoffman and Franke, 1986; Greenacre, 1993; Greenacre, 1994) were used extensively as primary sources both in the interpretation of the results of the correspondence analysis in this study, and in the discussion which follows.

The responses of 147 study participants were cross-classified in a two-way contingency table according to 5 levels of risk and 33 risk factors: Table 15 (pp.287-

288). The potential risk factors were identified from the medical literature, experts, and the popular media; they were then classified as 'related', 'unrelated' or 'controversial' for breast cancer development. Table 16 (pp.289-290) outlines the risk factors that were evaluated for possible inclusion in the validity scale, together with their abbreviation codes which appear in the graphs generated by CA.

The levels of risk assigned to the risk factors by the respondents corresponded to the women's perceptions of the etiologic importance (plausibility) of each factor in disease development (breast cancer). The exposure data validity scale questionnaire (Appendix 2) consisted of a 4-point scale, graded in etiologic importance as HIGH RISK (very important), MODERATE RISK (moderately important), LOW RISK (somewhat important), and NO RISK (not important) for breast cancer development. However, during the administration of the study questionnaire, participants indicated that for several of the risk factors, and in particular, the various drugs, they had no knowledge (familiarity), or opinion as to whether or not these factors played a role in the occurrence of breast cancer. Therefore, when screenees insisted that they could not respond using one of the existing scale levels, a decision was made to open another category (NO OPINION), resulting in a 5-point scale for the level of risk.

Only those study participants who spontaneously questioned the adequacy of a 4-point scale (N=47) were given the option of using the additional risk level category. The danger with the use of the 'forced choice' procedure was that some respondents may have felt pressured to give a specific answer (high, moderate, low or no risk), even though 'no opinion/did not know' was their proper response. This was considered a limitation in the development of the validity scale.

In any cell of the 33x5 contingency table, one finds the frequency with which participants ascribe a particular level of risk (etiologic importance) to a

specific risk factor. The abbreviations for the 33 risk factors are defined in Table 16 (pp.289-290).

The University of British Columbia computing services were used to analyze the contingency table data. As mentioned previously, these data were analyzed by means of correspondence analysis (CA).

CA is a descriptive multivariate statistical technique which transforms contingency table information (categorical variables) into joint graphical displays (in the same low-dimensional space), in order to detect and to interpret the relationships among the variables in the rectangular data matrix (Hoffman and Franke, 1986; Greenacre, 1994). Hoffman and Franke (1986) emphasized that correspondence analysis "shows *how* variables are related, not just that a relationship exists" (p.213).

CA is classified as a "geometric approach to multivariate descriptive (exploratory) data analysis" (Hoffman and Franke, 1986, p.214). The procedure was developed and popularized by the French analyst, Jean-Paul Benzécri in the early 1960s (i.e., "analyse factorielle des correspondances"). This technique has been used in other domains "under a variety of names, including: dual scaling, method of reciprocal averages, optimal scaling, canonical analysis of contingency tables, categorical discriminant analysis, homogeneity analysis, quantification of qualitative data, and simultaneous linear regression" (Hoffman and Franke, 1986, p.214).

In correspondence analysis, "numerical scores are assigned to the rows and columns of a data matrix so as to maximize their interrelationship. The scores are in corresponding units, allowing all the variables to be plotted in the same space for ease of interpretation" (Hoffman and Franke, 1986, p.215). These same authors stress that the goal of CA is "to obtain a graphical representation of both the rows and the columns of the original data matrix in terms of as few dimensions as

possible" (p.215). Furthermore, the resulting graphical display of each set of points (i.e., row and column profiles) "reveals the nature of similarities and variation within the set, and the joint display shows the correspondence between the sets" (p.219).

Prior to executing the correspondence analysis, the response frequencies for levels of risk for each risk factor were expressed as percentages of the marginal row totals. These data are outlined in Table 17 (pp.291-292). Here, the objective was to generate a crude estimate of which of the 33 risk factors could be defined as 'plausible' for the outcome event (breast cancer), based on the responses of the respondents. The decision rule for the inclusion of a factor as 'plausible' for disease development was that at least 70% of the study subjects evaluated the factor as being either high, moderate or low risk for breast cancer development. Based on this (subjective) inclusion rule, 19 of the 33 exposure (risk) factors qualified as potential items for inclusion in the 'exposure data validity scale'. The results of this crude and subjective analysis were compared with those derived from correspondence analysis to determine the accuracy and efficiency of the simpler, inclusion-rule based analysis compared to the formal (statistical) approach of correspondence analysis. This comparison can be found in Table 18 (p.293).

### **Correspondence Analysis of the Risk Factor by Level of Risk Matrix**

By means of CA, the categorical risk factors by attributes (level of risk) matrix will provide information on the positioning of each risk factor with respect to the level of etiologic importance selected to describe them. In terms of the 33 risk factors, in 5-dimensional attribute space (etiologic importance: high, moderate, low, no risk, no opinion), the levels of risk will be assigned differently

to each risk factor, and consequently, some risk factors will be classified more frequently as high risk rather than low risk.

"To perform a CA, one rescales the original data matrix so that the sum of the elements equals 1. These row and column sums are referred to as masses in CA. These masses enable the investigator to weight each row and column profile point in proportion to its frequency" (Hoffman and Franke, 1986, p.216). To graphically represent the distances between row (or column) profiles in k-dimensional subspaces, the configuration of points are positioned at the "center of gravity" of both sets (p.216). "The centroid of the set of column points in its space is equivalent to the vector of row masses" (p.216). To perform the analysis relative to the center of gravity, the correspondence matrix whose elements are the relative frequencies, is "centered 'symmetrically' by rows and columns...so that the origin corresponds to the average profile of both sets of points" (p.216).

The distances between row and column profiles are similar to Euclidean distances except that they are "weighted by the inverse of the relative frequency (mass)" which corresponds to the specific profile. Furthermore, these distances are defined as chi square distances because "it guarantees invariance according to the property of *distributional equivalence*:

- If two rows having identical column profiles are aggregated, the distances between columns remain unchanged.
- If two columns having identical row profiles are aggregated, the distances between rows remain unchanged.

Clearly, identical profiles imply equal or proportional raw data " (Hoffman and Franke, 1986, p.218). These authors also noted that "it is because of the geometric correspondence of the two sets of points, in position and inertia" that the separate displays of row and column profiles can be merged to represent the "varied features of the data" in a joint graphical display (pp.217-219). Furthermore, "the geometric display of each set of points reveals the nature of similarities and

variation within the set, and the joint display shows the correspondence between the sets" (p.219).

The numerical output from the correspondence analysis of the data matrix in Table 15 (pp.287-288) is reported in Table 20 (pp.295-296). This output reported the decomposition of inertia (i.e., spatial variation) among the rows (risk factors) and columns (perceived level of risk) along the first two principal axes (i.e., axis 1 and 2). The **total inertia** is defined as "the weighted sum of squared distances from the points to their respective centroids and is equivalent for both sets of points" (Hoffman and Franke, 1986, p.218). The spatial variation in the row and column profiles (as quantified by the total inertia term) aids in the interpretation of the joint display. The total inertia is also decomposed along the principal axes (eigenvalues) to "indicate the weighted variance (inertia) explained by the  $t^{\text{th}}$  principal axis of the display" (p.219).

**Interpretation of the Axes.** The eigenvalues and the percentages of variance (inertia) for 33x5 data matrix in this study are displayed in Table 19 (p.294). The first eigenvalue, also referred to as the first principal inertia (or principal axis), explained 72.5% of the total variance (i.e., spatial variation in the data); it was therefore considered to be highly significant (Lebart et al., 1984). Together, the first two axes explained 90.2% of the total inertia. These results suggested that two axes predominated, and the two-dimensional graphical display of the data points in Figure 2.1 (p.303) was the best simultaneous representation of the information contained in the contingency table (i.e., the rectangular data matrix of risk factor by attribute (etiologic importance)).

By scanning the quality (QLT), and the relative contribution of the principal axis to the respective row and column profiles (COR), the data indicated that the row and column points were well represented in two dimensions. The

only exceptions were the following row profiles: mastitis (which is best represented on axes 2 and 3), exposure to the chemicals in cleaning solutions (axis 3); and, the one column profile - low risk (axis 3). Therefore, the position of these latter points in the two-dimensional graphical display was viewed with caution, given the poor quality of their representation and/or the low correlation of these points with the first two dimensions. Apart from these profiles, the display was judged to be very informative for the analytical purposes proposed (i.e., the selection of exposure variables for inclusion in the validity scale).

Only 9.8% of the total inertia (variance) of the points was not represented by the first two principal axes. With only a small percentage of the row and column points lying along the third dimension, it was concluded that Figure 2.1 (p.303) was almost an exact representation of the row and column profiles. The original data matrix can be recovered and displayed (in this study) in two dimensions. However, it must be remembered that the row and column profiles are actually located in three-dimensional space. In addition, the two principal axes were assessed as stable because the two principal inertias were not close to each other -- 72.5 and 17.7 (Lebart et al., 1984).

The chi-square value Table 19 (p.294) proved to be highly statistically significant ( $p < .001$ ) indicating a lack of homogeneity (independence) in the contingency table data. Therefore, it could be concluded that a strong association existed between the row and column profiles (i.e., the 33 possible risk factors, and the (attribute) 'perceived' level of risk assigned to them by the study participants).

As mentioned previously, correspondence analysis is a statistical procedure concerned with the vectors of relative frequencies as points in multidimensional space (Lebart et al., 1984; Greenacre, 1984). These vectors are referred to as profiles, and the resulting graphical analyses represent the distances between profiles. The graphs illustrate the configuration of points (profiles) in projection

planes formed by the first principal axes taken two at a time (Lebart et al., 1984). Because these projections were only approximate low-dimensional geometrical displays of a set of high-dimensional points, the graphs were analyzed to determine where the display was either accurate or inaccurate. This assessment included a determination of which profiles were situated close to the plane and those which were not. The following guidelines were used to interpret the graphical results correctly, and ultimately to determine the quality of fit of the model to the data (Greenacre, 1984; Lebart et al., 1984; Moran et al., 1990).

1. **Axis Classification.** Axis classification was useful in the interpretation of the specific displays. The assigning of a descriptive name to the axis is determined by an assessment of which points (profiles) have contributed the most to the building of the axis (Greenacre, 1984). In this case, for example, the column profiles (risk levels) explain the inertia of the first principal axis the best. Specifically, the 147 respondents primarily perceive (classify) the risk factors as either etiologically important (high or moderate risk) or unimportant (no risk) in breast cancer development. In very general terms, the display can be used for informal classification of risk factors by level of risk, that is, factor classification according to etiologic importance. The column profile NO OPINION has contributed most to the construction of the second axis.

2. **Distance Among and Between Profiles.** As it is in principal components analysis, it is legitimate in correspondence analysis to interpret both the distance among the elements of one set of points, and the relative positions of one set with respect to all the points of the other set (Lebart et al., 1984). The separate displays of column and row profiles show the similarities and dispersion within the

respective profiles. The joint or simultaneous display indicates the correspondence between the clouds (of points).

The term 'correspondence analysis' is derived from the direct relationship that exists between the geometry of column and row profiles. This means, for example, that a row profile (risk factor) will tend to a (geometrical) position in its space which corresponds to the column profile (level of risk) that is prominent in that row profile. Similarly, given the display of risk factors, a particular level of risk will position itself along the principal axis in the direction of the risk factors that are predominant in that category (Lebart et al., 1984). In general, in correspondence analysis, one is looking for similarities among profiles. Specifically, risk factors (row profiles) whose points are close to one another have similar risk levels (column profiles).

For example, the fifth column profile - high risk, is situated on the positive side (0.702) of the first principal axis, and any risk factor (row profile) which loads relatively high on this risk level will also lie on the positive side of the first axis, in close proximity: smoking, consumption of dietary fats, breast implants, use of estrogens during menopause, exposure of the breast/chest area to ultraviolet radiation, and the use of oral contraceptives. By analogy, a particular category of 'perceived risk' lies along the principal axes in the direction of the risk factor profiles which are prominent in the category (Greenacre, 1984; Lebart et al., 1984).

Individual profiles similar to the marginal profile yield points close to the origin in correspondence analysis plots (Moran et al., 1990). This means that row points which lie close to the origin (i.e., the center of gravity of the profile points) have a mean profile of risk levels, and as such, are non-discriminating (Lebart et al., 1984).

In this analysis, the following column and row profiles appear to lie near zero when the graphical displays were inspected: low risk, mastitis, injury to the

breast/chest area, alcohol consumption, ingestion of food additives (MSG), exposure to chronic viruses (hepatitis, HIV), the use of artificial sweeteners, and exposure to cleaning solution chemicals. However, contrary to what is portrayed in the displays (Figures 1, 2.1 and 2.2), the position of these profiles was considered to be sufficiently distant from the origin: these points contributed to axis formation and cannot be dismissed as non-discriminating.

Figure 1 (p.302), Figure 2.1 (p.303) and Figure 2.2 (p.304) show the plot of column profiles using the first coordinate axis, and the first two coordinate axes, respectively. These displays compare risk factor proportions in the 5 risk levels. Here we see that the research subjects perceive the risk factors as being either etiologically important (high or moderate risk) or unimportant (no risk) for breast cancer development. Figure 1 (p.302) shows the scaling of risk levels, and reveals that the (approximate chi-square) 'distance' from no risk to high risk is more than three times the distance from moderate risk to high risk. Because the first axis accounts for most of the total inertia (spatial variance of the data), the separate plots provide good scaling for risk factor and level of perceived risk (etiologic importance).

### 3. CA Plots as a Measure of Association Between Rows and Columns

Moran et al. (1990) stressed that "it is not the closeness of a row point to a column point that determines their degree of association, but the comparison of their distances from the origin" (p.642). Large standardized deviate values indicate a strong association (either positive or negative) between row and column profiles. When the standardized deviate is close to zero (the origin), the association between the specific column and row profile is low. A standardized deviate is defined as the difference between the observed frequency and its expected value

when there is independence or lack of association between the row and column variables (Moran et al., 1990, p.641).

Figure 1 (p.302), the plot of the row and column coordinates on the same axis, helped to determine the association between rows and columns. One observes that the high risk column coordinate and six row coordinates which included: smoking, dietary fats, breast implants, estrogen replacement therapy, ultraviolet exposure, and the use of oral contraceptives were both large and positive. These data indicated a positive association between this column profile and the 6 row points. At the same time, 'elavil' (a row coordinate), was large and negative. Therefore, these data suggested that there was a large negative association between the 'high' risk category and the risk factor 'use of the drug elavil'.

Further analysis of Figure 1 (p.302) indicated that the row coordinates (mastitis, traumatic injury to breast/chest area, alcohol consumption, consumption of food additives) were close to zero, so that they will have lower associations with the column profiles. These row profiles were associated with low COR values (i.e., squared correlations -- the variance of a row or column profile which is explained by the principal axis). Consequently, these profiles were not well-represented in the first dimension. Other graphical displays (other dimensions) show these profiles to be located distant from the origin, and as such, these row profiles were considered to be differentiated among the 5 risk levels.

#### 4. Interpretation of the Axes

Figures 2.1 and 2.2 (pp.303-304) show the projection of the row and column profiles onto the plane; however, these displays do not indicate which risk factors and levels of risk have contributed the most in determining the orientation of the axes. "Because the total inertia of each set of points is decomposed along the

principal axes and among the points in similar and symmetric fashion, the inertia of each set of points can be decomposed in a manner analogous to the decomposition of variance. These decompositions are used to assist in the interpretation of the graphical display" (Hoffman and Franke, 1986, p.219). Therefore, to interpret the axes, Lebart et al. (1984) indicate that the following two coefficients must be calculated for each axis, and for all the row and column profiles, and then interpreted:

a. Absolute Contributions to Inertia (CTR). This coefficient is interpreted as "the percentage of (weighted) variance explained by each point in relation to each axis" (Hoffman and Franke, 1986, p.220). In addition, the CTR value "quantifies the importance of each point in determining the direction of the principal axes and serve as guides to the interpretation of each axis" (p.220). A large CTR indicates that the specific row or column profile has played a significant role in the building of the axis. A row or column profile can make a significant contribution to the inertia of a principal axis "when it has a large mass and/or when it is a large distance from the centroid, even when it has relatively low mass" (p.220);

b. Squared Correlations ( $COR = \cos^2 \theta$ ). This coefficient indicates the part of the variance of a row or column profile explained by the principal axis. In other words, the squared correlation indicates the correlation of the row or column profile with the principal axis. COR values are used to determine the "*quality*" of the representation of each point in the display. These values are independent of the point's mass and indicate how well each point is "fit" by the representation" (p.220). If the COR value is high, the profile is considered to be located near the principal axis (i.e., the axis explains the point's inertia very well); there is little error in its graphical display (Lebart et al., 1984; Hoffman and Franke, 1986). However, if the COR coefficient is low, "the profile vector 'correlates' with the

principal axis (i.e., it lies in the direction of the principal axis" (Greenacre, 1984, p.70).

Table 22 (pp.298-299) and Table 23 (pp.300-301) are used for the interpretation of the first two principal axes based on the consideration of the QLT values, as well as the absolute contributions (CTR) and the squared correlations (COR) of the respective row (risk factor) profiles and column profiles (levels of perceived risk).

If the COR ( $\cos^2\theta$ ) was  $> 70\%$ , the column or row profile was exclusive in the creation or building of the axis. Otherwise, it contributed more to the building of the other axes. In addition, if the CTR value was  $> 30\%$ , the specific profile had contributed strongly to the creation of the particular axis. QLT values greater than .600 indicated that the specific points were well-represented in the two-dimensional display shown in Figures 2.1 (p.303) and 2.2 (p.304). In these analyses, only 4 profiles were not adequately represented in Figures 2.1 and 2.2. These included mastitis, exposures to chemicals in cleaning solution, the ingestion of food additives such as MSG, and factors which were associated with the fourth profile (column profile) - the low risk category. Otherwise, the graphical displays were considered accurate for the remaining column and risk profiles.

Based on high COR values, the following row and column profiles were exclusive in their characterization of the first axis. Here, the first axis explained a significant proportion of the variance of these profiles, and as such, there was little error in their graphical display: smoking, dietary fats, maternal use of DES, stress, breast implants, ultraviolet exposure, chemical exposure, ureaformaldehyde exposure, asbestos exposure, exposure to herbicides (weed killers), use of an antihypertensive (reserpine, capoten), pain killers, diuretics, other medications for high blood pressure, elavil, menopausal estrogen, oral contraceptives, and both moderate and high risk categories ( a total of 21 out of 38 profiles). The row profile

- injury to the breast and chest area - contributed strongly to the creation of axis 2. Because the remaining profiles had low CTR but high COR values, it was concluded that these profiles did not have a strong influence in building axes 1 and 2 respectively.

The formation of the first two principal axes and the two-dimensional graphical projections are primarily based on the classification of factors by level of perceived risk for breast cancer development (i.e., classification by etiologic importance).

It appears that two axes predominate and account for 90.2% of the total variance as reported in Table 19 (p.294). As a result, the interpretation of results will be limited to the first two axes only. Figures 2.1 and 2.2 (pp.303-304) show the principal relationships among the profiles with respect to the first two axes. In viewing these representations, Lebart et al. (1984) noted that one must perform successive interpretations of proximities among rows, among columns, and finally, the simultaneous representation of both spaces.

Therefore, based on the simultaneous representation and interpretation of the column and row profiles, the classification of risk factors by level of perceived risk (etiologic importance/plausibility) were as follows:

#### ASSOCIATED LEVEL OF RISK

High Risk

Moderate Risk

#### RISK FACTOR

smoking  
dietary fats  
breast implants  
postmenopausal estrogen  
uv exposure  
oral contraceptives

nitrites/nitrites  
stress  
steroids  
DES (diethylstilbesterol)  
asbestos  
phenoxyherbicide

	coffee ureaformaldehyde mastitis alcohol food additives (MSG) injury to breast/chest exposure to chemicals
Low Risk	virus use of sweeteners cleaning solutions dysmenorrhea thyroxin
No Risk	reserpine deodorants elavil capoten diuretics antihypertensive drugs size of breast use of pain killers
No Opinion	fake drug (mellal)

The association between risk level and risk factors was the strongest for high risk, no risk, and no opinion categories (column profiles). Moderate risk took an intermediate position with respect to the association between risk level (column profiles) and risk factors (row profiles). There was a low association between the risk level profile (low risk) and the risk factors that lay close to it.

Based on the results of the preceding correspondence analysis, the following variables were selected as the validity scale items for inclusion in Q2:

- a. traumatic injury to breast or chest area (A);
- b. use of artificial sweeteners (C);
- c. nicotine exposure (smoking) - (D);
- d. dietary fat consumption (E);
- e. history of mastitis during breast feeding (H);
- f. maternal use of DES - diethylstilbesterol (J);
- g. significant periods of personal distress (L);

- h. sunbathing and overexposure to ultraviolet radiation (N);
- i. chemical exposure - paints, paint removers, solvents and wood stains (S);
- j. exposure to ureaformaldehyde (P);
- k. asbestos exposure (Q);
- l. exposure to phenoxyherbicides (chemical weed killers) - (R);
- m. fake drug - mellar (Y);
- n. drug exposure (steroids) - (CC);
- o. drug exposure (reserpine) - (V);
- p. use of oral contraceptives (GG);
- q. use of postmenopausal estrogen (FF);
- r. use of monosodium glutamate - MSG (U);
- s. consumption of foods containing nitrates/nitrites (EE); and,
- t. alcohol consumption (I).

Differential exposure misclassification (recall bias) was assessed to be present if any of three participant groups identified a larger number of the validity scale items in comparison to the other groups. It must be remembered that if an exposure variable is not related to disease outcome, the proportion of cases and controls exposed to this variable should be approximately the same. Therefore, the exposure-odds ratio would then approximate 1.00 (indicating no risk association with exposure). The risk estimates for the factors included in the validity scale are reported in Table 13 (pp.278-280). Only 4 factors had significant odds ratios: 1) a previous history of traumatic injury to the breast/chest area; 2) nicotine exposure (smoking); 3) consumption of foods containing nitrites and nitrates; and, 4) consumption of dietary fats (i.e., high, normal, low). These factors have received considerable popular media coverage as risk factors for a variety of diseases

including breast cancer. In addition, in Dr. Waxler-Morrison's research (personal communication), breast cancer patients have personally identified these factors as being responsible for their breast cancer. Therefore, these factors may be susceptible to exposure misclassification, and group differences may be noticed in the analysis of case-control differences with respect to the validity scale summary score (**VSSCORE**).

The total validity scores for cases and controls were calculated using the individual factor weights derived from the correspondence analysis of the 33x5 data matrix. A factor's weighting is the distance (FACT T) that the variable is located from the origin (i.e., its strength of association as measured by chi square). Rows and columns that are located together far from the origin are considered to be more strongly associated. The values for the factor weightings are found in Table 20 (FACT T), (pp.295-296).

The total validity scale score (**VSSCORE**) as a function of the risk factor items and their specific weightings can be summarized as follows:

**VSSCORE** (an additive model) = (.061)(history of breast injury) + [(0.795)(smoking history) + (.795)(amount smoked) + (.795)(duration (years) smoked)] + (.093)(history of mastitis) + (.785)(level of dietary fat consumption) + [(0.036)(history of alcohol consumption) + (.036)(amount consumed) + (.036)(duration (years) alcohol consumed)] + (.338)(use of DES by mother) + (.343)(history of stress) + (.601)(overexposure to ultraviolet radiation) + [(0.279)(exposure to chemicals in paints, paint removers, solvents and wood stains) + (.279)(frequency of exposure to chemicals in paints, paint removers, solvents and wood stains)] + [(0.212)(history of exposure to ureaformaldehyde) + (.212)(duration of exposure)] + [(0.338)(history of exposure to asbestos) + (.338)(frequency of exposure to asbestos)] + [(0.312)(exposure to chemical weed killers) + (.312)(frequency of exposure to chemical weed killers)] + (-.903)(use of mella) + (.341)(use of steroids) + (.411)(consumption of foods containing nitrites/nitrates) + [(0.630)(use of postmenopausal estrogen) + (.630)(duration of use (years))] + [(0.586)(use of oral contraceptives) + (.586)(duration of use (years))] + (.009)(consumption of MSG) + (-.134)(use of artificial sweeteners) + (-.660)(use of reserpine for elevated blood pressure).

Prior to the estimation of the VSSCOREs for the cases and control groups one and two, and the comparison of the study groups with respect to differences in the VSSCORE, it was necessary to perform an analysis of variance for each scale

item to determine if there were any significant group differences at the 0.05 level of significance. The presence of group differences could possibly interact (i.e., cancel out) when the results for each item are aggregated into the composite VSSCORE. Consequently, any comparison of VSSCORE for the three study groups would fail to detect differential reporting of exposure by the cases and the controls.

The results of this analysis can be found in Tables 14.1 and 14.2 (pp.281-284). The difference between these two tables is in the VSSCORE calculation: Table 14.1 (pp.281-282) applied the factor weightings to all the composite factors which included an assessment of exposure to the variable, as well as the frequency and duration of exposure (i.e.,  $[(.036)(\text{history of alcohol consumption}) + (.036)(\text{amount consumed}) + (.036)(\text{duration (years) alcohol consumed})]$ ). Unfortunately, during the administration of the validity scale questionnaire, the study subjects were not questioned on the importance of duration and frequency of use/exposure for the various study factors; consequently weighting for the specific factor components of duration and frequency were not obtained by correspondence analysis. It was not clear whether or not the same weighting factor should be applied uniformly to a factor consisting of exposure, duration and frequency components.

Therefore, for the specific group comparisons, the analysis of variance assessment (ANOVA) and VSSCORE determinations were done separately, as follows: 1) using the weighting factors for duration and frequency - Table 14.1 (pp.281-282); and, 2) not using the specific weighting factors - Table 14.2 (pp.283-284). A third analysis was completed for exposure variables identified by the breast cancer patients in Dr. Waxler-Morrison's research (personal communication) as being responsible for the development of their breast cancer. From this study, women identified 19 variables as being relevant. These included a

previous history of breast injury, the size of a woman's breasts, use of artificial sweeteners, cigarette smoking, dietary factors (including fats), history of fibrocystic disease, excessive coffee consumption, chronic breast infection (mastitis), regular use of alcohol, stress, silicone implants, use of estrogen containing drugs, exposure to chemical weed killers (as reported by the cases living in the Fraser Valley area), food additives, use of oral contraceptives, familial history of breast cancer, consumption of nitrites and nitrates in food products, and exposure to radiation and electromagnetic fields. Only the last two exposure variables (i.e., exposure to radiation and electromagnetic fields had not been included in the study questionnaire used to select exposures for inclusion in the validity scale (Appendix 2). Therefore, it was possible to analyze 17 of these variables by correspondence analysis to determine their specific weighting factors (etiologic importance). However, only 13 of these variables were included in Q2 (the study questionnaire), and could therefore be used in the determination of the VSSCOREs for group comparisons - Table 14.3 (pp.285-286).

The ANOVA indicated that the prevalence of significant differences between the groups on individual scale items is low (i.e., group differences were seen for 3 factors when weightings were used, and for 5 factors when the weightings were only applied to exposure history. For the analysis using those exposure variables identified by breast cancer patients, 3 factors (identical to the ones found in the original analysis - Table 14.1 (pp.281-282) were found to have group differences. These results suggest that the individual scale items can be aggregated into a summary score (VSSCORE), and that group scores may be compared to determine differences in recall accuracy, and possibly to detect differential exposure misclassification (i.e., recall bias).

The VSSCORE analyses - Tables 14.1, 14.2 and 14.3 (pp.281-286) indicated that there were no two groups different at the 0.05 level for the total VSSCORE.

The results were consistent with the main study results, suggesting that differential exposure misclassification bias was not operating to distort the estimated exposure-disease odds ratios in this case-control study of breast cancer. Specifically, cases and controls did not change their reports of exposure post-diagnosis in any systematic and biased manner as a result of the knowledge of their group status. Retrospectively collected exposure information was fairly reliable, and the knowledge of group status (i.e., diagnosis) did not appear to invalidate study conclusions regarding risk estimates. However, random exposure misclassification error was present in this study and the risk estimates were influenced overall, by the resulting measurement error, but not by bias.

If significant differential exposure misclassification (recall bias) had been detected, the OR estimates for the real study factors in this case-control study could have been adjusted proportionately to the group differences on the VSSCORE.

#### 4.9 Summary

In summary, a nested case-control study was conducted to determine overall, the reliability and validity (i.e., 'criterion-validity') of exposure information provided retrospectively by cases and controls after they were aware of their diagnosis. The specific objectives of this study were to determine if the cases and controls reported their past exposures significantly differently, to determine if non-differential and/or differential exposure misclassification were present, and to assess the impact of any resulting exposure misclassification on the estimates of relative risk (i.e., exposure-disease odds ratios) for each study factor. The results were then used to evaluate the effectiveness and efficiency of an

exposure data validity scale designed for the assessment and control of differential exposure misclassification.

This chapter provided evidence from the varied analyses performed that the retrospectively collected exposure data were recalled fairly reliably. Kappa coefficients  $>.75$  suggested a high level of consistency (reproducibility) between Q1 and Q2 reports of exposure. Any inconsistencies in the post-diagnostic (Q2) reports of exposure were random, and definitely not related to case-control status. However, these results are not conclusive. Due to the study limitations discussed in Chapter 5, the level of agreement between the prospective and retrospective reports of exposure may have been overestimated. The correlation or lack of independence between Q1 and Q2 response sets may have added sufficient 'noise' into the system to preclude the ability to detect non-differential and differential exposure misclassification, and to assess their impact on measures of association.

When the exposure-disease odds ratios and their 95% confidence intervals were compared for the Q1 and Q2 reports of exposure, the estimates of effect were found to be comparable; overlap of the confidence intervals suggested the absence of differential exposure misclassification. In addition, when the differential in reporting past exposure was compared directly for the cases and their matched controls, no significant differences in recall accuracy were observed. Thus, the accumulated results presented so far would suggest that there was little evidence to support a systematic shift in self-reported exposure by either the cases or the controls: none of the three study groups demonstrated a propensity to understate/underreport or overstate/overreport their exposure to the included study factors. It could be concluded that the risk estimates had not been biased: the exposure-disease odds ratios were only subject to random exposure misclassification error.

Overall, there was no strong and conclusive evidence to provide substantiation for the hypothesis that case-control studies are susceptible to differential exposure misclassification (recall bias). There was no reason to believe that the conclusions from case-control studies are invalid as a result of differential exposure misclassification bias. However, if this study was unable to detect NDEM, the 'noise' generated as a result of the presence of NDEM may have prevented the detection of DEM.

Study objective 4 -- to determine the factors that may be responsible for case-control differences in exposure recall, and for the subsequent occurrence of exposure misclassification -- could not be addressed in this study due to the fairly high reliability of Q2 exposure data, and the low frequency of changes in exposure between Q1 and Q2. For most factors, the percent discordance was < 10%, and for the remainder, discordance did not exceed 20%.

Lastly, the results of the exposure data validity scale concurred with the results from the main study. Specifically, the validity scale analysis suggested that the study groups did not systematically overreport or underreport exposure to the factors which were 'plausible but irrelevant' for breast cancer development. A comparison of the total VSSCORE for the study groups indicated some group differences possibly related to random error in exposure reporting -- differential exposure misclassification was not suggested by the comparison of the VSSCOREs (i.e., there were no significant group differences at the 0.05 level of significance).

The 'exposure data validity scale' appeared to be a successful design strategy for the measurement and the control of differential exposure misclassification. If used, the results of case-control studies could possibly be strengthened, and significant research findings in etiologic studies would not be rejected and abandoned due to criticism directed at the methodological limitations of retrospective case-control research. For example, significant research findings

would not be interpreted as "spurious associations, artifacts due only to the methodological limitations of the case-control design" (Mackenzie, 1986). The construction and use of an 'exposure data validity scale' in every case-control study would permit the routine validation of the quality of the exposure data collected in those studies. In this way, the researcher could comment directly on the validity of the study's conclusions, and correct estimates of risk according to the results of the validity scale analysis. Its usefulness and effectiveness must now be evaluated further, that is, replicated, to determine its utility (i.e., its strengths and limitations) within other research domains using other samples, exposures and health outcomes. One study is insufficient for concluding that an 'exposure data validity scale' is an appropriate design strategy for the measurement and the control of differential exposure misclassification in every case-control study. Replication studies are required to validate the scale and specifically, its ability to measure differential exposure misclassification (recall bias), and to improve its general methodology.

Table 3: Estimated Response Rates for the Three Study Comparison Groups

Study Group	Eligible Subjects	Respondents (N=1177)	Non-Respondents (N=217)	Response Rate By Study Group (%)
<b>CASES</b>  Women with a histologically confirmed diagnosis of breast cancer	280	234	46	83.6
<b>CONTROL GROUP ONE</b>  Women with an abnormal mammogram but no cancer (i.e., Anamnestic Controls)	556	464	92	83.5
<b>CONTROL GROUP TWO</b>  Normal mammogram, no breast cancer (i.e., healthy controls)	558	479	79	85.8

Table 4: Reasons Given for Non-Participation and Non-Response

REASON	(Number) and Overall Percentage (%) of the Total Number of Non-Respondents (N=217)	
1. Language - the inability to comprehend and to respond to the study questionnaire independently	(12)	5.5%
2. Age - the reported inability to remember the exposure information that was being requested in the study questionnaire	(17)	7.8%
3. Poor health - either self, or an immediate family member	(8)	3.7%
4. Unqualified refusal - no reason given for non-participation	(30)	13.8%
5. Qualified refusal - an explicit statement that the subject did not wish to participate in this research study	(47)	21.7%
6. Subject cannot be located or has moved out of the province	(60)	27.6%
7. Initial agreement to participate, but then the subject was lost-to-follow-up	(34)	15.7%
8. Away on vacation or out of the country	(3)	1.4%
9. Woman had died	(6)	2.8%

**Table 5: Comparison of the Study Population to the Screening Mammography First-time and Returning Participants Regarding Demographic Factors and the Risk Factor Profile**

<b>Characteristic</b>	<b>Study Population (%)</b>	<b>SMP BC First-Time Participants (%)</b>	<b>SMP BC Return Participants (%)</b>
<b>a. Age (years)(*)</b>			
• < 40	0.3	0.4	0.1
• 40-49	14.5	39.2	31.6
• 50-59	22.6	28.3	32.8
• 60-69	34.2	21.7	26.4
• ≥ 70	28.4	10.4	9.1
<b>b. Menopausal Status (*)</b>			
• Premenopausal	17.8	35.4	31.7
• Postmenopausal	82.2	55.1	61.7
<b>c. Marital Status</b>			
• Married (includes common-law)	69.4	72.5	73.1
• Divorced, widowed, separated	25.8	20.9	20.7
• Never married	4.8	5.7	6.7
<b>d. Ethnic Origin</b>			
• Caucasian	95.6	91.4	93.1
• Asian (Japanese, Chinese)	2.5	3.6	3.4
• Other	1.9	3.6	2.8
<b>e. Education (*)</b>			
<b>(categories determined by SMP BC)</b>			
• Did not graduate	-	24	17.9
• High school graduate	50.6	24.4	25.9
• Post-secondary	43.6	50.5	56.2
<b>(categories used in study protocol)</b>			
• Elementary school	5.8	-	-
• High school graduate	50.6	-	-
• Post-secondary education	43.6	-	-

Notes: 1. The Screening Mammography Program Population Statistics were compiled from the 1990-1991 Annual Report of the Screening Mammography Program of British Columbia, pp. 23-26.

2. (\*) denotes the study factors for which group differences were observed.

**Table 5 (continued): Comparison of the Study Population to the Screening Mammography First-time and Returning Participants Regarding Demographic Factors and the Risk Factor Profile**

Characteristic	Study Population (%)	SMP BC First-Time Participants (%)	SMP BC Return Participants (%)
<b>f. Parity</b>  <ul style="list-style-type: none"> <li>• Nulliparous</li> <li>• Parous</li> </ul>	16.3 83.7	15.5 83.9	34.9 65.1
<b>g. Age at first birth (years)</b>  <ul style="list-style-type: none"> <li>• &lt; 20</li> <li>• 20-29</li> <li>• ≥ 30</li> <li>• nulliparous</li> </ul>	15.2 53.8 14.7 16.3	12.3 59.4 11.6 15.5	6.9 47.4 10.8 34.9
<b>h. Family history of breast cancer (i.e., any first-degree relatives)</b>  <ul style="list-style-type: none"> <li>• No</li> <li>• Yes, lower risk (unilateral and post-menopausal)</li> <li>• Yes, higher risk (bilateral or pre-menopausal)</li> </ul>	82.7 4.9 3.1	87.3 5.7 4.9	84.5 6.6 5.7
<b>i. Oral contraceptives (*)</b>  <ul style="list-style-type: none"> <li>• Current or ex-user</li> <li>• Non-user</li> </ul>	56.2 43.8	43.3 55.3	59.5 40.3
<b>j. Estrogen-containing drugs (ERT) (*)</b>  <ul style="list-style-type: none"> <li>• Current or ex-user</li> <li>• Non-user</li> </ul>	66.2 33.8	70.6 27.7	69.0 30.7
<b>k. Cigarette smoking (*)</b>  <ul style="list-style-type: none"> <li>• Current or ex-smoker</li> <li>• Non-smoker</li> </ul>	42.2 57.8	49.1 49.8	62.3 37.7

Table 6: Demographic, Medical, Reproductive and Lifestyle Characteristics of the Study Comparison Groups

CHARACTERISTIC	CASES (N=234)	CONTROL GROUPS	
		GROUP ONE (Abnormal Mammogram/No Breast Cancer) (N=464)	GROUP TWO (Normal Mammogram/No Breast Cancer) (N=479)
DEMOGRAPHIC FACTORS			
a. Marital Status (%) • Ever married • Married (includes common-law) • Divorced • Separated • Widowed • Never Married	95.6 72.2 6.5 3.0 13.9 4.4	94.9 66.9 10.5 3.4 14.1 5.1	95.0 69.2 7.3 1.9 16.6 5.0
b. Ethnicity (%) • Caucasian (white) • Asian (Chinese, Japanese) • Other	94.8 2.6 2.6	96.3 2.1 1.6	95.9 2.7 1.4
c. Age Distribution (years) (%) • 30-49 • 50-59 • 60-69 • ≥ 70	14.3 19.2 35.2 31.3	16.6 26.6 34.5 22.3	13.6 22.1 32.8 31.5
Mean age (years) of each study group	63.2	61.3	63.1
d. Occupational Status (%) • Retired/Unemployed • Student • Housewife • Vocational/Trade • Professional • Clerical, Sales and Manufacturing	45.2 0.5 29.0 1.4 12.4 11.5	33.3 0.7 35.7 0.2 14.5 15.6	39.5 - 32.6 1.1 12.5 14.3
e. Education (%) (Highest school grade achieved) • Elementary school • Secondary school • Post-secondary education	6.7 48.0 45.3	4.3 52.5 43.3	6.3 51.3 42.5
f. Socioeconomic Status (%) • Upper • Middle • Lower	8.3 86.0 5.7	9.2 86.6 4.3	7.5 86.2 6.2

Table 6 (continued): Demographic, Medical, Reproductive and Lifestyle Characteristics of the Study Comparison Groups

CHARACTERISTIC	CASES (N=234)	CONTROL GROUPS	
		GROUP ONE (Abnormal Mammogram/No Breast Cancer) (N=464)	GROUP TWO (Normal Mammogram/No Breast Cancer) (N=479)
DEMOGRAPHIC FACTORS (continued)			
g. Location of Residence (%) • Urban • Rural	79.1 20.9	76.7 23.3	81.6 18.4
HEALTH PROMOTION PRACTICES AND DIAGNOSTIC PROCEDURES			
a. Practises breast self-examination (BSE) regularly (%)	72.8	74.3	76.6
b. Mean frequency of BSE over 12 months	11.0	12.5	9.4
c. (%) of subjects who have had at least one needle breast aspiration	9.2	11.4	6.9
d. (%) of subjects who have had a previous mammogram	39.6	48.1	55.4
MEDICAL HISTORY			
a. Mean age (years) at menarche	13.1	13.0	13.0
b. Mean age (years) at menopause	54.2	54.9	54.2
c. Mean (years) total ovulatory activity [Age menopause minus age menarche]	41.1	41.7	41.2
d. Present menopausal status (%) 1. Still menstruating 2. Menopausal	16.2 83.8	19.6 80.4	17.5 82.5
e. Type of menopause (%) 1. Natural 2. Prior hysterectomy 3. Prior oophorectomy	63.5 18.8 17.7	57.6 31.3 11.1	56.6 32.5 10.9

Table 6 (continued): Demographic, Medical, Reproductive and Lifestyle Characteristics of the Study Comparison Groups

CHARACTERISTIC	CASES (N=234)	CONTROL GROUPS	
		GROUP ONE (Abnormal Mammogram/No Breast Cancer) (N=464)	GROUP TWO (Normal Mammogram/No Breast Cancer) (N=479)
OCCURRENCE OF SOMATIC CHANGES POSSIBLY RELATED TO OVULATORY ACTIVITY			
a. Somatic changes preceded menses (%)			
1. No	9.0	7.4	7.3
2. Sometimes	38.4	42.9	44.1
3. Always	49.6	49.0	48.4
4. Uncertain	3.0	0.7	0.2
b. Breast swelling period-related (%)			
1. Yes	42.1	51.0	48.7
2. No	57.9	49.0	51.3
c. History of breast pain and tenderness (%)			
1. Yes	54.5	63.5	52.9
2. No	45.5	36.5	47.1
d. Symptoms related to breast tenderness (%)			
1. Pain	6.9	4.1	9.3
2. Tenderness	66.6	75.5	71.1
3. Both pain and tenderness	26.5	20.4	19.6
e. Breast pain and tenderness related to periods (%)			
1. Yes	44.8	49.3	41.1
2. No	55.2	50.7	58.9
REPRODUCTIVE HISTORY			
a. History of pregnancy (%)			
1. Ever pregnant	86.5	86.0	78.7
2. Never pregnant	13.5	14.0	21.3
b. Mean number of pregnancies (live births only)	2.2	2.4	2.3
c. Mean age (years) at first full term pregnancy	24.7	24.3	24.7

Table 6 (continued): Demographic, Medical, Reproductive and Lifestyle Characteristics of the Study Comparison Groups

CHARACTERISTIC	CASES (N=234)	CONTROL GROUPS	
		GROUP ONE (Abnormal Mammogram/No Breast Cancer) (N=464)	GROUP TWO (Normal Mammogram/No Breast Cancer) (N=479)
USE OF EXOGENOUS HORMONES			
a. Use of oral contraceptives (%) 1. Ever use 2. Never	38.6 61.4	48.8 51.2	44.1 55.9
b. Mean duration (years) of oral contraceptive use among ever users	5.2	5.4	5.7
c. Estrogen replacement therapy [ERT] (postmenopausal) (%) 1. Ever use 2. Never	33.3 66.7	29.7 70.3	38.5 61.5
d. Among ever users, mean duration (years) of ERT use	8.2	7.1	7.9
ANTHROPOMETRIC CHARACTERISTICS			
Quetelet's Index (body mass to height index defined as [kg/m <sup>2</sup> ]) (%) 1. Low (≤ 21) 2. Normal (22-24) 3. High (≥ 25)	14.0 20.1 65.9	18.3 29.2 52.5	16.5 30.3 53.2
FAMILY HISTORY OF BREAST CANCER (First Degree Relatives)			
a. Positive history of breast cancer in a first degree relative (i.e., mother and/or sister(s)) (%)	19.5	13.8	18.5
b. Mother with breast cancer (%)	9.0	6.9	7.4
c. Sister(s) with breast cancer (%)	10.5	6.9	11.1
d. Occurrence of premenopausal breast cancer in first degree relatives (%)	1.8	1.3	2.0
e. (%) First degree relatives diagnosed with: 1. unilateral breast cancer 2. bilateral breast cancer	8.6 0.5	5.5 0.7	6.4 0.9

Table 6 (continued): Demographic, Medical, Reproductive and Lifestyle Characteristics of the Study Comparison Groups

CHARACTERISTIC	CASES (N=234)	CONTROL GROUPS	
		GROUP ONE (Abnormal Mammogram/No Breast Cancer) (N=464)	GROUP TWO (Normal Mammogram/No Breast Cancer) (N=479)
SMOKING HISTORY			
a. % ever smoker	64.9	59.4	49.1
b. % never smoker	35.1	40.6	50.9
c. Mean number of cigarettes smoked per day	15.6	16.2	14.2
d. Mean duration (years) of smoking	16.3	15.9	12.2
ALCOHOL USE HISTORY			
a. % ever drank	69.5	62.5	66.2
b. % never drank	30.5	37.5	33.8
c. Mean number of drinks per month	13.5	13.8	11.7
d. Mean duration (years) of alcohol consumption	23.2	20.2	23.3
OTHER FACTORS			
Mean time (months) between the administration of Q1 and Q2	12.9	9.5	9.5

Table 7: Demographic, Medical, Reproductive and Lifestyle Characteristics of Study Respondents and Non-Respondents

Characteristic		Respondents (N=1177)	Non-Respondents (N=217)
<b>DEMOGRAPHIC FACTORS</b>			
a.	<b>Marital Status (%)</b> • Ever married • Married (includes common-law) • Divorced • Separated • Widowed • Never married	95.1 69.4 8.5 2.4 15.4 4.8	92.8 57.1 12.4 2.7 21.0 7.1
b.	<b>Ethnicity (%)</b> • Causasian (white) • Asian (Japanese, Chinese) • Other	95.6 2.5 1.9	92.8 4.4 2.8
c.	<b>Age Distribution (years) (%)</b> • 30-49 • 50-59 • 60-69 • ≥ 70	14.8 22.6 34.2 28.4	16.2 19.1 33.3 31.4
	Mean age (years)	62.4	62.2
d.	<b>Occupational Status (%)</b> • Retired/Unemployed • Student • Housewife • Vocational/Trade • Professional • Clerical, Sales, Manufacturing	39.3 0.6 32.4 0.9 13.0 13.8	36.5 0.5 33.0 2.0 10.0 18.0
e.	<b>Education (%)</b> (Highest school grade achieved) • Elementary school • Secondary school • Post-secondary education	5.8 50.6 43.6	13.7 52.9 33.4
f.	<b>Socioeconomic Status (SES) (%)</b> • Upper • Middle • Lower	8.3 86.0 5.7	7.4 78.9 13.7
g.	<b>Location of Residence (%)</b> • Urban • Rural	79.1 20.9	86.4 13.6

Table 7 (continued): Demographic, Medical, Reproductive and Lifestyle Characteristics of Study Respondents and Non-Respondents

	Respondents (N=1177)	Non-Respondents (N=217)
<b>HEALTH PROMOTION PRACTICES AND DIAGNOSTIC PROCEDURES</b>		
a. Practises breast self-examination (BSE) regularly (%)	74.6	66.5
b. Mean frequency of BSE over 12 months	10.7	11.0
c. (%) of subjects who have had at least one needle breast aspiration	9.2	5.8
d. (%) of subjects who have had a previous mammogram	47.7	37.7
<b>MEDICAL HISTORY</b>		
a. Mean age (years) at menarche	13.0	12.9
b. Mean age (years) at menopause	54.5	53.0
c. Mean (years) total ovulatory activity [age at menopause minus age at menarche]	41.3	40.0
d. Present Menopausal Status (%)		
1. Still menstruating	17.8	15.7
2. Menopausal	82.2	84.4
e. Type of Menopause (%)		
1. Natural	59.3	64.2
2. Prior hysterectomy	27.5	26.6
3. Prior oophorectomy	13.2	9.2
<b>OCCURRENCE OF SOMATIC CHANGES POSSIBLY RELATED TO OVULATORY ACTIVITY</b>		
a. Somatic changes preceded menses (%)		
1. No	7.9	11.8
2. Sometimes	41.8	36.9
3. Always	49.0	43.1
4. Uncertain	1.3	8.2
b. Breast swelling period-related (%)		
1. Yes	47.3	35.2
2. No	52.7	64.8
c. History of breast pain and tenderness (%)		
1. Yes	57.0	46.3
2. No	43.0	53.7

Table 7 (continued): Demographic, Medical, Reproductive and Lifestyle Characteristics of Study Respondents and Non-Respondents

Characteristic	Respondents (N=1177)	Non-Respondents (N=217)
<b>OCCURRENCE OF SOMATIC CHANGES POSSIBLY RELATED TO OVULATORY ACTIVITY (continued)</b>		
d. Symptoms related to breast tenderness (%)		
1. Pain	6.8	7.3
2. Tenderness	71.1	69.5
3. Both pain and tenderness	22.1	23.2
e. Breast pain and tenderness related to periods (%)		
1. Yes	45.1	59.3
2. No	54.9	40.7
<b>REPRODUCTIVE HISTORY</b>		
a. History of pregnancy (%)		
1. Ever pregnant	83.7	80.7
2. Never pregnant (i.e., nulliparous)	16.3	19.3
b. Mean number of pregnancies (live births only)	2.3	2.6
c. Mean age (years) at first full-term birth	24.5	23.7
d. Mean duration of breastfeeding among those who engaged in breast feeding (months)	4.8	1.5
<b>USE OF EXOGENOUS HORMONES</b>		
a. Use of oral contraceptives		
1. Ever use (%)	43.8	40.9
2. Never (%)	56.2	59.1
b. Mean duration of oral contraceptive use among ever users (years)	5.5	5.8
c. Estrogen replacement therapy [ERT] (postmenopausal)		
1. % ever menopausal hormone use	33.8	28.0
2. % never used	66.2	72.0
d. Among ever users mean duration (years) of ERT	7.7	8.8
<b>ANTHROPOMETRIC CHARACTERISTIC</b>		
Quetelet's Index (body mass to height index defined as $\text{kg/m}^2$ ) (%)		
1. Low ( $\leq 21$ )	16.3	15.1
2. Normal (22-24)	26.5	25.9
3. High ( $\geq 25$ )	57.2	59.0

Table 7 (continued): Demographic, Medical, Reproductive and Lifestyle Characteristics of Study Respondents and Non-Respondents

Characteristic	Respondents (N=1177)	Non-Respondents (N=217)
<b>FAMILY HISTORY OF BREAST CANCER</b> (First Degree Relatives)		
a. Positive history of breast cancer in a first degree relative (i.e., mother and/or sister(s)) - (%)	17.3	11.3
b. Mother with breast cancer (%)	7.8	6.9
c. Sister(s) with breast cancer (%)	9.5	5.4
d. Occurrence of premenopausal breast cancer in first degree relatives (%)	1.7	1.5
e. (%) First degree relatives diagnosed with: 1. unilateral breast cancer 2. bilateral breast cancer	6.80 0.70	5.45 0.99
<b>SMOKING HISTORY</b>		
a. % ever smoker	57.8	47.6
b. % never smoker	42.2	52.4
c. Mean number of cigarettes smoked per day	15.4	16.7
d. Mean duration (years) of smoking	26.2	26.1
<b>ALCOHOL USE HISTORY</b>		
a. % ever drank	66.1	55.6
b. % never drank	33.9	44.4
c. Mean number of drinks per month	12.9	13.9
d. Mean duration (years) of alcohol consumption	22.0	17.6

Table 8: Pearson Product Moment and Spearman Rank Correlations of Prospective and Retrospective Reports of Exposure for Variables and Conditions Possibly Related to Breast Cancer Risk - A Summary by Magnitude

(0.00 - 0.49) LITTLE/NO RELATIONSHIP LOW AGREEMENT/CORRELATION	(0.50 - 0.69) MODERATE AGREEMENT/CORRELATION	(0.70 - 1.00) HIGH LEVEL OF AGREEMENT TO PERFECT CORRELATION
		<b>DEMOGRAPHIC FACTORS</b> <ul style="list-style-type: none"> <li>• Age</li> <li>• Ethnicity</li> <li>• Marital Status</li> <li>• Education (i.e., highest level of schooling achieved)</li> </ul>
		<b>MEDICAL HISTORY</b> <ul style="list-style-type: none"> <li>• Age at first menstrual period</li> <li>• Menopausal status (pre- versus postmenopausal)</li> <li>• Among postmenopausal women, age (years) when period stopped</li> <li>• Type of menopause (natural, hysterectomy, bilateral oophorectomy)</li> <li>• Hysterectomy (yes versus no)</li> <li>• Bilateral oophorectomy (yes versus no)</li> </ul>
	<b>REPRODUCTIVE HISTORY</b> <ul style="list-style-type: none"> <li>• Pregnancy history (ever pregnant versus never) (nulliparous)) - <u>(Control Group Two)</u></li> </ul>	<b>REPRODUCTIVE HISTORY</b> <ul style="list-style-type: none"> <li>• Pregnancy history (ever pregnant versus never (nulliparous)) - <u>Cases and Control Group One</u></li> <li>• Parity (i.e., number of live full term births)</li> <li>• Age (years) at first full term pregnancy</li> </ul>

Note: The correlation coefficients are similar for all three comparison groups unless otherwise specified.

References: 1. Munro BH and Page BE. **Statistical Methods for Health Care Research**. Philadelphia: J.B. Lippincott Company, 1993, p. 181.  
2. Colton T. **Statistics in Medicine**. Boston: Little Brown and Company, 1974, pp. 210-211.

Table 8 (continued): Pearson Product Moment and Spearman Rank Correlations of Prospective and Retrospective Reports of Exposure for Variables and Conditions Possibly Related to Breast Cancer Risk - A Summary by Magnitude

<p>(0.00 - 0.49) LITTLE/NO RELATIONSHIP LOW AGREEMENT/CORRELATION</p>	<p>(0.50 - 0.69) MODERATE AGREEMENT/CORRELATION</p>	<p>(0.70 - 1.00) HIGH LEVEL OF AGREEMENT TO PERFECT CORRELATION</p>
		<p><b>FAMILY HISTORY OF BREAST CANCER</b></p> <ul style="list-style-type: none"> <li>• breast cancer history in any first degree relative (i.e., mother, sister(s))</li> <li>• Type and characteristics of the breast cancer diagnosed in first degree relatives (i.e., premenopausal or postmenopausal; unilateral or bilateral)</li> </ul>
		<p><b>SMOKING HISTORY</b></p> <ul style="list-style-type: none"> <li>• Ever versus never smoked</li> <li>• Among ever smokers, age (years ) when started and stopped smoking, as well as frequency (i.e., number of cigarettes smoked per day)</li> <li>• Duration of smoking - total number of years smoked</li> </ul>
	<p><b>HISTORY OF ALCOHOL CONSUMPTION</b></p> <ul style="list-style-type: none"> <li>• Ever versus never drinker</li> <li>• Among ever drinkers, age (years) when individual started and stopped drinking, the amount of alcohol consumed (i.e., the number of drinks per month), duration (i.e., the total number of years that an individual drank)</li> </ul>	

Table 8 (continued): Pearson Product Moment and Spearman Rank Correlations of Prospective and Retrospective Reports of Exposure for Variables and Conditions Possibly Related to Breast Cancer Risk - A Summary by Magnitude

<p>(0.00 - 0.49) LITTLE/NO RELATIONSHIP LOW AGREEMENT/CORRELATION</p>	<p>(0.50 - 0.69) MODERATE AGREEMENT/CORRELATION</p>	<p>(0.70 - 1.00) HIGH LEVEL OF AGREEMENT TO PERFECT CORRELATION</p>
	<p><b>HORMONE USE</b></p> <p>(2) Postmenopausal Estrogen Replacement Therapy (ERT)</p> <ul style="list-style-type: none"> <li>• Among ever users, age when ERT stopped</li> </ul>	<p><b>HORMONE USE</b></p> <p>(1) Oral Contraceptives</p> <ul style="list-style-type: none"> <li>• Ever versus never use</li> <li>• Among ever users, age (years) when started and stopped therapeutic use</li> <li>• Duration (years) that oral contraceptives were used</li> </ul> <p>(2) Postmenopausal Estrogen Replacement Therapy (ERT)</p> <ul style="list-style-type: none"> <li>• Ever versus never used</li> <li>• Among ever users, age (years) when woman commenced ERT</li> <li>• Duration (years) that ERT was used</li> </ul>

Table 8 (continued): Pearson Product Moment and Spearman Rank Correlations of Prospective and Retrospective Reports of Exposure for Variables and Conditions Possibly Related to Breast Cancer Risk - A Summary by Magnitude

<p>(0.00 - 0.49) LITTLE/NO RELATIONSHIP LOW AGREEMENT/CORRELATION</p>	<p>(0.50 - 0.69) MODERATE AGREEMENT/CORRELATION</p>	<p>(0.70 - 1.00) HIGH LEVEL OF AGREEMENT TO PERFECT CORRELATION</p>
<p><b>SOMATIC CHANGES POSSIBLY RELATED TO OVULATORY ACTIVITY</b></p> <ul style="list-style-type: none"> <li>• Among women who had experienced breast pain and tenderness, the reporting of the associated symptoms (<u>control groups one and two</u>)</li> <li>• Reported knowledge that period was coming (i.e., body changes preceded menses)</li> </ul>	<p><b>SOMATIC CHANGES POSSIBLY RELATED TO OVULATORY ACTIVITY</b></p> <ul style="list-style-type: none"> <li>• History of breast pain and tenderness (excluding during pregnancy) (Yes versus No)</li> <li>• Among women who had experienced breast pain and tenderness, the reporting of the associated symptoms (<u>cases only</u>)</li> <li>• Breast pain and tenderness related to periods (Yes versus No)</li> <li>• Period-related breast swelling (Yes versus No)</li> <li>• Among women who reported a history of breast swelling, age (years) when breast swelling started and stopped</li> </ul>	
<p><b>HEALTH PRACTICES AND PROCEDURES</b></p> <ul style="list-style-type: none"> <li>• Mammogram history - reported frequency and date of last mammogram</li> </ul>	<p><b>HEALTH PRACTICES AND PROCEDURES</b></p> <ul style="list-style-type: none"> <li>• The reported practice of regular breast self-examination (BSE)</li> <li>• Frequency of BSE (yearly)</li> <li>• Breast needle aspiration (Yes versus No)</li> <li>• Age at first needle breast aspiration and frequency - (<u>cases only</u>)</li> <li>• Mammogram history - frequency of previous mammograms</li> </ul>	<p><b>HEALTH PRACTICES AND PROCEDURES</b></p> <ul style="list-style-type: none"> <li>• Breast needle aspiration (Yes versus No); and</li> <li>• Age at first needle aspiration - (<u>control groups one and two</u>)</li> </ul>

Table 8 (continued): Pearson Product Moment and Spearman Rank Correlations of Prospective and Retrospective Reports of Exposure for Variables and Conditions Possibly Related to Breast Cancer Risk - A Summary by Magnitude

<p>(0.00 - 0.49) LITTLE/NO RELATIONSHIP LOW AGREEMENT/CORRELATION</p>	<p>(0.50 - 0.69) MODERATE AGREEMENT/CORRELATION</p>	<p>(0.70 - 1.00) HIGH LEVEL OF AGREEMENT TO PERFECT CORRELATION</p>
	<p>Quetelet's body to mass index (kg/m<sup>2</sup>)</p>	

Table 9: The Agreement Between Prospective and Retrospective Reports of Exposure as Measured by Kappa - A Summary by Magnitude.

(0.00 < k < 0.40) POOR TO MARGINAL REPRODUCIBILITY	(0.40 ≤ k ≤ 0.75) MODERATE TO SUBSTANTIAL REPRODUCIBILITY	(0.75 < k ≤ 1.0) EXCELLENT TO ALMOST PERFECT REPRODUCIBILITY
		<b>DEMOGRAPHIC VARIABLES</b> <ul style="list-style-type: none"> <li>• Age</li> <li>• Ethnicity</li> <li>• Marital Status</li> <li>• Education (i.e., highest level of schooling achieved)</li> </ul>
	<b>MEDICAL HISTORY</b> <ul style="list-style-type: none"> <li>• Age at first menstrual period</li> <li>• Among postmenopausal women, age (years) when periods stopped (<u>control group two</u>)</li> <li>• Type of menopause (i.e., natural vs hysterectomy) - (<u>control group two</u>)</li> </ul>	<b>MEDICAL HISTORY</b> <ul style="list-style-type: none"> <li>• Menopausal status (i.e., pre- or postmenopausal)</li> <li>• Among postmenopausal women, age (years) when periods stopped</li> <li>• Type of menopause (<u>cases/control group one</u>)</li> </ul>
		<b>REPRODUCTIVE HISTORY</b> <ul style="list-style-type: none"> <li>• Pregnancy history (i.e., ever versus never)</li> <li>• Parity (i.e., number of live full-term births)</li> <li>• Age at first full-term pregnancy</li> </ul>

Note: Kappa values are similar for all study groups unless otherwise specified.

References: 1. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics*, 1977; 33:159-174.  
2. Fleiss, JL. *Statistical Methods for Rates and Proportions* (Second Edition). New York: John Wiley and Sons, 1981, p. 218.

Table 9 (continued): The Agreement Between Prospective and Retrospective Reports of Exposure as Measured by Kappa - A Summay by Magnitude

(0.00 < k < 0.40) POOR TO MARGINAL REPRODUCIBILITY	(0.40 ≤ k ≤ 0.75) MODERATE TO SUBSTANTIAL REPRODUCIBILITY	(0.75 < k ≤ 1.0) EXCELLENT TO ALMOST PERFECT REPRODUCIBILITY
		<p><b>FAMILY HISTORY OF BREAST CANCER</b></p> <ul style="list-style-type: none"> <li>• Family history of breast cancer in any first degree relatives (i.e., mother, sister(s))</li> <li>• Type and characteristics of the breast cancer in the first degree relatives (i.e., premenopausal versus postmenopausal, and/or unilateral versus bilateral)</li> </ul>
		<p><b>SMOKING HISTORY</b></p> <ul style="list-style-type: none"> <li>• Ever versus never smoked</li> <li>• Frequency of use (i.e., specific number of cigarettes smoked per day)</li> <li>• Duration of smoking - (i.e., total number of years smoked)</li> </ul>
	<p><b>ALCOHOL CONSUMPTION</b></p> <ul style="list-style-type: none"> <li>• Ever versus never drinker</li> <li>• Among ever drinkers, age when individual started and stopped drinking</li> <li>• Amount consumed (i.e., number of drinks per month) - (<u>control group two</u>)</li> <li>• Duration (total number of years) - (<u>cases and control group one</u>)</li> </ul>	<p><b>ALCOHOL CONSUMPTION</b></p> <ul style="list-style-type: none"> <li>• Among ever drinkers, amount (i.e., number of drinks per month) consumed - (<u>cases and control group one</u>)</li> <li>• Duration (i.e., total number of years) - (<u>control group two</u>)</li> </ul>

Table 9 (continued): The Agreement Between Prospective and Retrospective Reports of Exposure as Measured by Kappa - A Summay by Magnitude

$(0.00 < k < 0.40)$ <b>POOR TO MARGINAL REPRODUCIBILITY</b>	$(0.40 \leq k \leq 0.75)$ <b>MODERATE TO SUBSTANTIAL REPRODUCIBILITY</b>	$(0.75 < k \leq 1.0)$ <b>EXCELLENT TO ALMOST PERFECT REPRODUCIBILITY</b>
	<b>HORMONE USE</b> <b>(1) Oral Contraceptives (OCs)</b> <ul style="list-style-type: none"> <li>• Duration of use (years)</li> <li>• Among ever users, age when OCs started and stopped</li> </ul> <b>(2) Postmenopausal Estrogen Replacement Therapy (ERT)</b> <ul style="list-style-type: none"> <li>• Among ever users, age (years) when ERT started and stopped</li> <li>• Duration of use (i.e., total number of years) - <u>(cases only)</u></li> </ul>	<b>HORMONE USE</b> <b>(1) Oral Contraceptives</b> <ul style="list-style-type: none"> <li>• Ever versus never use</li> </ul> <b>(2) Postmenopausal Estrogen Replacement Therapy (ERT)</b> <ul style="list-style-type: none"> <li>• Ever versus never use</li> <li>• Duration of use (i.e., total number of years) - <u>(control groups one and two)</u></li> </ul>
<b>SOMATIC CHANGES POSSIBLY RELATED TO OVULATORY ACTIVITY</b>  <ul style="list-style-type: none"> <li>• Reported knowledge that period was coming: body changes preceded menses <u>(case only)</u></li> </ul>	<b>SOMATIC CHANGES POSSIBLY RELATED TO OVULATORY ACTIVITY</b> <ul style="list-style-type: none"> <li>• History of breast pain and tenderness (excluding during pregnancy) (Yes versus No)</li> <li>• Symptoms reported among those who experienced breast pain and tenderness</li> <li>• Breast pain and tenderness related to periods (Yes versus No)</li> <li>• Reported knowledge that period was coming: body changes preceded menses - <u>(control groups one and two)</u></li> <li>• Breast swelling period-related (Yes versus No)</li> <li>• Among women who reported a history of breast swelling, age (years) when swelling started and stopped</li> </ul>	

Table 9 (continued): The Agreement Between Prospective and Retrospective Reports of Exposure as Measured by Kappa - A Summay by Magnitude

<p>(0.00 &lt; k &lt; 0.40) POOR TO MARGINAL REPRODUCIBILITY</p>	<p>(0.40 ≤ k ≤ 0.75) MODERATE TO SUBSTANTIAL REPRODUCIBILITY</p>	<p>(0.75 &lt; k ≤ 1.0) EXCELLENT TO ALMOST PERFECT REPRODUCIBILITY</p>
<p><b>HEALTH PRACTICES AND DIAGNOSTIC PROCEDURES</b></p> <ul style="list-style-type: none"> <li>• Mammogram history - reported frequency (i.e., the total number of previous mammograms) and the date of the last mammogram</li> <li>• Reported history of having a breast needle aspiration, as well as frequency (total number) and age at first aspiration (years) - (<u>cases only</u>)</li> </ul>	<p><b>HEALTH PRACTICES AND DIAGNOSTIC PROCEDURES</b></p> <ul style="list-style-type: none"> <li>• The reported practice of regular breast self-examination (BSE)</li> <li>• Frequency of BSE (yearly)</li> <li>• Breast needle aspiration (Yes versus No), frequency, and age at first aspiration - (<u>control group one</u>)</li> </ul>	
		<p><b>ANTHROPOMETRIC VARIABLE</b></p> <ul style="list-style-type: none"> <li>• Quetelet's body to mass index [kg/m<sup>2</sup>]</li> </ul>

Table 10.1: A Comparison of Prospective and Retrospective Reports of Exposure by Breast Cancer Cases: Level of Agreement Between Q1 and Q2 Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
1. Socioeconomic Status	High <sup>1</sup>	13	4	4.4	95.6	1.00	.719	No
	Low <sup>2</sup>	5	182					
2. Age at Menses	≤ 12 yrs <sup>1</sup>	67	5	9.9	90.1	.42	.801	No
	≥ 13 yrs <sup>2</sup>	9	60					
3. Age at Menopause	≤ 45 yrs <sup>1</sup>	52	6	8.9	91.1	.79	.811	No
	46-55 yrs <sup>2</sup>	8	92					
4. Total number of years of ovulatory activity	> 40 yrs <sup>1</sup>	43	11	11.2	88.8	.06	.768	No
	34-40 yrs <sup>2</sup>	3	68					
5. Oral contraceptive use (Ever vs Never)	No <sup>1</sup>	83	3	5.8	94.2	.09	.879	No
	Yes <sup>2</sup>	10	127					

Notes:

1. Questionnaire One (Q1) reports were used as the 'gold standard' or 'criterion' for the comparison of concordance between Q1 and Q2 exposure reports. Q1 reports were found to be unaffected by the knowledge of diagnosis.
2. (\*) Denotes a significant McNemar's Test - (p < 0.05).

Table 10.1 (continued): A Comparison of Prospective and Retrospective Reports of Exposure by Breast Cancer Cases: Level of Agreement Between Q1 and Q2 Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
6. Among ever users, the duration (years) that oral contraceptives were used	Never <sup>1</sup>	127	7	5.6	94.4	.34	.854	No
	1-5 yrs <sup>2</sup>	3	41					
7. Postmenopausal Estrogen Replacement (ERT)	Never <sup>1</sup>	134	11	11.6	88.4	.69	.734	No
	Ever <sup>2</sup>	14	57					
8. Among ever users duration (years) of Estrogen Replacement Therapy (ERT)	Never <sup>1</sup>	134	10	11.0	89.0	1.00	.660	No
	1-5 yrs <sup>2</sup>	10	27					
9. Pregnancy history	Ever <sup>1</sup>	194	4	6.1	93.9	1.00	.830	No
	Nullip <sup>2</sup>	5	24					
10. Quetelet's Index: body to mass index (kg/m <sup>2</sup> )	Low <sup>1</sup>	20	4	5.2	94.8	1.00	.802	No
	High <sup>2</sup>	4	24					
	High <sup>1</sup>	125	22	17.4	82.6	.08	.545	No
	Normal <sup>2</sup>	11	32					

Table 10.1 (continued): A Comparison of Prospective and Retrospective Reports of Exposure by Breast Cancer Cases: Level of Agreement Between Q1 and Q2 Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
11. History of breast cancer in a first degree relative	No <sup>1</sup>	123	5	3.8	96.2	.22	.880	No
	Yes <sup>2</sup>	1	28					
12. Smoking history	Never <sup>1</sup>	70	8	4.5	95.5	.11	.899	No
	Ever <sup>2</sup>	2	142					
13. Average number of cigarettes smoked per day	Never <sup>1</sup>	73	6	5.5	94.5	.13	.884	No
	Ever, 1-10 <sup>2</sup>	1	48					
14. Duration (years) of smoking among ever smokers	Never <sup>1</sup>	73	6	6.1	93.9	.03* Note 2	.825	No
	Ever, 1-10 <sup>2</sup>	0	19					
15. Alcohol Consumption	Never <sup>1</sup>	143	12	13.3	86.7	.36	.675	No
	Ever <sup>2</sup>	18	52					
16. Amount of alcohol consumed (drinks/month)	Never <sup>1</sup>	50	13	17.9	82.1	.38	.641	No
	Ever, 1-10 <sup>2</sup>	8	46					

Table 10.1 (continued): A Comparison of Prospective and Retrospective Reports of Exposure by Breast Cancer Cases: Level of Agreement Between Q1 and Q2 Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
17. Among ever drinkers, the duration (years) of alcohol consumption	Never <sup>1</sup>	51	3	7.0	93.0	.63	.465	No
	Ever, 1-10 <sup>2</sup>	1	2					
	Never <sup>1</sup>	51	12	14.9	85.1	.66	.697	No
	Ever, > 30 <sup>2</sup>	9	69					

Table 10.2: A Comparison of Prospective and Retrospective Reports of Exposure by Controls (Group One): Level of Agreement Between Q<sub>1</sub> and Q<sub>2</sub> Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
1. Socioeconomic Status	High <sup>1</sup>	39	2	3.4	96.6	(.007) *Note 2	.820	No
	Low <sup>2</sup>	13	378					
2. Age at Menses	≤ 12 yrs <sup>1</sup>	139	20	14.1	85.9	1.0	.716	No
	≥ 13 yrs <sup>2</sup>	21	111					
3. Age at Menopause	≤ 45 yrs <sup>1</sup>	114	16	6.9	93.1	.03*	.857	No
	46-55 yrs <sup>2</sup>	5	169					
4. Total number of years of ovulatory activity	> 40 yrs <sup>1</sup>	97	26	14.0	86.0	.01*	.719	No
	34-40 yrs <sup>2</sup>	10	125					
5. Oral contraceptive use (Ever vs Never)	No <sup>1</sup>	215	6	5.9	94.1	.007*	.881	No
	Yes <sup>2</sup>	21	213					

Notes:

1. Questionnaire One (Q1) reports were used as the 'gold standard' or 'criterion' for the comparison of concordance between Q1 and Q2 exposure reports. Q1 reports were considered to be unaffected by the knowledge of diagnosis.
2. (\*) Denotes a significant McNemar's Test - ( $p < 0.05$ ).

Table 10.2 (continued): A Comparison of Prospective and Retrospective Reports of Exposure by Controls (Group One): Level of Agreement Between Q<sub>1</sub> and Q<sub>2</sub> Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
6. Among ever users duration (years) that oral contraceptives were used	Never <sup>1</sup>	213	20	6.0	94.0	.002*	.838	No
	1-5 yrs <sup>2</sup>	4	97					
7. Postmenopausal Estrogen Replacement (ERT)	Never <sup>1</sup>	283	24	6.9	93.1	.002*	.838	No
	Ever <sup>2</sup>	6	119					
8. Among ever users, the duration (years) of Estrogen Replacement Therapy (ERT)	Never <sup>1</sup>	283	18	6.0	94.0	.004*	.808	No
	1-5 yrs <sup>2</sup>	4	60					
9. Pregnancy history	Ever <sup>1</sup>	396	2	3.3	96.7	.007*	.851	No
	Nullip <sup>2</sup>	13	50					

Notes:

1. Questionnaire One (Q1) reports were used as the 'gold standard' or 'criterion' for the comparison of concordance between Q1 and Q2 exposure reports. Q1 reports were considered to be unaffected by the knowledge of diagnosis.
2. (\*) Denotes a significant McNemar's Test - ( $p < 0.05$ ).

Table 10.2 (continued): A Comparison of Prospective and Retrospective Reports of Exposure by Controls (Group One): Level of Agreement Between Q<sub>1</sub> and Q<sub>2</sub> Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
10. Quetelet's Index: Body to mass index (kg/m <sup>2</sup> )	Low <sup>1</sup>	64	4	6.8	93.2	.02*	.825	No
	High <sup>2</sup>	15	195					
	High <sup>1</sup>	195	35	15.0	85.0	.03*	.680	No
	Normal <sup>2</sup>	18	105					
11. History of breast cancer in a first degree relative	No <sup>1</sup>	288	11	3.8	96.2	.02*	.835	No
	Yes <sup>2</sup>	2	39					
12. Smoking history	Never <sup>1</sup>	175	10	3.1	96.9	.18	.936	No
	Ever <sup>2</sup>	4	268					
13. Average number of cigarettes smoked per day	Never <sup>1</sup>	175	8	3.7	96.3	.11	.917	No
	Ever, 1-10 <sup>2</sup>	2	85					
14. Duration (years) of smoking among ever smokers	Never <sup>1</sup>	175	4	3.2	96.8	1.0	.892	No
	Ever, 1-10 <sup>2</sup>	3	36					

Table 10.2 (continued): A Comparison of Prospective and Retrospective Reports of Exposure by Controls (Group One): Level of Agreement Between Q<sub>1</sub> and Q<sub>2</sub> Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
15. Alcohol Consumption	Never <sup>1</sup>	259	24	16.5	83.5	.003*	.638	No
	Ever <sup>2</sup>	51	121					
16. Amount of alcohol consumed (drinks/month)	Never <sup>1</sup>	121	39	19.8	80.2	.008*	.806	No
	Ever, 1-10 <sup>2</sup>	18	110					
17. Among ever drinkers, the duration of Alcohol Consumption (years)	Never <sup>1</sup>	120	5	3.9	96.1	.06	.529	No
	Ever, 1-10 <sup>2</sup>	0	3					
	Never <sup>1</sup>	120	33	14.6	85.4	.001*	.709	No
	Ever, > 30 <sup>2</sup>	10	131					

Table 10.3: A Comparison of Prospective and Retrospective Reports of Exposure by Controls (Group Two): Level of Agreement Between Q<sub>1</sub> and Q<sub>2</sub> Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
1. Socioeconomic Status	High <sup>1</sup>	34	1	2.4	97.6	.022* Note 2	.859	No
	Low <sup>2</sup>	9	378					
2. Age at Menses	≤ 12 yrs <sup>1</sup>	138	13	14.7	85.3	.015*	.705	No
	≥ 13 yrs <sup>2</sup>	30	112					
3. Age at Menopause	≤ 45 yrs <sup>1</sup>	123	12	8.1	91.9	1.00	.836	No
	46-55 yrs <sup>2</sup>	13	162					
4. Total number of years of ovulatory activity	> 40 yrs <sup>1</sup>	23	100	12.9	87.1	.02*	.743	No
	34-40 yrs <sup>2</sup>	117	9					
5. Oral contraceptive use (Ever vs Never)	No <sup>1</sup>	196	8	4.1	95.9	.65	.916	No
	Yes <sup>2</sup>	11	245					

Notes:

1. Questionnaire One (Q1) reports were used as the 'gold standard' or 'criterion' for the comparison of concordance between Q<sub>1</sub> and Q<sub>2</sub> exposure reports. Q1 reports were considered to be unaffected by the knowledge of diagnosis.
2. (\*) Denotes a significant McNemar's Test - (p < 0.05).

Table 10.3 (continued): A Comparison of Prospective and Retrospective Reports of Exposure by Controls (Group Two): Level of Agreement Between Q<sub>1</sub> and Q<sub>2</sub> Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively (Q <sub>1</sub> ) Note 1	Reported Retrospectively (Q <sub>2</sub> )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
6. Among ever users, the duration (years) that oral contraceptives were used	Never <sup>1</sup>	246	7	3.7	96.3	1.00	.853	No
	1-5 yrs <sup>2</sup>	6	96					
7. Postmenopausal Estrogen Replacement (ERT)	Never <sup>1</sup>	243	27	8.0	92.0	.002*	.835	No
	Ever <sup>2</sup>	8	160					
8. Among ever users, the duration (years) of Estrogen Replacement Therapy (ERT)	Never <sup>1</sup>	243	24	9.0	91.0	.004*	.757	No
	1-5 yrs <sup>2</sup>	7	69					
9. Pregnancy history	Ever <sup>1</sup>	362	3	10.6	89.4	.000*	.650	No
	Nullip <sup>2</sup>	43	55					
10. Quetelet's Index: Body to mass index (kg/m <sup>2</sup> )	Low <sup>1</sup>	57	5	9.6	90.4	.004*	.756	No
	High <sup>2</sup>	20	178					
	High <sup>1</sup>	178	50	17.2	82.8	.000*	.650	No
	Normal <sup>2</sup>	11	116					

Table 10.3 (continued): A Comparison of Prospective and Retrospective Reports of Exposure by Controls (Group Two): Level of Agreement Between  $Q_1$  and  $Q_2$  Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively ( $Q_1$ ) Note 1	Reported Retrospectively ( $Q_2$ )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
11. History of breast cancer in a first degree relative	No <sup>1</sup>	277	5	2.4	97.6	.73	.916	No
	Yes <sup>2</sup>	3	53					
12. Smoking history	Never <sup>1</sup>	221	14	3.7	96.3	0.01*	.926	No
	Ever <sup>2</sup>	3	224					
13. Average number of cigarettes smoked per day	Never <sup>1</sup>	221	13	4.6	95.4	.007*	.892	No
	Ever, 1-10 <sup>2</sup>	2	92					
14. Duration (years) of smoking among ever smokers	Never <sup>1</sup>	221	10	4.2	95.8	.02*	.807	No
	Ever, 1-10 <sup>2</sup>	1	27					
15. Alcohol Consumption	Never <sup>1</sup>	284	22	12.8	87.2	.07	.708	No
	Ever <sup>2</sup>	37	119					
16. Amount of Alcohol Consumed (drinks/month)	Never <sup>1</sup>	120	29	16.9	83.1	.32	.662	No
	Ever, 1-10 <sup>2</sup>	21	126					

Table 10.3 (continued): A Comparison of Prospective and Retrospective Reports of Exposure by Controls (Group Two): Level of Agreement Between  $Q_1$  and  $Q_2$  Reports (Kappa) and Directional Discordance (McNemar's Test)

Variable	Reported Prospectively ( $Q_1$ ) Note 1	Reported Retrospectively ( $Q_2$ )		Percent Discordance (%)	Percent Concordance (%)	McNemar's Test (p-value)	Kappa	Changes in Exposure Reports Will Affect Odds Ratio Estimate
		1	2					
17. Among ever drinkers, the duration (years) of alcohol consumption	Never <sup>1</sup>	120	0	0.00	100	1.0	.100	No
	Ever, 1-10 <sup>2</sup>	0	6					
	Never <sup>1</sup>	120	18	10.8	89.2	.86	.780	No
	Ever, > 30 <sup>2</sup>	16	160					

Table 11: Exposure - Disease Odds Ratio Estimates for Prospective (Pre-Diagnostic) and Retrospective (Post-Diagnostic) Reports of Exposure for Study Factors Possibly Related to Breast Cancer Development

EXPOSURE VARIABLE ANTECEDENT EVENT/CONDITION	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR
	Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)		Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)	
DEMOGRAPHIC VARIABLES						
a. <b>Marital Status</b> Ever versus Never married	1.17 (.55-2.49)	1.28 (.53-3.10)	1.09	1.15 (.54-2.45)	1.50 (.63-3.58)	1.30
b. <b>Socioeconomic Status</b> • Lower • Middle • Upper	*Note 1 .62 (.31-1.25) .55 (.24-1.33)	* .72 (.37-1.40) .52 (.23-1.18)	1.16 .95	* .92 (.48-1.75) .99 (.43-2.31)	* .97 (.52-1.81) .86 (.38-1.92)	1.89 .87
c. <b>Education</b> • ≤ 8 years • 9-12 years • > 12 years (Post- secondary, including vocational/trade/ college/university	* .61 (.32-1.15) .71 (.37-1.34)	.68 (.35-1.33) .68 (.35-1.33)	1.11 .96	* .68 (.36-1.26) .78 (.42-1.47)	.95 (.50-1.79) .85 (.45-1.60)	1.40 1.09
d. <b>Location of Residence</b> • Urban vs Rural	1.15 (.78-1.69)	-	-	1.01 (.57-1.24)	-	-

Note: 1. (\*) Denotes reference category for the calculation of odds ratios.

Table 11 (continued): Exposure - Disease Odds Ratio Estimates for Prospective (Pre-Diagnostic) and Retrospective (Post-Diagnostic) Reports of Exposure for Study Factors Possibly Related to Breast Cancer Development

EXPOSURE VARIABLE ANTECEDENT EVENT/CONDITION	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR
	Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)		Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)	
MEDICAL HISTORY FACTORS						
a.    Age at menarche • ≤ 12 years • 13 years • ≥ 14 years	*  1.15 (.78-1.69) 1.08 (.73-1.6)	*  1.05 (.72-1.54) 1.06 (.72-1.57)	  .91 .98	*  1.08 (.70-1.52) 1.05 (.70-1.52)	*  1.21 (.77-1.65) 1.13 (.77-1.65)	  1.12 .93
b.    Age at menopause • ≤ 45 years • 46-55 years • > 55 years	*  1.09 (.75-1.58) 1.42 (.63-3.24)	*  1.01 (.69-1.48) 2.00 (.85-4.70)	  .93 1.41	*  1.19 (.82-1.73) 1.17 (.53-2.58)	*  1.24 (.86-1.81) 1.99 (.87-4.55)	  1.04 1.70
c.    Total Lifetime Ovulatory Activity • ≤ 33 years • 34-40 years • > 40 years	*  1.18 (.81-1.73) 1.01 (.68-1.54)	*  .91 (.62-1.34) .82 (.54-1.28)	  .77 .81	*  1.30 (.89-1.91) 1.03 (.68-1.56)	*  1.06 (.73-1.55) .87 (.57-1.35)	  .82 .84
d.    Hysterectomy • No versus Yes	1.68 (1.11-2.56)	1.51 (.99-2.29)	.90	1.84 (1.20-2.80)	1.56 (1.03-2.36)	.84
e.    Bilateral oophorectomy • No versus Yes	1.80 (1.19-2.72)	1.58 (1.05-2.37)	.88	1.96 (1.30-2.96)	1.64 (1.09-2.46)	.84

Table 11 (continued): Exposure - Disease Odds Ratio Estimates for Prospective (Pre-Diagnostic) and Retrospective (Post-Diagnostic) Reports of Exposure for Study Factors Possibly Related to Breast Cancer Development

EXPOSURE VARIABLE ANTECEDENT EVENT/CONDITION	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR
	Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)		Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)	
SOMATIC CHANGES POSSIBLY RELATED TO OVULATORY ACTIVITY (i.e., menses)						
a.    Breast swelling period-related • Yes versus No	1.43 (1.04-1.99)	1.20 (.87-1.65)	.84	1.10 (.93-1.29)	1.02 (.87-1.18)	.93
b.    Breast pain and tenderness period-related • Yes versus No	.69 (.50-.96)	.52 (.38-.73)	.75	1.06 (.77-1.47)	.72 (.52-.99)	.68
REPRODUCTIVE HISTORY						
a.    Ever pregnant • Yes versus No	.96 (.61-1.52)	1.19 (.74-1.93)	1.24	.59 (.37-.90)	1.05 (.65-1.68)	1.78
b.    Age at first full-term pregnancy • ≤ 20 years	*	*	*	*	*	1.00
• 21-24 years	1.17 (.71-1.93)	.99 (.60-1.66)	.85	.98 (.58-1.66)	.98 (.58-1.65)	1.00
• 25-29 years	1.36 (.80-2.29)	1.36 (.80-2.31)	1.00	.87 (.51-1.48)	.87 (.51-1.47)	1.55
• ≥ 30 years	1.35 (.82-2.24)	1.37 (.82-2.29)	1.01	.82 (.49-1.38)	1.27 (.76-2.14)	.98
• Nulliparous	1.92 (.41-9.07)	2.35 (.14-3.89)	1.22	1.11(.73-2.70)	1.09 (.83-3.20)	

Table 11 (continued): Exposure - Disease Odds Ratio Estimates for Prospective (Pre-Diagnostic) and Retrospective (Post-Diagnostic) Reports of Exposure for Study Factors Possibly Related to Breast Cancer Development

EXPOSURE VARIABLE ANTECEDENT EVENT/CONDITION	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR
	Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)		Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)	
REPRODUCTIVE HISTORY (continued)						
c. Parity (i.e., number of live births) • 1-2 • 3-4 • ≥ 5 • None	* .86 (.60-1.24) .61 (.30-1.21) 1.09 (.74-1.75)	* .84 (.58-1.21) .86 (.46-1.61) .96 (.59-1.63)	.98 1.41 .88	* .84 (.58-1.21) .56 (.28-1.12) 1.12 (.69-3.40)	* .74 (.52-1.07) .78 (.42-1.44) 1.04 (.86-2.40)	.88 1.39 .93
USE OF EXOGENOUS HORMONES						
a. Oral contraceptives • Yes vs No	.66 (.48-.91)	.65 (.47-.90)	.98	.80 (.58-1.10)	.85 (.62-1.18)	1.06
b. Duration (years) • 0 years • 1-5 years • > 5 years	* .75 (.51-1.09) .61 (.39-.96)	* .65 (.44-.94) .75 (.47-1.18)	.87 1.23	* .83 (.57-1.22) .79 (.50-1.25)	* .85 (.58-1.23) 1.04 (.65-1.65)	1.02 1.32
c. Postmenopausal Estrogen Replacement Therapy (ERT) • Ever vs Never	1.18 (.84-1.67)	.83 (.59-1.18)	.70	.80 (.57-1.12)	.59 (.42-.82)	.74

Table 11 (continued): Exposure - Disease Odds Ratio Estimates for Prospective (Pre-Diagnostic) and Retrospective (Post-Diagnostic) Reports of Exposure for Study Factors Possibly Related to Breast Cancer Development

EXPOSURE VARIABLE ANTECEDENT EVENT/CONDITION	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR
	Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)		Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)	
USE OF EXOGENOUS HORMONES (continued)						
d.   Duration (years) ERT • 0 years • 1-5 years • 6-10 years • > 10 years	*  1.13 (.74-1.73) 1.23 (.59-2.57) 1.26 (.71-2.22)	*  .91 (.60-1.38) .48 (.19-1.19) .89 (.50-1.56)	  .81 .39 1.42	*  .78 (.51-1.17) .76 (.38-1.52) .88 (.51-1.54)	*  .68 (.45-1.03) .25 (.10-.59) .66 (.38-1.15)	  .87 .33 .75
QUETELET'S INDEX (kg/m <sup>2</sup> ) • low (≤ 21) • normal (22-24) • high (≥ 25)	*  .90 (.53-1.52) 1.64 (1.05-2.58)	*  1.38 (.82-2.32) 2.24 (1.39-3.61)	  1.53 1.37	*  .79 (.46-1.33) 1.47 (1.01-2.32)	*  1.14 (.68-1.92) 2.42 (1.50-3.92)	  1.44 1.65
FAMILY HISTORY OF BREAST CANCER						
a.   Positive history of breast cancer in a first degree relative (i.e., mother/sister/both) • Yes versus No	  1.65 (.99-2.75)	  1.58 (.99-2.52)	  .96	  1.10 (.68-1.78)	  1.35 (.85-2.14)	  <i>1.23</i>
b.   Breast cancer in mother • Yes versus No	  1.47 (.83-2.60)	  1.51 (.90-2.55)	  1.03	  1.36 (.78-2.39)	  1.64 (.97-2.78)	  <i>1.21</i>

Table 11 (continued): Exposure - Disease Odds Ratio Estimates for Prospective (Pre-Diagnostic) and Retrospective (Post-Diagnostic) Reports of Exposure for Study Factors Possibly Related to Breast Cancer Development

EXPOSURE VARIABLE ANTECEDENT EVENT/CONDITION	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR
	Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)		Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)	
FAMILY HISTORY OF BREAST CANCER (continued)						
c. <b>Breast cancer in sister(s)</b> • Yes versus No	1.58 (.82-3.03)	1.48 (.81-2.70)	.94	.94 (.51-1.72)	1.01 (.57-1.78)	1.07
d. <b>Family history of breast cancer in first degree relatives</b> • No • Yes, lower risk (unilateral and postmenopausal) • Yes, higher risk (bilateral or premenopausal)	* 1.81 (.91-3.59)  .72 (.07-6.92)	* 1.32 (.65-2.68)  1.06 (.10-11.80)	.73  1.47	* 1.69 (.86-3.33)  .53 (.06-4.76)	* 1.38 (.67-2.83)  .35 (.04-2.96)	.82  .66
SMOKING HISTORY						
a. <b>Smoking status</b> • Ever smoker vs Never	1.21 (.87-1.68)	1.34 (.96-1.87)	1.11	1.83 (1.32-2.54)	1.91 (1.37-2.66)	1.04
b. <b>Cigarettes smoked per day</b> • None (never smoked) • 1-10 • 11-19 • 20-29 • ≥ 30	* 1.26 (.83-1.91) 1.22 (.73-2.05) 1.15 (.73-1.81) 1.30 (.67-2.52)	* 1.40 (.93-2.10) 1.62 (.93-2.82) .97 (.62-1.53) 1.77 (.90-3.45)	1.11 1.33 .84 1.36	* 1.51 (1.10-2.28) 2.82 (1.59-4.98) 1.73 (1.09-2.75) 2.74 (1.32-5.68)	* 1.71 (1.15-2.56) 2.56 (1.43-4.55) 1.48 (.94-2.35) 2.22 (1.58-6.98)	1.13 .91 .86 1.22

Table 11 (continued): Exposure - Disease Odds Ratio Estimates for Prospective (Pre-Diagnostic) and Retrospective (Post-Diagnostic) Reports of Exposure for Study Factors Possibly Related to Breast Cancer Development

EXPOSURE VARIABLE ANTECEDENT EVENT/CONDITION	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR
	Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)		Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)	
SMOKING HISTORY (continued)						
c.   Duration of smoking (years)	*	*		*	*	
• None (never smoked)	1.35 (.79-2.32)	1.62 (.93-2.8)	1.20	2.01 (1.16-3.50)	1.85 (1.08-3.17)	.92
• 1-10	1.64 (.92-2.90)	2.15 (1.15-3.99)	1.31	2.02 (1.15-3.58)	2.53 (1.37-4.68)	1.25
• 11-19	1.16 (.71-1.91)	.89 (.51-1.53)	.77	1.94 (1.16-3.24)	1.34 (.77-2.34)	.69
• 20-29	1.04 (.70-1.56)	1.14 (.76-1.70)	1.10	1.62 (1.08-2.40)	1.87 (1.24-2.80)	1.15
• ≥ 30						
ALCOHOL USE HISTORY						
a.   Alcohol consumption						
• Ever versus Never	1.37 (.97-1.92)	1.14 (.81-1.61)	.83	1.17 (.83-1.64)	1.08 (.77-1.54)	.92
b.   Number of drinks consumed per month						
• None (non-drinker)	*	*		*	*	
• 1-10	1.19 (.81-1.75)	.92 (.62-1.37)	.77	.97 (.66-1.43)	.92 (.62-1.38)	.95
• 11-20	1.67 (.93-3.00)	1.20 (.71-2.03)	.72	1.39 (.78-2.48)	1.39 (.81-2.38)	1.00
• 21-29	2.57 (.87-7.60)	.88 (.27-2.92)	.34	1.62 (.59-4.42)	.86 (.26-2.85)	.53
• ≥ 30	1.56 (.92-2.65)	1.86 (1.12-3.07)	1.19	1.26 (.75-2.12)	1.31 (.81-2.12)	1.04

Table 11 (continued): Exposure - Disease Odds Ratio Estimates for Prospective (Pre-Diagnostic) and Retrospective (Post-Diagnostic) Reports of Exposure for Study Factors Possibly Related to Breast Cancer Development

EXPOSURE VARIABLE ANTECEDENT EVENT/CONDITION	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR	Prospective Q <sub>1</sub>	Retrospective Q <sub>2</sub>	Ratio of Retrospective Odds Ratio (OR) to Prospective OR
	Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)		Odds Ratio (95% C.I.)	Odds Ratio (95% C.I.)	
ALCOHOL HISTORY (continued)						
c.   Duration of alcohol consumption (years)						
• None (non-drinker)	*	*		*	*	
• 1-10	1.27 (.42-3.84)	1.29 (.49-3.43)	1.02	1.63 (.50-5.30)	1.67 (.60-4.69)	1.02
• 11-20	1.56 (.79-3.09)	1.20 (.57-2.50)	.77	1.82 (.89-3.72)	1.27 (.60-2.68)	.70
• 21-29	1.23 (.73-2.06)	1.09 (.66-1.81)	.89	1.22 (.72-2.06)	1.11 (.67-1.85)	.91
• ≥ 30	1.39 (.95-2.02)	1.20 (.83-1.75)	.86	1.03 (.71-1.49)	1.11 (.77-1.62)	1.08

Table 12: A Comparison of the Prospective and Retrospective Exposure-Disease Odds Ratios: Study Factors for which the Odds Ratio Estimates Differed.

STUDY FACTOR	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Odds Ratio (95% C.I.)	Retrospective Odds Ratio (95% C.I.)	Direction of Change	Prospective Odds Ratio (95% C.I.)	Retrospective Odds Ratio (95% C.I.)	Direction of Change
1. Hysterectomy • No versus Yes	1.68 (1.11-2.56)	1.51 (.99-2.29)	Positive (+) association to no association	1.84 (1.2-2.8)	1.56 (1.03-2.36)	Both ORs support a positive (+) association
2. Bilateral oophorectomy • No versus Yes	1.80 (1.19 - 2.72)	1.58 (1.05-2.37)	Both ORs support a positive (+) association	1.96 (1.3-2.96)	1.64 (1.09-2.46)	Both ORs support a positive (+) association
3. Breast swelling which is reported to be related to menses • Yes vs No	1.43 (1.04-1.99)	1.20 (.87-1.65)	Positive (+) association to no association	1.10 (.93-1.29)	1.02 (.87-1.18)	No association
4. Breast pain and tenderness reported to be related to menses • Yes vs No	.69 (.50-.96)	.52 (.38-.73)	Both ORs support a negative (-) or protective association	1.06 (.77-1.47)	.72 (.52-.83)	No association to a negative (-) association
5. Pregnancy history • Yes versus Nulliparous	.96 (.61-1.52)	1.19 (.74-1.93)	No association	.59 (.37-.90)	1.05 (.65-1.68)	Negative (-) association to no association
6. Oral contraceptives • Yes versus No	.66 (.48-.91)	.65 (.47-.90)	Negative (-) association	.80 (.58-1.10)	.85 (.58-1.73)	No association

Table 12 (continued): A Comparison of the Prospective and Retrospective Exposure-Disease Odds Ratios: Study Factors for which the Odds Ratio Estimates Differed.

STUDY FACTOR	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Odds Ratio (95% C.I.)	Retrospective Odds Ratio (95% C.I.)	Direction of Change	Prospective Odds Ratio (95% C.I.)	Retrospective Odds Ratio (95% C.I.)	Direction of Change
7. Among ever users, duration of Oral Contraceptive use (years) • Never versus Ever, (1-5) years  • Never versus Ever, > 5 years	.75 (.51-1.09)  .61 (.39-.96)	.65 (.44-.94)  .75 (.47-1.18)	No association to a negative (-) association  Negative (-) association to no association	.83 (.57-1.22)  .79 (.50-1.25)	.85 (.58-1.23)  1.04 (.65-1.65)	No association  No association
8. Postmenopausal Estrogen Replacement Therapy (ERT) • Ever versus Never	1.18 (.84-1.67)	.83 (.59-1.18)	No association	.80 (.57-1.12)	.59 (.42-.82)	No association to a negative (-) association
9. Duration (years) of ERT among ever users • Never vs (6-10) years	1.23 (.59-2.57)	.48 (.19-1.19)	No association	.76 (.38-1.52)	.25 (.10-.59)	No association to a negative (-) association
10. Quetelet's Index • High versus low body mass index	1.64 (1.05-2.58)	2.24 (1.39-3.62)	Positive (+) association	1.47 (1.91-2.32)	2.42 (1.50-3.92)	Positive (+) association
11. Smoking history • Ever versus Never	1.21 (.87-1.68)	1.34 (.96-1.87)	No association	1.83 (1.32-2.54)	1.91 (1.37-2.66)	Positive (+) association

Table 12 (continued): A Comparison of the Prospective and Retrospective Exposure-Disease Odds Ratios: Study Factors for which the Odds Ratio Estimates Differed..

STUDY FACTOR	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Odds Ratio (95% C.I.)	Retrospective Odds Ratio (95% C.I.)	Direction of Change	Prospective Odds Ratio (95% C.I.)	Retrospective Odds Ratio (95% C.I.)	Direction of Change
12. Number of cigarettes smoked (on average) per day						
• Ever (1-10) vs Never	1.26 (.83-1.91)	1.40 (.93-2.10)	No association	2.82 (1.59-4.98)	2.56 (1.43-4.55)	Positive (+) association
• Ever (11-19) vs Never	1.22 (.73-2.05)	1.62 (.93-2.82)	No association	1.73 (1.09-2.75)	1.48 (.94-2.35)	Positive (+) to no association
• Ever (20-29) vs Never	1.15 (.73-1.81)	.97 (.62-1.53)	No association	2.74 (1.32-5.68)	3.33 (1.58-6.98)	Positive (+) association
• Ever ( $\geq 30$ vs Never)	1.30 (.67-2.52)	1.77 (.90-3.45)	No association	2.01 (1.16-3.5)	1.85 (1.08-3.17)	Positive (+) association

Table 12 (continued): A Comparison of the Prospective and Retrospective Exposure-Disease Odds Ratios: Study Factors for which the Odds Ratio Estimates Differed.

STUDY FACTOR	CASE-CONTROL GROUP ONE COMPARISON			CASE-CONTROL GROUP TWO COMPARISON		
	Prospective Odds Ratio (95% C.I.)	Retrospective Odds Ratio (95% C.I.)	Direction of Change	Prospective Odds Ratio (95% C.I.)	Retrospective Odds Ratio (95% C.I.)	Direction of Change
13. Among ever smokers the duration (years) of smoking history						
• Ever (1-10) vs Never	1.35 (.79-2.32)	1.62 (.93-2.8)	No association	2.01 (1.16-3.5)	1.85 (1.08-3.17)	Positive (+) association
• Ever (11-19) vs Never	1.64 (.92-2.90)	2.15 (1.15-3.99)	No association to a positive (+) association	2.20 (1.15-3.58)	2.53 (1.37-4.68)	Positive (+) association
• Ever (20-29) vs Never	1.16 (.71-1.91)	.89 (.51-1.53)	No association	1.94 (1.16-3.24)	1.34 (.77-2.34)	Positive (+) association to no association
• Ever ( $\geq 30$ ) vs Never	1.04 (.70-1.56)	1.14 (.76-1.70)	No association	1.62 (1.08-2.4)	1.87 (1.24-2.8)	Positive (+) association

Table 13: Estimation of Exposure-Disease Odds Ratios for the Selected Exposures and Conditions Included in the Exposure Data Validity Scale

Exposure Variable/Antecedent Event/Condition	Case-Control Comparison Groups	
	Case-Control Group One (Exposure-Disease) Odds Ratio (95% confidence interval)	Case-Control Group Two (Exposure-Disease) Odds Ratio (95% confidence interval)
1. History of dysmenorrhea before age 40 years • Yes vs No	.87 (.63-1.20)	.85 (.62-1.17)
2. A previous history of traumatic injury to the chest or breast area • Yes vs No	1.75 (1.09-2.83)	1.81 (1.12-2.93)
3. Use of artificial sweeteners • Yes vs No	.81 (.59-1.11)	1.02 (.74-1.40)
4. Consumption of food items containing MSG (monosodium glutamate) • Yes vs No	.83 (.47-1.48)	1.57 (.83-2.95)
5. Exposure to nicotine (cigarette smoking) • Yes vs No	1.34 (.96-1.82)	1.91 (1.37-2.66)
6. Drug exposures: (Yes vs No)		
a. elavil (mood elevator)	.56 (.26-1.20)	.91 (.41-2.04)
b. reserpine (an anti-hypertensive drug)	1.12 (.53-2.38)	1.17 (.55-2.50)
c. steroids such as cortisone	1.50 (.94-2.40)	1.19 (.76-1.87)
d. mellal (fake drug)	1.16 (.48-2.82)	2.80 (.95-8.17)
e. thyroxin replacement therapy	1.74 (.57-5.29)	2.15 (.72-6.42)

Note 1: (\*) Denotes reference category for the calculation of exposure-disease odds ratios.

Table 13: (continued) Estimation of Exposure-Disease Odds Ratios for the Selected Exposures and Conditions Included in the Exposure Validity Scale

Exposure Variable/Antecedent Event/Condition	Case-Control Comparison Groups	
	Case-Control Group One (Exposure-Disease) Odds Ratio (95% confidence interval)	Case-Control Group Two (Exposure-Disease) Odds Ratio (95% confidence interval)
7. Hormone use: a. oral contraceptives • Yes vs No  b. estrogen replacement therapy (post-menopause use) • Yes vs No	.65 (.47-.90)  .82 (.58-1.15)	.85 (.62-1.18)  .58 (.41-.81)
8. A history of an underactive thyroid • Yes vs No	1.32 (.84-2.67)	1.17 (.75-1.82)
9. Exposure to food additives (nitrates) • Yes vs No	8.2 (1.02-4.25)	3.19 (1.40-6.76)
10. Experiencing prolonged periods of personal stress • Yes vs No	1.15 (.83-1.60)	1.36 (.98-1.89)
11. Regular sunbathing • Yes vs No	.76 (.55-1.05)	.71 (.51-.98)
12. Overexposure to ultraviolet radiation • Yes vs No	.87 (.49-1.55)	1.25 (.68-2.30)
13. Exposure to phenoxy-herbicide (chemical weed killers) • Yes vs No	.66 (.44-.99)	.72 (.47-1.08)
14. Exposure to ureaformaldehyde • Yes vs No	2.19 (.97-4.96)	1.21 (.58-2.51)
15. Exposure to the chemicals used in the manufacture of paints, paint removers, solvents, varnishes, wood stains • Yes vs No	2.18 (.96-4.96)	1.21 (.58-2.51)

Table 13: (continued) Estimation of Exposure-Disease Odds Ratios for the Selected Exposures and Conditions Included in the Exposure Validity Scale

Exposure Variable/Antecedent Event/Condition	Case-Control Comparison Groups	
	Case-Control Group One (Exposure-Disease) Odds Ratio (95% confidence interval)	Case-Control Group Two (Exposure-Disease) Odds Ratio (95% confidence interval)
16. History of mastitis during breast-feeding. • Yes vs No	1.06 (.60-1.86)	.89 (.51-1.55)
17. Exposure to asbestos • Yes vs No	.85 (.61-1.18)	.91 (.66-1.28)
18. Consumption of Dietary Fat: • Low  • Normal  • High	* (Note 1)  1.20 (.86-1.69)  1.69 (.74-3.85)	*  1.49 (1.02-2.01)  3.17 (1.26-7.94)
19. Exposure of mother to diethylstilbesterol (DES), a drug to prevent miscarriage • Yes vs No	.99 (.99-1.00)	.99 (.98-1.00)
20. History of breastfeeding • Yes vs No	.83 (.53-1.32)	.86 (.52-1.42)
21. Alcohol Consumption • Yes vs No	1.12 (.79-1.58)	1.09 (.77-1.55)
22. Breast-feeding • Yes vs No	1.08 (.79-1.49)	.76 (.55-1.04)

Table 14.1: Exposure-Data Validity Scale Analysis: The Determination of Study Group Differences For Individual Scale Exposure Variables and the Aggregate Validity Scale Summary Score (VSSCORE) - (Part 1 - Use of Weighting Factors)

Validity Scale Exposure Factors		Factor Weighting	Analysis of Variance of Study Group Differences (ANOVA-F PROB)	Significant Group Differences
Correspondence Analysis Abbreviation code	Variable Definition			
1. A	Breast trauma/injury	.061	.025*** <sup>1</sup>	Case-Control Group 2
2. D	Smoking (nicotine exposure)	.795	.021**	Case-Control Group 1
3. H	Mastitis/breast infection	.093	.072	
4. E	Dietary fat consumption	.785	.472	
5. I	Alcohol consumption	.036	.314	
6. J	Maternal exposure to (DES) Diethylstilbestrol	.338	.293	
7. L	Personal stress	.343	.156	
8. N	Exposure to ultraviolet radiation	.601	.253	
9. S	Exposure to various chemicals in paints, paint removers, solvents, varnishes, wood stains	.279	.415	
10. P	Ureaformaldehyde exposure	.212	.257	
11. Q	Asbestos exposure	.338	.486	
12. R	Chemical weed killers (phenoxy-herbicides)	.312	.129	

Notes: 1. Statistically significant group differences exist at the 0.05 level (\*\*)  
2. None of the study groups are significantly different at the 0.05 level for the VSSCORE (\*)

Table 14.1 (continued): Exposure-Data Validity Scale Analysis: The Determination of Study Group Differences for Individual Scale Exposure Variables and the Aggregate Validity Scale Summary Score (VSSCORE)

Validity Scale Exposure Factors		Factor Weighting	Analysis of Variance of Study Group Differences (ANOVA-F PROB)	Significant Group Differences
Correspondence Analysis Abbreviation code	Variable Definition			
13. Y	Fake drug (mellal)	-(.903)	.110	
14. CC	Use of steroids	.341	.208	
15. EE	Nitrates/nitrites in foods	.411	.016 **	Case Control Group 2
16. FF	Postmenopausal estrogen replacement therapy	.630	.661	
17. GG	Oral contraceptive use	.586	.643	
18. U	MSG/food additives	.009	.062	
19. C	Artificial sweeteners	-(.134)	.119	
20. V	Antihypertensive drug - reserpine	-(.660)	.921	
VALIDITY SCALE SUMMARY SCORE (VSSCORE)		.823.* <sup>2</sup>		

Table 14.2: Exposure-Data Validity Scale Analysis: The Determination of Study Group Differences For Individual Scale Exposure Variables and the Aggregate Validity Scale Summary Score (VSSCORE) - (Part 2 - Weighting Factors Not Applied)

Validity Scale Exposure Factors		Factor Weighting	Analysis of Variance of Study Group Differences (ANOVA-F PROB)	Significant Group Differences
Correspondence Analysis Abbreviation code	Variable Definition			
1. A	Breast trauma/injury	.061	.025 ** <sup>1</sup>	Case-Control Group 2
2. D	Smoking (nicotine exposure)	.795	.021 **	Control Groups 1 and 2
3. H	Mastitis/breast infection	.093	.072	
4. E	Dietary fat consumption	.785	.472	
5. I	Alcohol consumption	.036	.822	
6. J	Maternal exposure to (DES) Diethylstilbestrol	.338	.293	
7. L	Personal stress	.343	.156	
8. N	Exposure to ultraviolet radiation	.601	.174	
9. S	Exposure to various chemicals in paints, paint removers, solvents, varnishes, wood stains	.279	.943	
10. P	Ureaformaldehyde exposure	.212	.126	
11. Q	Asbestos exposure	.338	.715	
12. R	phenoxyherbicides	.312	.131	
13. Y	Fake drug (mellal)	-(.903)	.110	

Notes: 1. Statistically significant group differences exist at the 0.05 level (\*\*)  
2. None of the study groups are significantly different at the 0.05 level for the VSSCORE (\*)

Table 14.2 (continued): Exposure-Data Validity Scale Analysis: The Determination of Study Group Differences for Individual Scale Exposure Variables and the Aggregate Validity Scale Summary Score (VSSCORE) (Part 2 - Weighting Factors Applied)

Validity Scale Exposure Factors		Factor Weighting	Analysis of Variance of Study Group Differences (ANOVA-F PROB)	Significant Group Differences
Correspondence Analysis Abbreviation code	Variable Definition			
14. CC	Use of steroids	.341	.208	
15. EE	Nitrates/nitrites	.411	.016**	Case-Control Group 2
16. FF	Postmenopausal estrogen replacement therapy	.630	.002**	Case Control Group 2
17. GG	Oral contraceptive use	.586	.017**	Control Groups 1 and 2
18. U	MSG/food additives	.009	.062	
19. C	Artificial sweeteners	-(.134)	.119	
20. V	Antihypertensive - reserpine	-(.660)	.921	
VALIDITY SCALE SUMMARY SCORE (VSSCORE)		.229* <sup>2</sup>		

Table 14.3: Exposure-Data Validity Scale Analysis (Part 3): Exposure Factors Identified by Breast Cancer Subjects as Being Relevant for Disease Occurrence

Validity Scale Exposure Factors		Factor Weighting	Analysis of Variance of Study Group Differences (ANOVA-F PROB)	Significant Group Differences
Correspondence Analysis Abbreviation code	Variable Definition			
1. A	Traumatic injury to breast/chest area	.273	.025** <sup>1</sup>	Case-Control Group 2
2. D	Smoking history (Yes vs No, amount, duration)	-(.577)	.022**	Control Groups 1 and 2
3. E	Dietary fat consumption	-(.561)	.445	
4. I	Alcohol consumption (Yes vs No, amount, duration)	.327	.304	
5. L	Personal stress	.066	.152	
6. R	Phenoxyherbicides (weed killers)	.081	.125	
7. EE	Exposure to nitrites and nitrates	-(.248)	.016**	Case Control Group 2
8. FF	Postmenopausal estrogen replacement therapy	-(.294)	.664	
9. GG	Oral contraceptive	-(.249)	.633	
10. U	Food additives: monosodium glutamate	.401	.219	
11. C	Consumption of artificial sweeteners	.490	.122	
12. H	Chronic breast infection: mastitis	.265	.069	

Notes: 1. Statistically significant group differences exist at the 0.05 level (\*\*)  
2. None of the study groups are significantly different at the 0.05 level for the VSSCORE (\*)

Table 14.3 (continued): Exposure-Data Validity Scale Analysis (Part 3): Exposure Factors Identified by Breast Cancer Subjects as Being Relevant for Disease Occurrence

Validity Scale Exposure Factors		Factor Weighting	Analysis of Variance of Study Group Differences (ANOVA-F PROB)	Significant Group Differences
Correspondence Analysis Abbreviation code	Variable Definition			
13. Family History	Family history of breast cancer in first degree relatives (mother/sister)	-(.248)	.317	
VALIDITY SCALE SUMMARY SCORE (VSSCORE)		.616* <sup>2</sup>		

Table 15: Contingency Table of Subject Responses Classified by Risk Factor and Perceived Level of Risk (Etiologic Importance)

OBSERVED FREQUENCY\*\*\*

Risk Factor	Perceived Level of Risk (Etiologic Importance)					
	None	No Opinion	Low	Moderate	High	Total
A	36	3	43	35	30	147
B	96	3	28	13	7	147
C	40	6	49	34	18	147
D	11	0	20	37	79	147
E	12	0	18	40	77	147
F	50	4	48	35	10	147
G	20	1	44	51	31	147
H	31	4	40	47	25	147
I	37	2	40	47	21	147
J	15	10	31	46	45	147
K	47	5	42	26	27	147
L	29	0	29	46	43	147
M	7	3	25	48	64	147
N	15	1	27	42	62	147
O	24	2	36	51	34	147
P	26	7	32	45	37	147
Q	25	4	31	39	48	147
R	20	6	36	41	44	147
S	38	6	52	38	13	147
T	78	4	38	21	6	147
U	30	4	49	47	17	147

Note: This table outlines the responses of 147 Screening Mammography subjects in the assessment of the level of risk associated with 33 controversial/unrelated factors for breast cancer development. Each factor has been rated on a 5-point scale: high, moderate, low, no risk or no opinion.

Table 15 (continued): Contingency Table of Subject Responses Classified by Risk Factor and Perceived Level of Risk (Etiologic Importance)

Risk Factor	Perceived Level of Risk (Etiologic Importance)					
	None	No Opinion	Low	Moderate	High	Total
V	51	29	42	18	7	147
W	72	19	38	14	4	147
X	52	27	47	16	5	147
Y	48	47	35	14	3	147
Z	50	33	39	19	6	147
AA	53	30	37	22	5	147
BB	48	29	37	26	7	147
CC	8	15	32	45	47	147
DD	35	18	53	28	13	147
EE	16	1	33	62	35	147
FF	7	2	27	58	53	147
GG	8	1	36	48	54	147
TOTAL	1135	326	1214	1199	977	4851

Table 16: Definition of Risk Factor Codes

Code Abbreviation	Code Description
A	injury to the breast/chest area
B	breast size
C	use of artificial sweeteners
D	smoking (nicotine exposure)
E	dietary fats/cholesterol
F	history of dysmenorrhea
G	coffee (caffeine/xanthines)
H	history of chronic mastitis
I	alcohol consumption
J	maternal exposure to DES (diethylstilbesterol)
K	viral infections (hepatitis, HIV, etc.)
L	personal stress
M	the presence of breast implants
N	over-exposure to ultraviolet radiation
O	exposure (at work or home) to various chemicals
P	exposure to ureformaldehyde
Q	exposure to asbestos
R	exposure to phenoxyherbicides (i.e., weed killers)
S	exposure to the chemicals in cleaning solutions
T	use of underarm deodorants
U	consumption of food additives
V	reserpine (an antihypertensive drug)
W	analgesics (pain medications)
X	use of diuretics (i.e., lasix)
Y	use of a 'fake' drug (i.e., mellal)
Z	capoten (an antihypertensive drug)

Table 16 (continued): Definition of Risk Factor Codes

Code Abbreviation	Code Description
AA	high blood pressure medications
BB	mood elevating drugs
CC	steroids
DD	thyroxin replacement therapy
EE	ingestion of nitrates/nitrites
FF	postmenopausal estrogen therapy
GG	oral contraceptives

Table 17: Response Frequency Expressed as Percentages of Marginal Row Totals

PERCENTS OF ROW TOTALS\*\*\*

Risk Factor	Perceived Level of Risk (Etiologic Importance)						Inclusion Rule (%)
	None	No Opinion	Low	Moderate	High	Total	
A	24.5	2.0	29.3	23.8	20.4	100	73.47*
B	65.3	2.0	19.0	8.8	4.8	100	32.65
C	27.2	4.1	33.3	23.1	12.2	100	68.70
D	7.5	0.0	13.6	25.2	53.7	100	92.52*
E	8.2	0.0	12.2	27.2	52.4	100	91.83*
F	34.0	2.7	32.7	23.8	6.8	100	63.26
G	13.6	0.7	29.9	34.7	21.1	100	85.71*
H	21.1	2.7	27.2	32.0	17.0	100	76.19*
I	25.2	1.4	27.2	32.0	14.3	100	73.47*
J	10.2	6.8	21.1	31.3	30.6	100	82.99*
K	32.0	3.4	28.6	17.7	18.4	100	64.63
L	19.7	0.0	19.7	31.3	29.3	100	80.27*
M	4.8	2.0	17.0	32.7	43.5	100	93.20*
O	10.2	0.7	18.4	28.6	42.2	100	89.12*
P	16.3	1.4	24.5	34.7	23.1	100	82.40*
Q	17.7	4.8	21.8	30.6	25.2	100	77.50*
R	17.0	2.7	21.1	26.5	32.7	100	80.27*
S	13.6	4.1	24.5	27.9	29.9	100	82.31*
T	25.9	4.1	35.4	25.9	8.8	100	70.06*
U	53.1	2.7	25.9	14.3	4.1	100	44.20
V	20.4	2.7	33.3	32.0	11.6	100	76.86*
W	34.7	19.7	28.6	12.2	4.8	100	45.57
X	49.0	12.9	25.9	9.5	2.7	100	38.09
Y	35.4	18.4	32.0	10.9	3.4	100	46.25

Table 17 (continued): Response Frequency Expressed as Percentages of Marginal Row Totals

Risk Factor	Perceived Level of Risk (Etiologic Importance)						Inclusion Rule (%)
	None	No Opinion	Low	Moderate	High	Total	
Z	32.7	32.0	23.8	9.5	2.0	100	35.37
AA	34.0	22.4	28.5	12.9	4.1	100	43.54
BB	36.1	20.4	25.2	15.0	3.4	100	43.54
CC	32.7	19.7	25.2	17.7	4.8	100	47.62
DD	5.4	10.2	21.8	30.6	32.0	100	84.35*
EE	23.8	12.2	36.1	19.0	8.8	100	63.94
FF	10.9	0.7	22.4	42.2	23.8	100	88.44*
GG	4.8	1.4	18.4	39.5	36.1	100	93.88*
HH	5.4	0.7	24.5	32.7	35.7	100	93.87*
Total	23.4	6.7	25.0	24.7	20.1	100	-

Table 18: Analysis of the Frequency of Response Percentages: Factors Identified as 'Plausible' Risk Factors for Breast Cancer Development Using the Subjective Inclusion Rule

Factor	Percentage Response (%)	Classification by Inclusion Rule	Factor Risk Classification by Correspondence Analysis
FF	93.88	High Risk	High Risk
GG	93.87	High Risk	High Risk
M	93.20	High Risk	High Risk
D	92.52	High Risk	High Risk
E	91.83	High Risk	High Risk
N	89.12	Moderate Risk	High Risk
EE	88.40	Moderate Risk	Moderate Risk
G	84.35	Moderate Risk	Moderate Risk
CC	84.35	Moderate Risk	Moderate Risk
J	82.99	Moderate Risk	Moderate Risk
O	82.40	Moderate Risk	Moderate Risk
R	82.31	Moderate Risk	Moderate Risk
Q	80.27	Moderate Risk	Moderate Risk
L	80.27	Moderate Risk	Moderate Risk
P	77.50	Low Risk	Moderate Risk
U	76.86	Low Risk	Moderate Risk
H	76.19	Low Risk	Moderate Risk
A	73.47	Low Risk	Moderate Risk
S	70.06	Low Risk	Low Risk

Notes: The Inclusion Rule

1. By the inclusion rule, a study factor is considered 'plausible' as a risk factor for breast cancer development if a minimum of 70% of respondents endorse the factor as either high, moderate or low risk for the disease outcome (i.e., breast cancer).
2. If the overall response rate is greater than 90% - the factor is then classified as high risk overall; 80-90% (moderate risk); 70-70.99% (low risk).
3. The classification rule, like the inclusion rule, is SUBJECTIVE. However, in general, there is good agreement between classification and inclusion of risk factors based on the simple but crude subjective rules when compared to the formal statistical analysis procedure -- correspondence analysis.

Table 19: Numerical Output from the Correspondence Analysis: The Principal Inertia (Eigenvalues) and Total Inertia, the Percentages of Inertia and the Cumulative Percentages.

# ANALYSIS OF THE OBSERVED FREQUENCY TABLE

Total Inertia = Sum of Eigenvalues = 0.3458

Axis	Eigenvalue	Percent of Inertia	Cumulative Percentage
1	0.25072	72.5	72.5
2	0.06116	17.7	90.2
3	0.2866	8.3	98.5
4	0.00530	1.5	100.0

Maximum number of factors to extract . . . . . 4

Cut-off tolerance . . . . . 90.00%

Number of factors accounting for 90% of inertia . . . . . 2

Number of factors actually extracted . . . . . 2

Chi-square value with 128 DF = 1677.664

Chi-square associated P-value 0.000

**Table 20: The Analysis of Row and Column Coefficients: Absolute Contributions (CTR), Squared Correlations (COR), Distance of the Profiles from the Origin (FACT T), Profile Masses (MASS), and the Quality of the Representation of the Row and Column Profiles (QLT)**

ROW	NAME	MASS	QLT	INR	FACT T	COR	CTR	FACT T	COR	CTR
					AXIS -- 1			AXIS -- 2		
1	A	0.030	0.787	0.001	0.061	0.091	0.000	0.168	0.697	0.014
2	B	0.030	0.821	0.031	-0.666	0.436	0.054	0.625	0.385	0.194
3	C	0.030	0.653	0.002	-0.134	0.237	0.002	0.178	0.415	0.016
4	D	0.030	0.825	0.024	0.795	0.803	0.076	-0.133	0.022	0.009
5	E	0.030	0.841	0.023	0.785	0.822	0.075	-0.120	0.019	0.007
6	F	0.030	0.895	0.006	-0.253	0.349	0.008	0.317	0.546	0.050
7	G	0.030	0.615	0.004	0.278	0.532	0.009	0.110	0.083	0.006
8	H	0.030	0.483	0.002	0.093	0.161	0.001	0.132	0.322	0.009
9	I	0.030	0.646	0.003	0.036	0.015	0.000	0.231	0.631	0.026
10	J	0.030	0.986	0.005	0.338	0.750	0.014	-0.190	0.237	0.018
11	K	0.030	0.717	0.002	-0.110	0.162	0.001	0.203	0.556	0.020
12	L	0.030	0.943	0.004	0.343	0.825	0.014	0.130	0.118	0.008
13	M	0.030	0.978	0.015	0.683	0.926	0.056	-0.162	0.052	0.013
14	N	0.030	0.927	0.012	0.601	0.918	0.044	-0.058	0.008	0.002
15	O	0.030	0.796	0.003	0.279	0.714	0.009	0.094	0.082	0.004
16	P	0.030	0.902	0.002	0.212	0.894	0.005	-0.021	0.008	0.000
17	Q	0.030	0.902	0.004	0.338	0.902	0.014	-0.004	0.000	0.000
18	R	0.030	0.978	0.003	0.312	0.942	0.012	-0.061	0.036	0.002
19	S	0.030	0.496	0.004	-0.156	0.202	0.003	0.187	0.293	0.017
20	T	0.030	0.947	0.017	-0.539	0.508	0.035	0.501	0.439	0.124
21	U	0.030	0.239	0.003	0.009	0.001	0.000	0.164	0.238	0.013
22	V	0.030	0.996	0.015	-0.660	0.886	0.053	-0.233	0.111	0.027
23	W	0.030	0.945	0.018	-0.729	0.914	0.064	0.133	0.030	0.009
24	X	0.030	0.981	0.015	-0.679	0.923	0.056	-0.170	0.058	0.014
25	Y	0.030	0.994	0.038	-0.903	0.657	0.099	-0.647	0.337	0.208
26	Z	0.030	0.998	0.018	-0.702	0.820	0.060	-0.328	0.179	0.053
27	AA	0.030	0.991	0.016	-0.680	-0.882	0.056	-0.239	0.109	0.028

Note: This is the Numerical Output from the Correspondence Analysis of the Contingency Table Data in Table 15: The Decomposition of Inertia Among the Rows (Risk Factors) and Columns (Perceived Level of Risk) Along the First Two Principal Axes (Axes 1 and 2).

Table 20 (continued): The Analysis of Row and Column Coefficients: Absolute Contributions (CTR), Squared Correlations (COR), Distance of the Profiles from the Origin (FACT T), Profile Masses (MASS), and the Quality of the Representation of the Row and Column Profiles (QLT)

ROW	NAME	MASS	QLT	INR	FACT T	COR	CTR	FACT T	COR	CTR
					AXIS -- 1			AXIS -- 2		
28	BB	0.030	0.986	0.013	-0.595	0.833	0.043	-0.256	0.154	0.032
29	CC	0.030	0.991	0.007	0.341	0.477	0.014	-0.354	0.514	0.062
30	DD	0.030	0.683	0.005	-0.326	0.623	0.013	-0.101	0.060	0.005
31	EE	0.030	0.680	0.008	0.411	0.666	0.020	0.061	0.014	0.002
32	FF	0.030	0.959	0.013	0.630	0.938	0.048	-0.094	0.021	0.004
33	GG	0.030	0.981	0.011	0.586	0.970	0.042	-0.063	0.011	0.002
1	None	0.234	0.930	0.089	-0.539	0.763	0.271	0.252	0.167	0.244
2	No- Opinion	0.067	0.999	0.098	-0.909	0.565	0.221	-0.796	0.433	0.696
3	Low	0.250	0.334	0.014	-0.125	0.287	0.016	0.051	0.047	0.011
4	Moderate	0.247	0.724	0.034	0.312	0.720	0.096	0.024	0.004	0.002
5	High	0.201	0.917	0.111	0.702	0.891	0.395	-0.121	0.026	0.048

Table 21: Definitions of the Column and Row Coefficients

<p>COR = <math>\cos^2\theta</math></p>	<p>The value of the <u>Squared Correlations</u> (COR) represent the relative contribution of the principal axis to the respective column or row point's inertia.</p> <p><u>Interpretation</u> (COR). A <b>high</b> <math>\cos^2\theta</math> value indicates that a point is practically on the principal axis, and there is little error in its geographical display. When <math>\cos^2\theta</math> is high, the axis explains the point's inertia well. If the COR value is <b>low</b>, the PROFILE VECTOR is said "to lie in the direction of the respective axis". In other words, it correlates with the axis.</p>
<p>QLT</p>	<p>The value of the <b>Quality</b> (QLT) term indicates how well points are represented in the geographical display. (<math>QLT = COR^1 + COR^2</math>)</p> <p><b>Low</b> QLT values indicate poor representation. This analysis varies depending on the subspace chosen.</p>
<p>MASS</p>	<p>The value of the <b>MASS</b> term indicates the weight of a point. The assigning of different masses to the frequency vectors amounts to attaching different degrees of importance to the positions of the respective row/column points in low-dimension spatial representation.</p> <p>In Correspondence Analysis, one identifies a low-dimensional sub-space which lies closest to all the data points. Because data points have different masses, the subspace will lie closer to the points of <b>higher</b> mass.</p>
<p>CTR</p>	<p>The value of the Absolute Contribution (CTR) represents the rows or columns relative inertia contribution to the axis.</p>

Note: These definitions and interpretations have been taken directly from the following sources on Correspondence Analysis: Lebart et al. (1984); Greenacre (1984, 1993); BMDP Statistical Software (1990).

Table 22: The Assessment of the Accuracy of the Two-Dimensional Graphical Representation: QLT Analysis

ROW	NAME	MASS	QLT	INR	FACT T	COR	CTR	FACT T	COR	CTR
					AXIS -- 1			AXIS -- 2		
1	A	0.030	0.787	0.001	0.061	0.091	0.000	0.168	0.697	0.014
2	B	0.030	0.821	0.031	-0.666	0.436	0.054	0.625	0.385	0.194
3	C	0.030	0.653	0.002	-0.134	0.237	0.002	0.178	0.415	0.016
4	D	0.030	0.825	0.024	0.795	0.803	0.076	-0.133	0.022	0.009
5	E	0.030	0.841	0.023	0.785	0.822	0.075	-0.120	0.019	0.007
6	F	0.030	0.895	0.006	-0.253	0.349	0.008	0.317	0.546	0.050
7	G	0.030	0.615	0.004	0.278	0.532	0.009	0.110	0.083	0.006
8	H	0.030	0.483	0.002	0.093	0.161	0.001	0.132	0.322	0.009
9	I	0.030	0.646	0.003	0.036	0.015	0.000	0.231	0.631	0.026
10	J	0.030	0.986	0.005	0.338	0.750	0.014	-0.190	0.237	0.018
11	K	0.030	0.717	0.002	-0.110	0.162	0.001	0.203	0.556	0.020
12	L	0.030	0.943	0.004	0.343	0.825	0.014	0.130	0.118	0.008
13	M	0.030	0.978	0.015	0.683	0.926	0.056	-0.162	0.052	0.013
14	N	0.030	0.927	0.012	0.601	0.918	0.044	-0.058	0.008	0.002
15	O	0.030	0.796	0.003	0.279	0.714	0.009	0.094	0.082	0.004
16	P	0.030	0.902	0.002	0.212	0.894	0.005	-0.021	0.008	0.000
17	Q	0.030	0.902	0.004	0.338	0.902	0.014	-0.004	0.000	0.000
18	R	0.030	0.978	0.003	0.312	0.942	0.012	-0.061	0.036	0.002
19	S	0.030	0.496	0.004	-0.156	0.202	0.003	0.187	0.293	0.017
20	T	0.030	0.947	0.017	-0.539	0.508	0.035	0.501	0.439	0.124
21	U	0.030	0.239	0.003	0.009	0.001	0.000	0.164	0.238	0.013
22	V	0.030	0.996	0.015	-0.660	0.886	0.053	-0.233	0.111	0.027
23	W	0.030	0.945	0.018	-0.729	0.914	0.064	0.133	0.030	0.009
24	X	0.030	0.981	0.015	-0.679	0.923	0.056	-0.170	0.058	0.014

Note: The 'Guidelines for Interpretation of QLT Values': If all the row and column profiles have a QLT value which lies close to 1.0, the researcher can conclude that the profiles are well represented by the two-dimensional display. Conversely, low QLT values indicate that the profiles lie outside the two-dimensional plane. None of the references on Correspondence Analysis have defined the lower boundary of QLT values which would be classified as LOW. Therefore, for this analysis, the QLT boundaries have been established arbitrarily as follows: QLT > .600 (high) and QLT < .600 (low). These profiles are not well represented in the two dimensional display - (Figure 2.1).

By scanning the QLT values for the included row and column profiles, one can conclude the majority of the profiles are well represented by the projection of the points in two dimensions.

Table 22 (continued): The Assessment of the Accuracy of the Two-Dimensional Graphical Representation: QLT Analysis

ROW	NAME	MASS	QLT	INR	FACT T	COR	CTR	FACT T	COR	CTR
					AXIS -- 1			AXIS -- 2		
25	Y	0.030	0.994	0.038	-0.903	0.657	0.099	-0.647	0.337	0.208
26	Z	0.030	0.998	0.018	-0.702	0.820	0.060	-0.328	0.179	0.053
27	AA	0.030	0.991	0.016	-0.680	-0.882	0.056	-0.239	0.109	0.028
28	BB	0.030	0.986	0.013	-0.595	0.833	0.043	-0.256	0.154	0.032
29	CC	0.030	0.991	0.007	0.341	0.477	0.014	-0.354	0.514	0.062
30	DD	0.030	0.683	0.005	-0.326	0.623	0.013	-0.101	0.060	0.005
31	EE	0.030	0.680	0.008	0.411	0.666	0.020	0.061	0.014	0.002
32	FF	0.030	0.959	0.013	0.630	0.938	0.048	-0.094	0.021	0.004
33	GG	0.030	0.981	0.011	0.586	0.970	0.042	-0.063	0.011	0.002
1	None	0.234	0.930	0.089	-0.539	0.763	0.271	0.252	0.167	0.244
2	No- Opinion	0.067	0.999	0.098	-0.909	0.565	0.221	-0.796	0.433	0.696
3	Low	0.250	0.334	0.014	-0.125	0.287	0.016	0.051	0.047	0.011
4	Moderate	0.247	0.724	0.034	0.312	0.720	0.096	0.024	0.004	0.002
5	High	0.201	0.917	0.111	0.702	0.891	0.395	-0.121	0.026	0.048

Table 23: The Evaluation of the Accuracy of the Two-Dimensional Graphical Display:  
CTR (Contributions to Inertia) and COR (Contributions to the Principal Axis)  
Analyses

ROW	NAME	MASS	QLT	INR	FACT T	COR	CTR	FACT T	COR	CTR
					AXIS -- 1			AXIS -- 2		
1	A	0.030	0.787	0.001	0.061	0.091	0.000	0.168	0.697	0.014
2	B	0.030	0.821	0.031	-0.666	0.436	0.054	0.625	0.385	0.194
3	C	0.030	0.653	0.002	-0.134	0.237	0.002	0.178	0.415	0.016
4	D	0.030	0.825	0.024	0.795	0.803	0.076	-0.133	0.022	0.009
5	E	0.030	0.841	0.023	0.785	0.822	0.075	-0.120	0.019	0.007
6	F	0.030	0.895	0.006	-0.253	0.349	0.008	0.317	0.546	0.050
7	G	0.030	0.615	0.004	0.278	0.532	0.009	0.110	0.083	0.006
8	H	0.030	0.483	0.002	0.093	0.161	0.001	0.132	0.322	0.009
9	I	0.030	0.646	0.003	0.036	0.015	0.000	0.231	0.631	0.026
10	J	0.030	0.986	0.005	0.338	0.750	0.014	-0.190	0.237	0.018
11	K	0.030	0.717	0.002	-0.110	0.162	0.001	0.203	0.556	0.020
12	L	0.030	0.943	0.004	0.343	0.825	0.014	0.130	0.118	0.008
13	M	0.030	0.978	0.015	0.683	0.926	0.056	-0.162	0.052	0.013
14	N	0.030	0.927	0.012	0.601	0.918	0.044	-0.058	0.008	0.002
15	O	0.030	0.796	0.003	0.279	0.714	0.009	0.094	0.082	0.004
16	P	0.030	0.902	0.002	0.212	0.894	0.005	-0.021	0.008	0.000
17	Q	0.030	0.902	0.004	0.338	0.902	0.014	-0.004	0.000	0.000
18	R	0.030	0.978	0.003	0.312	0.942	0.012	-0.061	0.036	0.002
19	S	0.030	0.496	0.004	-0.156	0.202	0.003	0.187	0.293	0.017
20	T	0.030	0.947	0.017	-0.539	0.508	0.035	0.501	0.439	0.124
21	U	0.030	0.239	0.003	0.009	0.001	0.000	0.164	0.238	0.013
22	V	0.030	0.996	0.015	-0.660	0.886	0.053	-0.233	0.111	0.027
23	W	0.030	0.945	0.018	-0.729	0.914	0.064	0.133	0.030	0.009
24	X	0.030	0.981	0.015	-0.679	0.923	0.056	-0.170	0.058	0.014

Note: CTR and COR values are interpreted as follows:

1. If CTR > 30%, the profile contributes strongly in the creation of the axis; otherwise, it does not contribute to axis creation; and,
2. If COR > 70%, the specific variable is exclusive in its characterization of the specific axis.

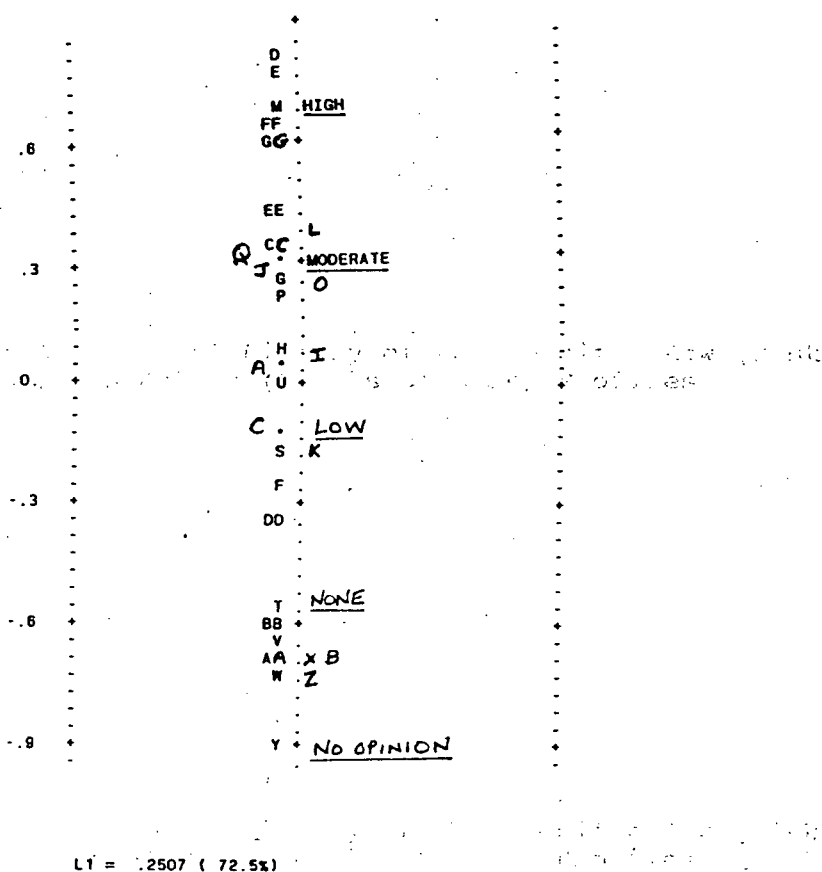
(\*) Denotes the highest COR value for that axis.

(\*\*) The factor contributes strongly to the creation of axis 3 (i.e., COR values are > 40% for the third axis.

Table 23 (continued): The Evaluation of the Accuracy of the Two-Dimensional Graphical Display: CTR (Contributions to Inertia) and COR (Contributions to the Principal Axis) Analyses

ROW	NAME	MASS	QLT	INR	FACT T	COR	CTR	FACT T	COR	CTR
					AXIS -- 1			AXIS -- 2		
25	Y	0.030	0.994	0.038	-0.903	0.657	0.099	-0.647	0.337	0.208
26	Z	0.030	0.998	0.018	-0.702	0.820	0.060	-0.328	0.179	0.053
27	AA	0.030	0.991	0.016	-0.680	-0.882	0.056	-0.239	0.109	0.028
28	BB	0.030	0.986	0.013	-0.595	0.833	0.043	-0.256	0.154	0.032
29	CC	0.030	0.991	0.007	0.341	0.477	0.014	-0.354	0.514	0.062
30	DD	0.030	0.683	0.005	-0.326	0.623	0.013	-0.101	0.060	0.005
31	EE	0.030	0.680	0.008	0.411	0.666	0.020	0.061	0.014	0.002
32	FF	0.030	0.959	0.013	0.630	0.938	0.048	-0.094	0.021	0.004
33	GG	0.030	0.981	0.011	0.586	0.970	0.042	-0.063	0.011	0.002
1	None	0.234	0.930	0.089	-0.539	0.763	0.271	0.252	0.167	0.244
2	No- Opinion	0.067	0.999	0.098	-0.909	0.565	0.221	-0.796	0.433	0.696
3	Low	0.250	0.334	0.014	-0.125	0.287	0.016	0.051	0.047	0.011
4	Moderate	0.247	0.724	0.034	0.312	0.720	0.096	0.024	0.004	0.002
5	High	0.201	0.917	0.111	0.702	0.891	0.395	-0.121	0.026	0.048

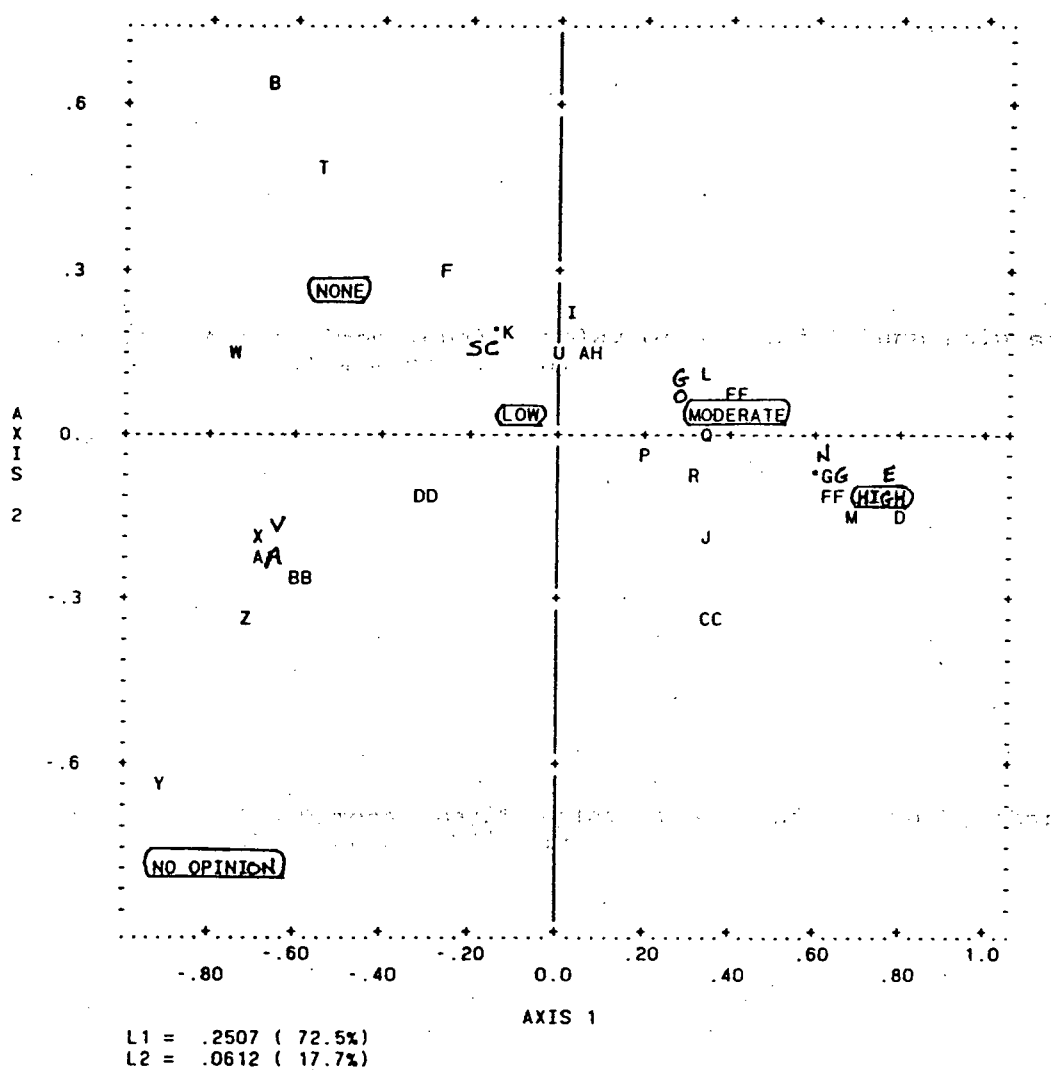
Figure 1 A One-Dimensional Display of the Specific Row (Risk Factor) and Column (Levels of Risk) Profiles



Notes: This one-dimensional display represents an approximate view of the proximities among and between row and column profiles. For example, smoking, consumption of dietary fats, breast implants, use of estrogens post-menopausally, overexposure to ultraviolet radiation and the use of oral contraceptives lie close together because the women perceived them as HIGH risk for breast cancer.

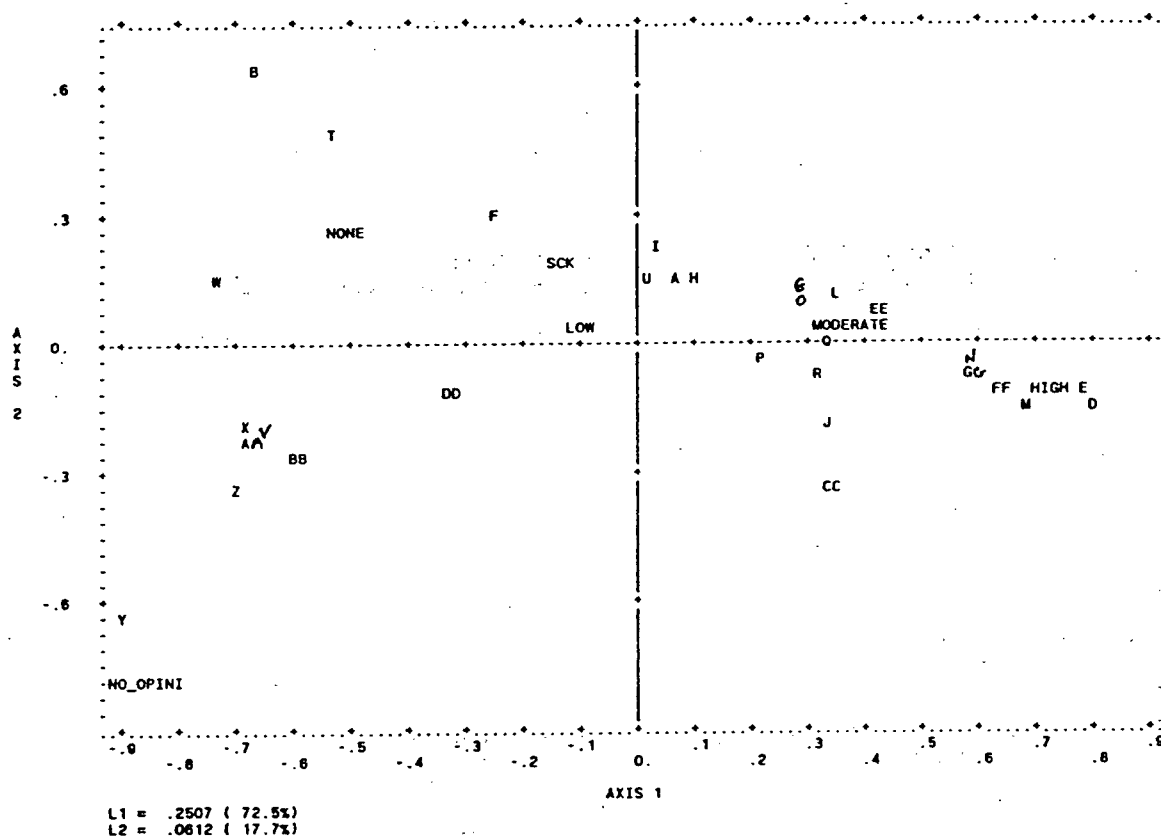
When looking at one-dimensional plots, one does not interpret points associated with low COR values because these points are not represented by the first axis.

Figure 2.1 A Two-Dimensional Display of Row and Column Points:  
A Simultaneous Display



Note: This is an optimal two-dimensional display of the contingency table data (Table 15) created by 'Correspondence Analysis'.

Figure 2.2 A Two-Dimensional Display of Row and Column Points:  
A Simultaneous Display Increased by Magnification



## Chapter 5

### DISCUSSION

#### **5.1 The Evidence for Non-Differential and Differential Exposure Misclassification in a Case-Control Study of Breast Cancer**

The critics of case-control studies strongly oppose the use of this design in epidemiological investigations, including studies of disease etiology, because of its perceived methodological limitations. Opponents base their objections on firmly held beliefs that exposure data collected retrospectively after the cases and the controls are aware of their diagnosis are subject to differential exposure misclassification. The resulting misclassification then biases the estimates of effect (exposure-disease odds ratio) in unpredictable ways (i.e., either towards or away from the null hypothesis), and consequently, invalidates the study conclusions.

Literature from cognitive and experimental psychology, the social sciences, and from survey research provide support to the critic's viewpoint by delineating numerous factors and response bias variables which could result in differential recall by the subjects in a case-control study of past exposures and antecedent events. These include the search for 'cause' cognitive processes, the salience of the event to be remembered, the methodology used to collect the exposure information, the motivation of the respondent to invest the time and energy to provide precise answers, the impact of threatening or embarrassing questions on the truthfulness of the respondent's answers, questionnaire design, the length of time since the event occurred, human memory and judgment factors, and so on. Scenarios, wherein cases and controls would remember and report exposures

differently are easy to construct; and the damaging impact of these differences (i.e., the biasing of the odds ratio) has been shown algebraically.

It is understandable why there are opponents of case-control methodology, and why they strongly believe that the exposure data in case-control studies are neither reliable nor valid, and that case-control study conclusions must be interpreted with much caution.

However, these beliefs are intuitive, and have not been confirmed through research. In fact, there has been little empirical research to establish the reliability and validity of exposure data collected in case-control studies. The relatively few studies which have been completed have been conducted primarily in reproductive medicine. Given the possibility that differential recall may be exposure and disease-specific, there is an ongoing requirement to investigate the reliability and validity of exposure data in different domains -- for different diseases and exposures, and under varying circumstances (Werler et al., 1989; Coughlin, 1990).

The studies which have addressed reliability and validity of exposure data, as well as recall bias, have not provided strong evidence to suggest that differential exposure misclassification is a significant problem, and that exposure misclassification has biased the estimates of effect so as to render the study conclusions invalid.

Therefore, this study was designed primarily to investigate the reliability and validity of exposure information collected retrospectively in a nested case-control study of breast cancer. The specific objectives were to determine if cases and controls reported their past exposures differently, to assess the impact of any resulting exposure misclassification (non-differential and/or differential) on the estimates of effect, and (most importantly), to develop and evaluate an 'exposure data validity scale' as a possible design standard for the measurement and the

control of exposure misclassification bias (should it be found to exist) in this study, and possibly, in future case-control studies.

Four different types of analyses were undertaken to address the specific research questions: 1) a test-retest reliability (correlation) analysis to determine the magnitude of the agreement between the prospective (pre-diagnostic) and retrospective (post-diagnostic) reports of exposure; 2) Cohen's Kappa analysis to determine the agreement between the two reports beyond that expected on the basis of chance alone; 3) McNemar's analysis to explore the frequency and direction of the changes in the levels of exposure reported in the two reports (Q2 versus Q1), and to determine if the study groups differentially recall their past exposure; and, 4) the estimation and comparison of the prospective and retrospective relative risk estimates (i.e., exposure-disease odds ratios). Significant odds ratios and their direction of change would indicate the presence of either non-differential or differential exposure misclassification. The three study groups were evaluated for any tendency (propensity) to overreport/overstate or underreport/understate exposure over several study factors in an attempt to determine if systematic biasing of the estimates of effect existed, and was related to case-control status.

Overall, the data in this nested case-control study of breast cancer suggested that the retrospective (Q2) exposure reports had adequate reliability and validity: the data provided no strong and conclusive evidence for the existence of either differential or non-differential exposure misclassification, or for the systematic biasing of the risk estimates (i.e., exposure-disease odds ratios) either towards or away from the null value (i.e., RR (OR)=1.0 -- (no risk or association between exposure and disease occurrence)).

The test-retest reliability analysis demonstrated that there was moderate to high correlation (consistency) between the prospective (Q1) and retrospective (Q2)

reports of exposure for most of the study variables (i.e., the correlation coefficients ranged from 0.50 to 0.97). The high levels of correlation may be partially attributable to the study limitations which include:

- 1) the lack of independence (i.e., the existing correlation between Q1 and Q2 exposure reports);
- 2) the possibility that the estimated correlation coefficients reflect shared error rather than consistency of reporting only; and,
- 3) the relatively short time interval between the administration of Q1 and Q2.

These limitations would influence the magnitude of the correlation coefficients and other measures of agreement, such as the Kappa statistic in the direction of 'overestimation'.

The fact that moderate correlations and moderate Kappa coefficients were found for some of the study variables demonstrated that there were some differences (i.e., inconsistencies) between the prospective (Q1) and retrospective (Q2) reports; however, these inconsistencies were similar among the three study groups, and were not found to be related to the subjects' knowledge about diagnosis. Prior exposures were remembered and reported with the same relative accuracy within the three study comparison groups. As such, there was no evidence that the study groups systematically overreported/overstated or underreported/understated their exposure to the various study variables: the differences or inconsistencies could be attributed to random (exposure misclassification) errors.

The assessment of the 'agreement' between the prospective (Q1) and retrospective (Q2) reports of exposure by Kappa analysis provided further evidence that the post-diagnostic exposure assessments provided a level of consistency (reproducibility) better than would be expected on the basis of chance. In summary, there was no conclusive evidence for the existence of differential

exposure misclassification (i.e., recall bias) resulting from the correlation and Kappa tests performed.

The McNemar's analysis explored the direction, prevalence and importance of any changes in reporting of exposure by the three study groups. For the majority of the study factors, the discordance rate was insignificant when compared to the overall concordance between the two reports on a particular study factor. Therefore, the data demonstrated that there was no differential exposure misclassification; the data from the McNemar's analysis did not suggest that any of the three groups exhibited a systematic shift in self-reported exposure status which was associated with the group's knowledge of diagnosis. Discrepancies between Q1 and Q2 reports were related to the exposure being reported on: it was also not possible to predict with certainty which variables were subject to error.

A comparison of the prospective and retrospective exposure-disease odds ratios for the two case-control comparison groups permitted an assessment of exposure misclassification and its impact on the estimates of effect. The prospective (Q1) and retrospective (Q2) odds ratio estimates were similar, with considerable overlap of the 95% confidence intervals for the two study comparison groups. This suggested that the retrospective exposure assessments were subject to random measurement error rather than systematic measurement bias. For the estimated Q1 and Q2 odds ratios, it was also observed that for most of the study factors, the estimates were in the same range and direction, and the increments or decrements of risk between the levels of exposure were maintained. These observations would suggest the absence of bias, as well as no compromise to the validity of the study conclusions.

These comparisons of prospective and retrospective odds ratio estimates and their 95% confidence intervals were also designed to detect the presence, if

any, of differential (DEM) and non-differential (NDEM) exposure misclassification. Evidence of DEM would consist of systematic group differences in the Q1 and Q2 odds ratio estimates in one of the following directions:

- 1) Q1 (OR) = 1.00 and Q2 (OR) > 1.00 -- a shift from no association to a positive association; and,
- 2) Q1 (OR) = 1.00 and Q2 (OR) < 1.00 -- a shift from no association to a negative association.

NDEM would be present if the Q2 (OR) tended toward an assessment of no association (i.e., Q2 (OR) = 1.00) when the respective Q1 (OR) had indicated either a positive (i.e., Q2 (OR) > 1.00) or a negative association (i.e., Q2 (OR) < 1.00).

Using the previous guidelines to assess for the presence of DEM and NDEM, the data provided no strong and conclusive evidence for the existence of either type of exposure misclassification.

In conclusion, the data and the analyses of the data in this study indicated that retrospectively collected exposure information in a case-control study of breast cancer has adequate reliability and validity, subject only to random exposure misclassification. The odds ratio estimates did not appear to have been biased by either systematic overreporting/overstating and/or underreporting/understating of exposure by the cases and controls.

From the results of this study, there was no evidence that case-control studies are unsuitable for the investigation of questions related to human health and disease. The favourable results of this study regarding the overall reliability and validity of the retrospectively collected exposure information cannot be generalized definitively to other case-control studies. Concerns regarding the reliability and validity of exposure data must always be considered by the researcher in the design, implementation and analysis of future case-control studies. Although the research literature would recommend the use of 'anamnestic

equivalents' as the most appropriate control group of choice for case-control comparisons, the results of this study do not provide supporting evidence for this recommendation. The exposure reports of control group one were not found to be more consistent (reliable) than those provided by control group two.

The correlation analysis reported similar correlations (agreement) between Q1 and Q2 reports of exposure among the three study groups. The cases, control group one and control group two subjects all showed some inconsistencies in the reporting of past exposure. No advantage was obtained by virtue of the fact that control group one was anamnesticallly equivalent to the cases except for diagnosis. That is, the members of control group one had experienced the same trauma (abnormal mammogram), had undergone the same diagnostic procedures to determine disease outcome, and had the same motivation to participate and to report their past exposures.

This was the first study which directly explored the impact of different control groups on the estimates of association between potential risk factors and disease (breast cancer), and whether or not a particular control group would have a tendency to bias the odds ratio estimates.

Lastly, Table 24 (p.334) provides a comparison of the prospective and retrospective risk estimates of this study with the magnitude of the retrospective risk associations reported in both Friedenrich's case-control study (1991) and Kelsey's summary of "generally considered established" risk factors for female breast cancer (1993). This table permits a general assessment of the consistency between the data set in this study and the known trends found in previous investigations. This is a particularly interesting, but seldom available opportunity, to compare the retrospective to prospective risk estimates.

Overall, the prospective data did not provide any evidence for increased risk for breast cancer as a result of exposure to any of the risk factors outlined in

Table 24 (p.319), with the exception of 'smoking (ever vs never)' for the case-control group two comparison: i.e., OR (95% confidence interval) = 1.91 (1.37 - 2.66). This was the only significant trend for increased risk due to exposure noted in this study.

For the retrospective data in this study, there are no significant associations between the reported risk factors (Table 24, p.319) and the outcome event (breast cancer). Unlike the findings in other case-control studies, no significant trends for increased risk were found for a family history of breast cancer (i.e., any first degree relative, mother or sister with breast cancer), late age at first full-term birth, early age at menarche and late age at menopause.

It must be noted that for case-control group one comparisons, significant risk increments for some of the exposure variables may have been missed if the control subjects were found to have a high incidence of benign breast disease. The research literature would indicate that breast cancer and benign breast disease may share common risk factors. For these factors, risk differentials between the cases and their matched controls would remain undetected, and the resulting odds ratio estimates would not be statistically significant.

This study did not find a significant association between Kelsey's "established" risk factors and breast cancer; nor were the magnitudes of the odds ratios comparable. These differences may be due to one of the following: 1) the small sample size may have resulted in insufficient study power to detect significant risk increments; 2) the cohort from which the cases and controls were selected (i.e., women participating in the Screening Mammography Program of BC) may be more similar (homogeneous) regarding exposure histories (especially the cases and control group one subjects) when compared to the women who were not eligible for this study (i.e., the BC female population who do not participate in SMP BC -- only 20 per 1000 women participate in the provincial

screening program); and, 3) the study design (including sample size, definition and selection of both the cases and the controls, the data collection method used to collect the exposure data, etc.,) may not have been similar enough to permit legitimate comparisons between the findings of this study, and those of previous investigations.

When the prospective and retrospective exposure-disease odds ratios of this study were then placed against the results of Friedenreich (1990), the risk estimates were comparable, with the exception of pregnancy history and parity (nulliparous vs parous). For these two risk factors, Friedenreich (1990) found significant odds ratio and an association between the risk factor and breast cancer. This comparability is probably due to similarities in overall design, whereas Kelsey's analysis is drawn apparently from several studies, with varying designs and possible methodological limitations.

In conclusion, the data from this study were not comparable regarding the range of risk estimates reported for the various study factors in the compilation by Kelsey (1993); however, both the prospective and retrospective risk estimates were comparable to the majority of the retrospective odds ratios reported in Friedenreich (1990). The inconsistency between this study and other studies may be due to either methodological differences or methodological limitations.

## **5.2 Evaluation of the Effectiveness of the 'Exposure Data Validity Scale'**

Case-control findings regarding the relationship between suspected exposures and the risk of disease have often been inconsistent, sometimes misleading, but most of all controversial. Consequently, the research community, medical practitioners, the media, and health care consumers are quite ambivalent as to the strength of the associations between the exposures and conditions

believed to be risk factors, and their actual association. The suitability of the case-control study to address questions of disease etiology rests on the ability of the cases and controls to provide complete and accurate exposure histories, and the extent to which the reported exposure data is valid (i.e., a reflection of the 'true' exposure).

These concerns have been raised by researchers such as Jick and Vessey (1978) when they urged researchers to consider the question of 'validity' when interpreting the results of case-control studies. They remarked that researchers must address "the question of the extent to which bias or imprecise exposure information may have influenced the results" (p.5). Without such measurements, 'uncertainty' leads to decreased confidence in the interpretation of the results (p.5). In addition, researchers such as Gordis (1979) have noted that inadequate effort and attention have been paid in assessing the quality (i.e., the reliability and validity) of the exposure data collected in case-control studies -- the data upon which the study conclusions depend (p.21). He remarked specifically, that "the need for a critical examination of the quality of epidemiological data and the methods which can be used to improve the validity of these data is particularly urgent" (p.24). Scientific rigor requires vigilance in seeking out and correcting methodological errors and bias.

These comments suggest the ongoing need to provide empirical evidence regarding the reliability and validity of exposure data within the context of each case-control study, as well as the means to correct estimates of association that have been distorted due to the presence of exposure misclassification bias.

Raphael's (1987) proposal for the construction and use of a validity scale to measure and to control recall bias offered the possibility for a case-control design strategy to empirically assess the quality of exposure data in a case-control study,

and to evaluate and correct the precision of the odds ratio estimates, if bias was found to exist.

The literature did not provide any evidence that such a scale had ever been constructed or used to measure and to control recall bias (i.e., differential exposure misclassification). Therefore, this study was significant in its attempt to design, implement and evaluate a 'validity scale' (as conceptualized by Raphael, 1987) for the measurement and the control of differential exposure misclassification (recall bias). As discussed in Section 2.7, the validity scale would consist of exposures that are 'plausible' as risk factors for the disease under investigation, but be unrelated etiologically. Because the validity scale items chosen are not associated with the disease being investigated, the proportion of exposed cases and controls should be approximately the same, and the estimate of effect among the various validity scale exposures and the occurrence of disease should equal 1.00, indicating no association. By analogy, if the "case respondents positively endorsed an excessively large number of validity scale items in comparison to the control respondents", Raphael (1987) argued that the differential endorsement by the cases would suggest "overreporting recall bias rather than actual higher rates of exposure" (p.169). She suggested that researchers could "compare the TOTAL validity scores for the cases and controls" to detect the propensity of cases and controls to remember and report their past exposures differently, and to assess the impact of the differential exposure misclassification on the odds ratio estimates. Furthermore, the resulting distortion in the estimates of effect (i.e., odds ratios) could be statistically corrected because "the validity scale score is a function of the extent of each respondent's recall bias" (p.169). The summary within groups validity score "may be entered into the final analysis as a statistical control for recall bias" (Raphael, 1987, p.169).

If an exposure data validity scale could be constructed efficiently, and be shown to be effective regarding its stated purpose -- the measurement and control of recall bias -- it could be used routinely as a design strategy within case-control research to provide the direct, empirical evidence regarding the 'quality' of the exposure data, and the validity of the estimates of association (i.e., exposure-disease odds ratios).

The specific steps in the construction of the exposure data validity scale included: 1) the search for the candidate variables (i.e., exposures that were 'plausible but unrelated' exposures to the disease under investigation); 2) the selection of the exposure variables for inclusion in the validity scale, and the assignment of exposure-specific weighting factors -- a measure of the variables' 'plausibility' as a real risk factor the disease in question; and, 3) the comparison of total validity scale score (VSSCORE) between the cases and the controls to detect group differences, which would be indicative of differential exposure misclassification.

Overall, the scale was simple to construct, and neither labor intensive, complex nor expensive; however, there were a few challenges in the development and analysis phases of the scale construction that had to be met. In the construction of the scale, it was very difficult to find exposure variables which had been shown empirically to be definitively and unequivocally unrelated to breast cancer. As previously mentioned, there is still a great deal of controversy and conflicting results in the epidemiological literature regarding the specific etiology of breast cancer. However, 45 variables were selected from the literature on the basis of showing no known biological plausibility with the occurrence of disease. With the assistance and input of a senior epidemiologist at the British Columbia Cancer Agency, 33 variables were selected for evaluation as candidate variables for inclusion in the validity scale. The process used to select the exposures that

were 'plausible risk factors for breast cancer' was fairly responsive and successful in retrospect. Proof was evident when the variables selected in this study were compared to the list of variables identified by breast cancer patients as being etiologically important for the development of breast cancer. These women identified 19 variables which they believed were responsible for disease occurrence; and, 17 of these variables had been evaluated in this study as candidate exposures for inclusion in the validity scale (personal communication - Dr. N. Waxler-Morrison). Therefore, it can be concluded that the search for equally 'plausible' exposure variables is possible, and the process is neither time-intensive or costly. The more familiar a researcher is with the empirical studies relating to a particular disease domain and/or the natural history of the disease itself, the easier the task will be.

The next challenge faced in the development of the validity scale was to determine which of the variables would be considered "equally" plausible when compared to the identified risk factors. Raphael (1987) stated that unless the exposures were of approximately the same plausibility, "the validity scale will not appropriately measure 'search for cause' cognitive processes" (p.169). In addition, it was necessary to find the means for differentiating between the candidate variables with respect to their relative 'plausibility' - that is, the importance attributed to them as risk factors for breast cancer development. This was accomplished by asking a sample of the target population to assign a level of risk (i.e., high risk, moderate risk, low risk and no risk) for each of the 33 candidate variables) that would correspond to their perception of the risk posed by that factor in the development of breast cancer. The responses of 147 participants were cross-classified in a two-way contingency table according to the 5 levels of risk and the 33 exposures, and analyzed by correspondence analysis to select those variables considered most important for breast cancer development, together with

the relative importance ascribed to them (i.e., the weighting factor that would be included in the estimation of the total validity scale score). A weighting factor was deemed necessary because the exposures would not be perceived to be the 'same' regarding the degree or level of risk posed by them in the development of breast cancer. Here, correspondence analysis was extremely useful in selecting the variables for the validity scale and their specific weighting factors. Excellent resources were also available to assist in the application of the procedure, and most importantly, in the interpretation of the analysis of the 33x5 data matrix, the selection of the scale items and their weighting factors.

In general, the task of scale development was not formidable, although very challenging. It is evaluated as being easy to develop, implement and analyze. It did not take excessive amounts of research time and monies to develop: the analysis was not overly complex once the multivariate technique of correspondence analysis was understood. The amount of extra effort required by other researchers to include a comparable validity scale in their research is minimal given the perceived benefits (i.e., the ability to assess for differential exposure misclassification, and, if present, to adjust the odds ratio estimates to remove the resulting bias).

The results of the validity scale analysis reported in Chapter 4 indicated that it was suitable for its stated task -- the assessment and control of recall bias. Here, it was noted that the results of the validity scale analysis were consistent with the findings of the main study which investigated the reliability and validity of the exposure data and the respective risk estimates (i.e., exposure-disease odds ratios).

The comparison of the total validity scale score (VSSCORE) by study group indicated that the groups were similar in their reporting of exposure: there was no differential endorsement of the exposures in the validity scale to suggest that

recall bias (differential exposure misclassification) was present. In this study, the validity scale was evaluated as an effective means of assessing whether or not exposure data were reported differently by the cases and the controls, and whether or not the estimates of association between risk factor and disease could have been subject to distortion (bias) as a result of differential exposure misclassification (i.e., recall bias).

There is an ongoing requirement to develop, implement and evaluate the usefulness and effectiveness of an exposure data validity scale in other research domains to determine the scale's utility as a design standard for the measurement and control of recall bias. As noted by Raphael (personal discussion of this doctoral thesis, December 1995), "one study is insufficient to determine if the method is worth the effort". At the same time, the investigator must conduct validity substudies in order to determine if the specific 'validity scale' is doing what it is intended to do. As noted in Section 3.11.3, validity scale evidence for non-differential recall of the innocent exposures does not rule out the possibility of differential reporting of true risk factors, and different exposures may behave differently regarding recall bias (Neugebauer and Ng, 1990; Werler et al., 1989; Coughlin, 1990). More research is required to determine if the scale results reflect the study results for 'true' risk factors.

### **5.3 Study Validity**

An evaluation of study validity involved the consideration of both internal and external validity, and a determination of whether or not the study had been adversely affected by any systematic error - such as sample distortion bias (also referred to as selection bias), information bias and confounding bias (Kleinbaum et al., 1982; Rothman, 1986; Kramer, 1988; Checkoway et al., 1989; Friedenreich, 1990). Prior to analyzing this study for any validity problems, definitions of

internal and external validity, and the potential biases that could result in systematic error will be provided.

Internal validity refers to the extent to which "the analytic inference derived from the study sample is correct for the target population" (Kramer, 1982, p.48). In this study, the target population consisted of women enrolled in the Screening Mammography Program of British Columbia.

External validity refers to the generalizability of the study results to the larger 'external population' or the subjects outside the study population (Kramer, 1982, p.48).

### **5.3.1 Internal Validity**

To determine if the internal validity of the study had been compromised, it was important to assess whether or not the study had been influenced by any of the following biases: sample distortion (selection) bias, information bias and confounding.

**Sampling Distortion (Selection) Bias** - if present, results in a biased estimate of effect. This bias would occur if the procedure by which the study sample is selected is distorted, and the subjects who were recruited for the study were not representative of the target population "with respect to the joint distribution of exposure and outcome" (Kramer, 1982, p.49; Rothman, 1986). Consequently, the effect estimate would differ from that which would have been generated had it been possible to study the entire target population -- all women enrolled in the Screening Mammography Program in BC (Rothman, 1986, p.83). There are many possible sources of selection bias including design deficiencies (i.e., the selection of the sampling frame, the choice of the controls for case-control comparisons, misclassification of the disease status of control subjects, and non-response (non-participation)).

A review of the study design suggests that if selection bias had occurred, it would have been minimal, and therefore, no biasing of the study results would have occurred. Many aspects of the design chosen for use in this study (i.e., nested case-control or case-control within a cohort) protected against the problem of selection bias. First of all, it was possible to enumerate all the cases and controls: delineation of a complete sampling frame, and access to all the potential subjects was possible. There was no need to use a convenience sample: the target population and sampling frame were detailed completely. From the sampling frame, every case (i.e., women with histologically confirmed breast cancer) was selected for the study, and by means of incidence sampling, two eligible controls (from all eligible controls who were free of disease, and matched to each case with respect to the matching factors -- age ( $\pm 5$  years), mammography clinic, date of visit to clinic, and ethnicity, were randomly selected for study. Therefore, the procedures used to select the study population appear to be unbiased.

The controls that were chosen were suitable because they were all selected from the same study frame (i.e., a cohort of British Columbia women enrolled in the screening mammography program). Furthermore, two differently constituted control groups were assembled to determine if the characteristics of these groups had any negative or biasing effect on the estimates of effect for the various study factors. The results of the study demonstrated that the two control groups were similar regarding their motivation to remember and report past exposure, and regarding the consistency with which they reported past exposure once diagnostic outcome was known.

As well, subjects were not chosen because they were known to have been, or not to have been exposed to a specific factor. As a result of this design feature, the potential for bias because exposure may have been linked to other factors that may determine likelihood of disease was minimized. In this study, all cases were

approached for recruitment and participation -- the problem of self-selection was avoided.

Another strength of this design was the fact that all potential cases and controls received a mammogram to screen for the possibility of breast cancer (i.e., equal diagnostic surveillance). For those women who had an abnormal mammogram, once again, the same diagnostics were used to confirm the presence or absence of disease. Thus, selection bias due to problems of unequal diagnostic surveillance was not considered relevant in this study.

The problem of sample distortion (selection) bias would also seem to be less probable when one considers the high rate of response which was comparable for the three study groups. With significant non-response, one would be concerned that "the relationship between exposure and disease is not the same for subjects who participate, and those who would be theoretically eligible for study but do not participate" (Rothman, 1986, p.84; Greenland, 1977). Here, Greenland (1977) noted that "selection bias is a theoretical possibility whenever correlates of the outcome [an exposure in the context of case-control studies] capable of influencing study participation are existent in some individuals at the beginning of the study" (p.187). The analysis of the non-respondents in Chapter 4 also provided evidence that the respondents and non-respondents were similar regarding the various study factors (i.e., the prevalence of exposure and non-exposure).

In summary, the choice of a nested case-control study design minimized or avoided the problem of selection bias. It is not reasonable to believe that such bias exists and has adversely affected the study conclusions.

**Information bias** - refers to a distortion in the estimate of effect due to "measurement error or misclassification of the subjects on one or more variables" (Kleinbaum et al., 1982, p.191). Many sources of information bias exist. These

include: invalid measurement -- the misclassification of subjects on one or more variables (i.e., exposure and/or disease), and unequal diagnostic surveillance which results in detection bias. This last factor would also play a role in producing sampling distortion (selection) bias.

In this study, exposure data were collected both prospectively and retrospectively by the same method (i.e., a self-administered questionnaire). Because the measurement procedure used was identical, the design attempted to protect against measurement errors, and specifically, differences in measurement of exposure which could be related to different methods of data collection rather than to group differences in recall accuracy. The measurement procedures were kept identical to help protect against invalid measurements of exposure.

Misclassification of subjects with respect to the presence and absence of disease was of no concern in this study. By means of mammography and subsequent diagnostics for those subjects with an abnormal mammogram, cases were histologically confirmed cases of breast cancer, and the controls were found to be free of disease. After completion of the study, none of the control group one subjects were later found to be positive for breast cancer.

One limitation in this study was non-access to the screening enrollment questionnaires for the study population. As a result, there was no way of ensuring that the responses of the subjects with respect to exposure for the various study factors had been entered into the data base without error. It is assumed that if there was error in data entry, it was random and minimal given the fact that each screening centre followed the same protocol for the administration of the screening questionnaire, the coding, and the data entry of the exposure responses.

In summary, a problem with information bias seems unlikely, but if it did exist, it would be limited in this study.

**Confounding** - Checkoway et al. (1989) in referring to an article on epidemiological confounding by Greenland and Robins (1986) stated that 'confounding' can occur "when the exposed and non-exposed groups are not comparable because of differences in background disease risk" (p.84). In other words, one group is more susceptible (i.e., at greater risk) at the beginning of the study for developing the outcome or disease (Rothman, 1986, p.54). Rothman (1986, p.89) further described confounding as a "mixing of effects" -- the exposure(s) of interest and extraneous factors which are predictive of the occurrence of disease. The effect of the extraneous factors "need not be causal" (Rothman, 1986, p.89). Confounding leads to the biasing of the estimates of effect of the exposure(s) of interest and the generation of invalid study conclusions.

A comparison of the study population (i.e., the cases and control groups one and two), as well as the respondents and non-respondents regarding the various study factors was conducted to assess for differences that could possibly affect the results of the study. The resulting analysis demonstrated that the study groups were comparable on the majority of the factors -- suggesting minimal problems due to confounding. In addition, the study design was also used to minimize and control for confounding. In this case, potential confounders (i.e., age, ethnicity, clinic, date of mammogram) were used as matching factors for the selection of appropriate controls for the case-control comparisons.

In summary, the compromise of internal validity as a result of confounding is not considered to be a significant problem.

### **5.3.2 External Validity**

An evaluation of external validity would focus on the degree to which our study results (using screening mammography subjects in British Columbia) can be generalized to women attending other screening programs, and possibly,

Canadian women in general. The results of this study can certainly be generalized to other women attending screening mammography programs, but probably not to the general female, Canadian population which would include women from British Columbia. (It must be noted that only 20/1000 BC women participate in the provincial mammography screening program). This exception (i.e., non-generalizability of the study findings to non-participants) is related to the fact that women who attend screening programs differ from the general population on two important factors -- their interest in their personal health, and their participation in the mammography screening program. As such, these women may be better motivated and more capable of remembering and reporting their past exposures than women who do not participate. These women may also differ from the general population regarding their overall exposure history. Concern about their health may also have been translated into specific changes in their lifestyle prior to, or at the time of enrollment in SMP BC. Therefore, this study may not have shown that differential exposure misclassification was a problem because the cases and the controls were too similar (**homogeneous**) with respect to their motivation (health-seeking behaviour), their recall ability as well as their exposure history. By taking both our cases and controls from this special cohort of women (i.e., mammography screening participants), there may have been little opportunity to observe and to assess the impact of differential recall on the reporting of past exposures by the cases and the controls. This was a limitation of this study, and was possibly a significant reason for not detecting differential exposure misclassification (i.e., recall bias). It would be advantageous to complete this study using a heterogeneous population to determine if the cases and the controls differed regarding their ability to remember and to report past exposures.

In conclusion, the potential sources of bias which may have compromised study validity have been assessed. No significant limitations have been identified.

Although bias is probably present in some degree, as it is in every study, it is not considered to be so pervasive that the study conclusions have been invalidated.

#### **5.4 Limitations of the Study**

As discussed in Section 5.4, the results of the study provided no evidence for the existence of either non-differential or differential exposure misclassification, and the subsequent biasing of the estimates of effect. These results cannot be generalized to other research domains because the presence of differential recall by cases and controls is possibly exposure and disease-specific. In addition, several other factors may be responsible for producing exposure misclassification (i.e., the method used to collect data, the detail of the information that is required, the perceived level of threat or sensitivity of the required information, etc.). The results of this study apply only to women enrolled in SMP BC, as well as to other women enrolled in other such programs. Furthermore, these results can only be directly applied to other case-control studies which used a self-administered questionnaire to collect exposure and other related information.

As mentioned previously, there was no way to determine the accuracy with which prospective (Q1) reports of exposure were coded and entered into the SMP BC data base. Depending on the prevalence of error associated with this process, the retrospective (Q2) reports of exposure may be found to be more or less consistent than was the case in this present study. Inconsistency in retrospective exposure reports could possibly be related to errors in the coding of prospective (Q1) responses, and data entry rather than to exposure misclassification due to any of the response variables discussed in Chapter 2, including 'faulty memory'.

Another limitation of this study was observed by Raphael (1995) and discussed in Section 4.4. She noted that Q1 and Q2 responses were probably correlated (i.e., Q1 responses regarding prior exposure "may have served as a rehearsal" for the Q2 reports of exposure). Consequently, "what was recalled at Q2 was probably not lifetime exposure history (although this is what the researcher wants to assess in the case-control study); but instead, what was reported just several (approximately 6 months) earlier at Q1". Therefore, the results of this study probably overestimated the reliability (overall consistency) of the retrospective (Q2) reports of exposure when compared to the prospective (Q1) responses.

Non-differential exposure misclassification which results in a type II error, and the failure to show a true association, may have been missed as the result of insufficient study power. In this study, sample size calculations were based on the ability to detect a relative risk of 1.6. Subtler risks may have required more subjects in order to detect non-differential misclassification which may have masked the risk -- biased the odds ratio toward the null value. In general, the study power of 76% may have been insufficient to detect either DEM or NDEM (i.e., resulting in a type II error). In addition, if this study seriously underestimated the occurrence and extent of NDEM, the resulting 'noise' in the system may have been sufficient to prevent the detection of DEM (i.e., recall bias). Both NDEM and DEM may have been underestimated due to the correlation between the Q1 and Q2 responses. The odds ratio estimates reported in this study were based on the assumption of the independence between Q1 and Q2 response sets for the various study factors. As such, this resulted in an extremely conservative test of the comparability of the prospective and retrospective odds ratio estimates.

As discussed in Chapter 2, the Screening Mammography Program of British Columbia made changes in the enrollment questionnaire prior to the end of subject recruitment. Consequently, prevalent cases also had to be recruited.

This design change would only be a limitation in this study if the duration of illness, or length of time since diagnosis had any impact on recall accuracy. However, it is recommended that future studies use only incident cases.

The fact that neither differential or non-differential exposure misclassification were found could also possibly be related to the use of a homogeneous sampling frame that was composed of women who were probably more aware than the general population of their personal health and disease prevention/detection (as measured by their participation in the mammography screening program and this research study). Health-seeking behavior (including changes in exposure history) and personal awareness may have been responsible for the consistency of the post-diagnostic exposure reports, and the absence of non-differential and differential exposure misclassification. Exposure misclassification may have been more prevalent if cases and controls had been selected from the general population.

## **5.5 Recommendations for Future Research**

As mentioned previously, exposure misclassification is postulated to be exposure and disease-specific (Werler et al., 1989; Coughlin, 1990). Therefore, it is important that other studies be completed in other research domains and under varying conditions to assess the reliability and validity of exposure data. These studies will also provide the opportunity for researchers to determine if cases and controls differentially report past exposures, and if they do, under what conditions; they will also evaluate if the estimates of effect have been biased, rendering the conclusions invalid. Thus, the various errors and biases that may possibly affect case-control research can be studied and catalogued while concurrently providing strategies for their minimization and/or prevention.

Future research must be designed to determine if case-control studies are suitable for the study of disease etiology, and if appropriate, there must then be an effort to establish methodological standards for the conduct of future case-control studies, and ongoing refinement and improvement of the general design. These efforts might very well be done within an interdisciplinary framework.

As discussed in Chapter 2, the gathering of information (both factual and attitudinal) by means of self-reports is widely used in the social sciences. Social scientists have also acknowledged that self-reports may be highly inaccurate, and by means of research, they have identified several response bias variables as possible sources of invalid conclusions for studies. As well, they have completed extensive methodological research with respect to the validity and precision of survey methods. However, their findings and recommendations have not been integrated across the disciplines. Researchers in general, are not fully aware of what is available in subject areas other than their own. Thus, interdisciplinary research seems to be an important consideration for future research endeavours.

The epidemiologist could work with the experimental psychologist to understand how the human memory stores health-related information, including exposure histories, as well as the psychological impact of disease on both the remembering and reporting of exposure data. In cooperation with the social psychologist and survey methodologist, these same researchers could then extend their knowledge on the factors that affect recall, and collectively determine the best methodology (i.e., data collection procedures) to use in order to obtain recalled information which is reliable, accurate and complete.

A final recommendation is in the area of national policy development. Researchers must have easy access to good (valid) data bases in the conduct of their research. It is important that agencies and institutes that are being funded provincially or federally for health care and related programs such as the

Screening Mammography Program of BC, be mandated to participate in research, and specifically, to collect demographic, medical, reproductive, lifestyle, occupational, and other required data on the Canadian consumer who uses their services.

There would also be an equivalent expectation for Canadians using the government funded health care programs to provide the level of personal information required, with an understanding of how important such data are for the planning of health care programs, promotion of general health and disease prevention programs, and overall, for the improvement of a Canadian's level of health and well-being. It would also be beneficial if there was an integrated approach to what information was required, the detail of information to be collected, where and how it should be collected, etc. Information collected on individuals at the time of its occurrence, with a standard format for data collection, will provide data that will be more reliable and valid for use by researchers. Subsequent research will be better able to ascertain the determinants of health and disease, to evaluate treatment, screening and prevention programs, etc. Research networks (centres across the country) rather than individuals or isolated research groups studying the same phenomenon should be constituted, and a phased approach to expanding the knowledge in that area of investigation should be endeavoured. An integrated approach to health care and health care research in times of fiscal restraint is integral to professional responsibility, and sound government practice in the allocation of health care monies, including research grants.

## **5.6 Conclusions**

The data from this case-control study of breast cancer indicated that the exposure data collected after the diagnosis of breast cancer are adequately reliable

(consistent) and valid when compared with the prospective exposure data collected before the subject is aware of their case-control status; Q1 was the 'gold' standard or 'criterion' for this comparison assessment.

The level of agreement or consistency between the prospective and retrospective exposure reports for the cases and controls was comparable. When group differences were observed, there was no systematic pattern of misclassification noted. Specifically, there was no tendency for the cases or control groups to overreport/overstate or underreport/understate their exposure(s) retrospectively, once their diagnosis was known. As discussed in Chapter 4, some differential and non-differential exposure misclassification was noted when the odds ratio estimates were compared. However, this occurrence was fairly isolated, and involved only a few of the study factors. Overall, there was no observed systematic shift in reported exposures by one of the study groups with respect to several of the included study factors. Thus, the data in this study provide no strong and consistent evidence for the existence of either differential or non-differential exposure misclassification. Furthermore, the estimates of effect (i.e., odds ratio estimates) have not been biased as a result of exposure misclassification; and therefore, the findings with respect to the strength of the association between the various study factors and breast cancer are valid. Consideration of the differences in the prospective and retrospective odds ratio estimates, while observing the overlap of their respective confidence intervals, would suggest that random exposure misclassification is present, and measurement error is responsible for the inconsistencies in the estimates of effect. The odds ratio estimates are subject to random measurement error rather than bias.

The use of 'anamnestic equivalents' provided no real advantages for case-control comparisons. As mentioned previously, control groups one and two did not differ with respect to the consistency of their exposure reports; the test-retest

reliability analysis, Kappa analysis, and McNemar test indicated that all three groups were comparable regarding the level of agreement or consistency between the prospective and retrospective reports of exposure. In addition, differential recall (i.e., systematic changes in the level of exposure) was absent.

It is recommended that more than one control group be used in future case-control studies, when possible, to minimize the potential for biased estimates of effect, as well as to demonstrate the consistency of the study findings (Mantel and Haenszel, 1959; Ibrahim and Spitzer, 1979).

Finally, the use of a validity scale shows significant promise as a design strategy for the measurement and control of recall bias (differential exposure misclassification). The 'validity scale' concept, originally proposed by Raphael (1987), has been implemented successfully, and a specific methodology for its construction and analysis has been established for future use and refinement. It is now important for a similar scale to be constructed, implemented and evaluated in other studies, to validate its effectiveness in measuring differential exposure misclassification (recall bias), and to improve upon its general methodology. An exposure data validity substudy would have to be done concurrently with each case-control study so as to provide a comparison baseline for the evaluation of the particular validity scale. This would mean more work for the investigator in the interim, but if the validity scale proves to be consistent and effective with respect to its stated purpose, the case-control researcher will gain the longterm benefits, that is, a design strategy that can be used to address directly the reliability of the exposure data collected, and the validity of study conclusions which rest on the completeness and accuracy of the exposure data provided by the cases and controls.

In conclusion, the results of this study demonstrated that neither non-differential or differential exposure misclassification were present to any

significant extent; nor were the odds ratio estimates biased. This study provided no evidence that case-control studies were susceptible to exposure misclassification biases which would invalidate study conclusions.

Table 24 (continued): A Comparison of the Prospective Risk Estimates of this Study with the Retrospective Estimates of Association Reported in Other Case-Control Studies

<b>Risk Factor</b>	<b>Present Study</b> Odds Ratio Estimate (95% Confidence Interval) <b>Prospective</b>	<b>Present Study</b> Odds Ratio Estimate (95% Confidence Interval) <b>Retrospective</b>	<b>Kelsey JL (1993)</b> Odds Ratio Estimate <b>Retrospective</b>	<b>Friedenreich CM (1990)</b> Odds Ratio Estimate (95% Confidence Interval) <b>Retrospective</b>
1. Marital Status (Never vs Ever)	* 1.17 (.55 - 2.49) ** 1.15 (.54 - 2.45)	* 1.28 (.53 - 3.10) ** 1.50 (.63 - 3.58)	1.1 - 2.0	1.32 (.84 - 2.06)
2. Age at Menopause ≥ 55 years (High Risk)	* 1.42 (.63 - 3.24) ** 1.17 (.53 - 2.58)	* 2.00 (.85 - 4.70) ** 1.99 (.87 - 4.55)	1.1 - 2.0	.95 (.62 - 1.54)
3. Socioeconomic Status (High vs Low)	* 0.55 (.24 - 1.33) ** 0.99 (.43 - 2.31)	* 0.52 (.23 - 1.18) ** 0.86 (.38 - 1.92)	1.1 - 2.0	-
4. Location of Residence (Urban vs Rural)	* 1.15 (.78 - 1.69) ** 1.01 (.57 - 1.24)	-	1.1 - 2.0	-
5. Family History of Breast Cancer in any First Degree Relative (i.e., mother/sister)	* 1.65 (.99 - 2.75) ** 1.10 (.68 - 1.78)	* 1.58 (.99 - 2.50) ** 1.35 (.85 - 2.14)	> 4.0	1.27 (1.00 - 1.63)
6. Family History - (mother with breast cancer)	* 1.47 (.83 - 2.60) ** 1.36 (.75 - 2.39)	* 1.51 (.90 - 2.55) ** 1.64 (.97 - 2.78)	2.1 - 4.0	1.40 (.97 - 2.03)
7. Family History - (sister with breast cancer)	* 1.58 (.82 - 3.03) ** 0.94 (.51 - 1.72)	* 1.48 (.81 - 2.70) ** 1.01 (.57 - 1.78)	2.1 - 4.0	1.32 (.79 - 2.18)
8. Pregnancy History (Ever vs Never)	* 0.96 (.61 - 1.52) ** 0.59 (.37 - 0.90)	* 1.19 (.74 - 1.93) ** 1.05 (.65 - 1.68)	-	1.39 (1.01 - 1.91)

Table 24 (continued): A Comparison of the Prospective Risk Estimates of this Study with the Retrospective Estimates of Association Reported in Other Case-Control Studies

9. Age at first full-term Pregnancy				
a. $\geq 30$ years (High Risk)	a. * 1.35 (.82 - 2.24) ** 0.82 (.49 - 1.38)	a. * 1.37 (.82 - 2.29) ** 1.27 (.76 - 2.14)	1.1 - 2.0	a. 1.18 (.67 - 2.07)
b. Nulliparous (High Risk)	b. * 1.92 (.41 - 9.07) ** 1.11 (.73 - 2.65)	b. * 2.35 (.14 - 3.89) ** 1.09 (.83 - 3.20)		b. 1.73 (1.04 - 2.89)
10. Use of oral contraceptives ( $> 5$ years)	* 0.61 (.39 - .960) ** 0.79 (.50 - 1.25)	* 0.65 (.47 - .90) ** 0.85 (.62 - 1.18)	-	0.81 (.58 - 1.13)
11. Postmenopausal Estrogen Use (Ever vs Never)	* 1.18 (.84 - 1.67) ** 0.80 (.57 - 1.12)	* 0.83 (.59 - 1.18) ** 0.59 (.42 - 0.82)	-	0.89 (.68 - 1.17)
12. Smoking History (Ever vs Never)	* 1.21 (0.87 - 1.68) ** 1.83 (1.32 - 2.54)	* 1.34 (0.96 - 1.87) ** 1.91 (1.37 - 2.66)	-	0.82 (.65 - 1.03)

Notes:

1. \* Denotes the odds ratio estimate for the case-control group one comparison.
2. \*\* Denotes the odds ratio estimate for the case-control group two comparison.

## BIBLIOGRAPHY

- Abelson RP. Beliefs are like possessions. **Journal for the Theory of Social Behavior** 1986; 16: 223-250.
- Adam SA, Sheaves JK, Wright NH, et al. A case-control study of the possible association between oral contraceptives and malignant melanoma. **Br J Cancer** 1981; 44: 45-49.
- Adami H-O, Lund E, Bergstrom R, Meirik O. Cigarette smoking, alcohol consumption and risk of breast cancer in young women. **Br J Cancer** 1988; 58: 832-837.
- Adams RM, Smelseer NJ, Treiman D (eds). **Behavioral and Social Sciences Research: A National Resource, Part I**. Washington: National Academy Press, 1982.
- Alba JW, Hasher L. Is memory schematic? **Psychological Bulletin** 1983; 93: 203-231.
- Anderson RC, Pichert JW. Recall of previously unrecalable information following a shift in perspective. **Journal of Verbal Learning and Verbal Behavior** 1978; 17: 1-12.
- Armitage P, Berry G. **Statistical Methods in Medical Research**. Oxford: Blackwell Scientific Publications, 1971, pp.123-125.
- Armstrong JS, Denniston WB, Gordon MM. The use of the decomposition principle in making judgments. **Organizational Behavior and Human Performance** 1975; 14: 257-263.
- Austin H, Hill HA, Flanders D, Greenberg RS. Limitations in the Application of Case-Control Methodology. **Epidemiologic Reviews** 1994; 16: 65-75.
- Austin MA, Criqui MH, Barrett-Connor E, Holdbrook MJ. The effect of response bias on the odds ratio. **Am J Epidemiol** 1981; 114: 137-143.
- Axelsson A. The case-control study: Valid selection of subjects. **J Chron Dis** 1985; 38: 553-558.
- Axelsson G, Rylander R. Validation of questionnaire-reported miscarriage, malformation and birth weight. **Int J Epidemiol** 1984; 13: 94-98.
- Babbie ER. **Survey Research Methods**. Belmont, CA: Wadsworth, 1973.
- Baddeley A. The limitations of human memory: Implications for the design of retrospective surveys. In L Moss, H Goldstein (eds), **The Recall Method in Social Surveys**. University of London: Institute of Education, 1979, pp.13-27.
- Bahrack HP. The accuracy of recalling information from long-term memory. In GH Bower (ed), **Psychology of Learning and Motivation**. New York: Academic Press, 1983.

- Bailey KD. **Methods of Social Research**. (Third Edition). New York: The Free Press, 1987.
- Barron BA. The effects of misclassification on the estimation of relative risk. **Biometrics** 1977; 26: 414-418.
- Barsalou LW. In U Neisser (ed), **Real Events Remembered: Ecological Approaches to the Study of Memory**. Cambridge University, England: Cambridge University Press, 1987.
- Bartlett FC. **Remembering: A Study in Experimental and Social Psychology**. Cambridge, England: Cambridge University Press, 1932.
- Baumgarten M, Siemiatycki J, Gibbs GW. Validity of work histories obtained by interview for epidemiologic purposes. **Am J Epidemiol** 1983; 118: 583-591.
- BC Cancer Agency, BC Cancer Foundation, BC Cancer Research Centre. **Annual Report 1993-1994**. Vancouver, 1994.
- Bean JA, Leeper JD, Wallace RB, Sherman BM, Jagger H. Variations in the reporting of menstrual histories. **Am J Epidemiol** 1979; 109: 181-185.
- Beard CM, Yunginger JW, Reed CE, et al. Interobserver variability in medical record review: An epidemiological study of asthma. **J Clin Epidemiol** 1992; 45: 1013-1020.
- Benzecri J-P. **Cours de Linguistique Mathématique**. Universite de Rennes, France, 1963.
- Benzecri J-P. Statistical Analysis as a Tool to Make Patterns Emerge from Data. In **Methodologies of Pattern Recognition**, S. Watanabe (Ed). New York: Academic Press, Inc., 1969, pp. 35-74.
- Benzécri J-P. **L'Analyse des Données. Vol. I, La Taxinomie**. Paris: Dunod, 1973a.
- Benzécri J-P. **L'Analyse des Données. Vol. II, L'Analyse des Correspondances**. Paris: Dunod, 1973b.
- Beresford SAA, Coker AL. Pictorially assisted recall of past hormone use in case-control studies. **Am J Epidemiol** 1989; 130: 202-205.
- Bisch L, Lee JI, Mark E. Major Causes of Death, Canada, 1989. **Chronic Diseases in Canada** 1989; 10:22-24.
- Black JB, Bower GH. Episodes as chunks in narrative memory. **Journal of Verbal Learning and Verbal Behavior** 1979; 18: 309-318.
- Blackburn H. (Commentary) Observation versus Experiment. **Statistics in Medicine** 1984; 3: 401-403.
- Blair A, Hoar Zahm S. Methodologic issues in exposure assessment for case-control studies of cancer and herbicides. **Am J Indust Med** 1990; 18: 285-293.

- Blair SN, Dowda M, Pate RR, et al. Reliability of long-term recall of participation in physical activity by middle-aged men and women. **Am J Epidemiol** 1991; 133: 266-275.
- Blaney PH. Affect and memory: A review. **Psychol Bull** 1986; 99: 229-246.
- BMDP Statistical Software, Inc. **BMDP Statistical Software Manual, Volume 2**, WJ Dixon, (ed). Los Angeles, CA: BMDP Statistical Software, Inc, 1992.
- Bower GH, Black JB, Turner TJ. Scripts in memory for text. **Cognitive Psychology** 1979; 11: 177-220.
- Boyle P, Leake R. (Review) Progress in understanding breast cancer: Epidemiological and biological interactions. **Breast Cancer Research and Treatment** 1988; 11: 91-112.
- Bracken MB, Vita K. Frequency of non-hormonal contraception around conception and association with congenital malformations in offspring. **Am J Epidemiol** 1983; 117: 281-291.
- Bracken MB. Methodologic issues in the epidemiologic investigation of drug induced congenital malformations. In MB Bracken (ed), **Perinatal Epidemiology**. New York: Oxford University Press, 1984, pp.436-440.
- Bradburn NM, Miles C. Vague quantifiers. **Public Opinion Quarterly** 1979; 43: 92-101.
- Bradburn NM, Sudman S. **Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions**. San Francisco: Jossey-Bass, 1979.
- Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: The impact of memory and inference on surveys. **Science** 1987; 239: 157-161.
- Bradley B, Matthews A. A negative self-schemata in clinical depression. **Br J Clin Psychol** 1983; 22: 173-181.
- Brandt LPA, Nielsen CV. Job stress and adverse outcome of pregnancy: A causal link or recall bias. **Am J Epidemiol** 1992; 135: 302-311.
- Bransford JD, Johnson MK. Considerations of some problems of comprehension. In WG Chase (ed), **Verbal Information Processing**. New York: Academic Press, 1973.
- Bransford JD, Nitsch KE. Coming to understand the things we could not previously understand. In JF Kavanagh, W Strange (eds), **Speech and Language in the Laboratory, School and Clinic**. Cambridge, Mass: MIT Press, 1978.
- Brenner H. Use and limitations of dual measurements in correcting for nondifferential exposure misclassification. **Epidemiology** 1992; 3: 216-222.

- Brenner H, Blettner M. Misclassification bias arising from random error in exposure measurement: Implications for dual measurement strategies. *Am J Epidemiol* 1993; 138:453-461.
- Breslow NE, Day NE, Halvorsen RT, Prentice RL, Sabai C. Estimation of multiple relative risk functions in matched case-control studies. *Am J Epidemiol* 1978; 108: 299-307.
- Breslow NE, Day NE. **Statistical Methods in Cancer Research (Volume 1 - The analysis of case-control studies)**. Lyon, France: International Agency for Research on Cancer, 1980, pp.91-126.
- Breslow N. Design and analysis of case-control studies. *Ann Rev Public Health* 1982; 3: 29-54.
- British Columbia Cancer Registry. **British Columbia Cancer Statistics 1992**. Vancouver, 1993, p.2.
- Bross I. Misclassification in 2x2 tables. *Biometrics*, 1954; 10: 478-86.
- Brown NR, Shevell SK, Ripps LJ. The subjective dates of natural events in very long-term memory. *Cognitive Psychology* 1985; 17: 139-177.
- Bryant HE, Love EJ. Video display terminal use and spontaneous abortion risk. *Int J Epidemiol* 1989; 18: 132-138.
- Bryant HE, Visser N, Love EJ. Records, recall loss and recall bias in pregnancy: A comparison of interview and medical records data of pregnant and postnatal women. *Am J Public Health* 1989; 79: 78-80.
- Buell P, Dunn JE. The dilution effect of misclassification. *AJPH* 1964; 54: 598-602.
- Byers TE, Rosenthal RI, Marshall JR, et al. Dietary history from the distant past: A methodological study. *Nutr Cancer* 1983; 5: 69-77.
- Byers TE, Marshall JR, Anthony E, et al. The reliability of dietary history from the distant past. *Am J Epidemiol* 1987; 125: 999-1011.
- Byrt T, Bishop J, Carlin J. Bias, prevalence and kappa. *J Clin Epidemiol* 1993; 46: 423-429.
- Caan B, Hiatt RA, Owen AM. Mailed dietary surveys: Response rates and the effect of omitted food items on nutrient values. *Epidemiology* 1991; 2: 430-436.
- Cannell CF, Kahn RL. Interviewing. In G Lindzey, E Aronson (eds), **The Handbook of Social Psychology (Volume 2): Research Methods**. (Second Edition). Reading, Mass: Addison-Wesley, 1968, pp. 526-595.
- Cannell CF, Henson R. Incentives, motives and response bias. *Annals of Economic and Social Measurement* 1974; 32: 307-317.
- Cannell CF, Oksenberg L, Converse JM. Striving for response accuracy: Experiments in new interviewing techniques. *Journal of Marketing Research* 1977; 14: 306-315.

- Cannell CF. Reporting of health events in household interviews: Data evaluation and methods research: A summary of Studies of interviewing methodology. **Vital and Health Statistics**. PHS Pub. No. 1000-Series 2-No. 69. DHEW Pub No. (HRA) 77-1342 and 77-1343. Health Resources Administration - National Center for Health Statistics. Washington. U.S. Government Printing Office, March 1977.
- Cannell CF, Miller PV, Oksenberg L. Research on interviewing techniques. In S Leinhardt (ed), **Sociological Methodology**. San Francisco: Josey-Bass, 1981.
- Carmines EG, Zeller RA. **Reliability and Validity Assessment**. Beverly Hills, London: Sage Publications, 1979.
- Carr LG. The Srole items and acquiescence. **American Psychological Review** 1971; 36: 287-293.
- Chambers LW, Spitzer WO. A method of estimating risk for occupational factors using multiple data sources: The Newfoundland lip cancer study. **Am J Public Health** 1977; 67: 176-179.
- Checkoway H, Pearce NE, Crawford-Brown DJ. **Research Methods in Occupational Epidemiology**. New York: Oxford University Press, 1989.
- Chouinard E, Walter S. Recall bias in case-control studies: An empirical analysis and theoretical framework. **J Clin Epidemiol** 1995; 48:245-254.
- Chu SY, Lee NC, Wingo PA, Webster LA. Alcohol consumption and the risk of breast cancer. **Am J Epidemiol** 1989; 130: 867-877.
- Cicchetti DV. Assessing inter-rater reliability for rating scales: Resolving some basic issues. **Br J Psychiat** 1976; 129: 452-456.
- Clark HH, Schreuder R, Buttrick S. Common ground and the understanding of demonstrative reference. **Journal of Verbal Learning and Verbal Behavior** 1983; 22: 245-258.
- Clarke M, Clayton D. The design and interpretation of case-control studies of perinatal mortality. **Am J Epidemiol** 1981; 113: 636-645.
- Clayton D. Using test-retest reliability data to improve estimates of relative risk: An application of latent class analysis. **Statistics in Medicine**. 1985; 4: 445-455.
- Cobb S, Cannell CF. Some thoughts about interview data. **Int J Epidemiol Assoc Bulletin** 1966; 13: 43-54.
- Cochran WG. Matching in analytical studies. **AJPH** 1953; 43: 684-691.
- Cochran WG. The planning of observational studies of human populations. **J Royal Statist Soc** 1965; 128: 234-265.

- Codero SF, Oakley GP, Greenberg F, et al. Is Bendectin a teratogen? **JAMA** 1981; 245: 2307-2310.
- Cohen J. A coefficient of agreement for nominal scales. **Educ Psychol Meas** 1960; 20: 37-46.
- Colditz GA, Martin P, Stampfer MJ, Willett WC, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. **Am J Epidemiol** 1986; 123: 894-900.
- Colditz GA, Stampfer MJ, Willett WC, et al. Reproducibility and validity of self-reported menopausal status in a prospective cohort study. **Am J Epidemiol** 1987; 126: 319-325.
- Cole P. The evolving case-control study. **J Chron Dis** 1979; 32: 15-27.
- Collins M, Butcher B. Interviewer and clustering effects in attitude surveys. **Journal of the Market Research Society** 1982; 25: 221-228.
- Colton T. **Statistics in Medicine**. Boston: Little Brown and Company, 1974.
- Converse PE. The nature of belief systems in the mass public. In D Apter (ed), **Ideology and Discontent**. New York: Free Press, 1964.
- Converse PE. Attitudes and nonattitudes: Continuation of a dialogue. In ER Tufte (ed), **The Quantitative Analysis of Social Problems**. Reading, MA: Addison-Wesley, 1970.
- Converse PE, Traugott, MW. Assessing the accuracy of polls and surveys. **Science** 1986; 234: 1094-1098.
- Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. **Am J Epidemiol** 1977; 105: 488-495.
- Cornfield J, Haenszel W. Some aspects of retrospective studies. **J Chron Dis** 1960; 11: 523-534.
- Cornfield J. A method for estimating comparative rates for clinical data. **JNCI** 1950-1951; 11: 1269-1275.
- Cornfield JA statistical problem arising from retrospective studies. In J Neyman (ed), **Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley: University of California Press, 1956, pp. 135-148.
- Corwin RG, Krober M, Roth HP. Patients' accuracy in reporting their past medical history: A study of 90 patients with peptic ulcer. **J Chronic Dis** 1971; 23: 875-879.
- Couch A, Keniston K. Yeasayers and naysayers: Agreeing response set as a personality variable. **Journal of Abnormal and Social Psychology** 1960; 60: 151-174.

- Coughlin SS. Recall bias in epidemiologic studies. **J Clin Epidemiol** 1990; 43: 87-91.
- Coughlin SS. (Letter) Recall bias in epidemiologic studies. **J Clin Epidemiol** 1990; 43: 1431-1433.
- Coughlin SS, Beauchamp TL. Ethics, scientific validity, and the design of epidemiologic studies. **Epidemiology** 1992; 3: 343-347.
- Coulter A, McPherson K, Elliott S, et al. Accuracy of recall of surgical histories: A comparison of postal survey data and general practice records. **Community Med** 1985; 7: 186-189.
- Coulter A, Vessey M, McPherson K. The ability of women to recall their oral contraceptive histories. **Contraception** 1986; 33: 127-137.
- Craik FIM. Age Differences in human memory. In J Birren, KW Shore (eds), **Handbook of Psychology of Aging**. New York: Van Nostrand Reinhold, 1977.
- Crawford A. Bias in the survey of drinking habits. **Alcohol and Alcoholism** 1987; 22: 167-179.
- Crichton NJ, Hinde JP. Correspondence analysis as a screening method for indicants for clinical diagnosis. **Statistics in Medicine** 1989; 8: 1351-1362.
- Crichton NJ, Hinde JP. Correspondence analysis and Independent Bayes for clinical diagnosis. In F Dunstan, J Pickles (eds), **Statistics in Medicine**. Oxford: Clarendon Press, 1991, pp.234-247.
- Criqui, MH. Response bias and risk ratios in epidemiologic studies. **Am J Epidemiol** 1979; 109: 394-399.
- Damon A, Bajema CJ. Age at menarche: Accuracy of recall after thirty-nine years, **Hum Biol** 1974; 46: 381-384.
- Davoli M, Perucci CA, Sangalli M, et al. Reliability of sexual behavior data among high school students in Rome. **Epidemiology** 1992; 3: 531-535.
- Dawes RM. **Rational Choice in an Uncertain World**. San Diego: Harcourt, Brace, Jovanovich, 1988.
- Dawson, DA. Breast cancer risk factors. **Vital and Health Statistics**. National Center for Health Statistics, series 10; 1987.
- Day Baird D, Weinberg CR, Rowland AS. Reporting errors in time-to-pregnancy data collected with a short questionnaire. **Am J Epidemiol** 1991; 133: 1282-1290.
- De Jong PCM, Huijsmans AA, Nienhuis HE, et al. Validation of a questionnaire on medical drug use during pregnancy. **Am J Epidemiol** 1991; 134: 998-1002.
- De Waard F, Trichopoulos D. A unifying concept of the aetiology of breast cancer. **Int J Cancer** 1988; 41: 666-669.

- DeMaio TJ. Social desirability and survey measurement: A review. In CF Turner, E Martin (eds), **Surveying Subjective Phenomena**. New York: Russell Sage Foundation, 1984, pp. 257-282.
- Diamond EL, Lillienfeld AM. Effects of errors in classification and diagnosis in various types of epidemiological studies. **AJPH** 1962; 52: 1137-1144.
- Diamond EL, Lillienfeld AM. Misclassification errors in 2X2 tables with one margin fixed: Some further comments. **AJPH** 1962; 52: 2107-2110.
- Dillman D. **Mail and Telephone Surveys**. New York: John Wiley and Sons, 1978.
- Dorn HF. Some applications of biometry in the collection and evaluation of medical data. **J Chron Dis** 1959a; 1: 638-665.
- Dorn HF. Some problems arising in prospective and retrospective studies of the etiology of disease. **N Engl J Med** 1959b; 261: 571-579.
- Drasar BS, Irving D. Environmental factors and cancer of the colon and breast. **Br J Cancer** 1973; 27: 167-172.
- Drews CD, Kraus JF, Greenland S. Recall bias in a case-control study of sudden infant death syndrome. **Int J Epidemiol** 1990a; 19: 405-411.
- Drews CD, Greenland S. The impact of differential recall on the results of case-control studies. **Int J Epidemiol** 1990b; 19: 1107-1112.
- Duncan GJ, Hill, DH. An investigation of the extent and consequences of measurement error in labor economic survey data. **Journal of Labor Economics** 1985; 3: 508-532.
- Dunn B, Hislop TG, Anthony V. **Breast Cancer Risk and Prognosis**. (Handout - BCCA).
- Dwyer JT, Krall EA, Coleman KA. The problem of memory in nutritional epidemiology research. **J Am Diet Assoc** 1987; 87: 1509-1512.
- Edwards AL. **The Social Desirability Variable in Personality Assessment and Research**. New York: Dryden, 1957.
- Eliasziw M, Donner A. A cost-function approach to the design of reliability studies. **Statistics in Medicine** 1987; 6: 647-655.
- Elwood JM. **Causal Relationships in Medicine: A practical system for critical appraisal**. Oxford: Oxford University Press, 1988.
- Ericson A, Kallen B. An epidemiological study of work with video screens and pregnancy outcome: II. A case-control study. **Am J Ind Med** 1986; 9: 459-475.
- Evans SJW. Uses and abuses of multivariate methods in epidemiology. **J of Epidemiol and Community Health** 1988; 42: 311-315.

- Feinstein AR. The epidemiologic trohoc, the ablative risk ratio, and 'retrospective' research. **Clin Pharmacol Ther** 1973; 14: 291-307.
- Feinstein AR. Methodologic problems and standards in case-control research. **J Chron Dis** 1979a; 32: 35-41.
- Feinstein AR. Scientific standards vs. statistical associations and biologic logic in the analysis of causation. **Clin Pharmacol Ther** 1979b; 20: 481-492.
- Feinstein AR, Horwitz RI, Spitzer WO, Battista RN. Coffee and pancreatic cancer: The problems of etiologic science and epidemiologic case-control research. **JAMA** 1981; 246: 957-961.
- Feinstein AR, Horwitz RI. Double standards, scientific methods, and epidemiologic research. **N Engl J Med** 1982; 307: 1611-1617.
- Feinstein AR. **Clinical Epidemiology: The architecture of clinical research**. Philadelphia: Saunders, 1985a.
- Feinstein AR. Experimental requirements and scientific principles in case-control studies. **J Chron Dis** 1985b; 38: 127-133.
- Feinstein AR. Scientific standards and epidemiologic methods. **Am J Clin Nutr** 1987; 45: 1080-1088.
- Feinstein AR. (Editorial) Directionality in epidemiologic research. **J Clin Epidemiol** 1988; 41: 705-707.
- Feinstein AR. Directionality and scientific inference. **J Clin Epidemiol** 1989; 42: 829-833.
- Feinstein AR, Cicchetti DV. High agreement but low kappa: The problems of two paradoxes. **J Clin Epidemiol** 1990; 43: 543-549.
- Feldman Y, Koren G, Mattice D, Shear H, et al. Determinants of recall and recall bias in studying drug and chemical exposure in pregnancy. **Teratology** 1989; 40: 37-45.
- Fenster L, Swan SH, Windham GC, Neutra RR. Assessment of reporting consistency in a case-control study of spontaneous abortion. **Am J Epidemiol** 1991; 133: 477-488.
- Fenster L, Windham GC, Swan SH, et al. Tap or bottled water consumption and spontaneous abortion in a case-control study of reporting consistency. **Epidemiology** 1992; 3: 120-124.
- Fienberg SE, Tanur JM. Large scale social surveys: Perspectives, problems and prospects. **Behavioral Science** 1983; 28: 135-153.
- Fienberg SE, Loftus EF, Tanur JM. Cognitive aspects of health survey methodology. **Millbank Memorial Fund Quarterly - Health and Society** 1985; 63: 547-597.

- Fischhoff B. Perceived informativeness of facts. **Journal of Experimental Psychology: Human Perception and Performance** 1977; 3: 349-358.
- Flanders WD, Drews CD, Kosinski MS. Methodology to correct for differential misclassification. **Epidemiology** 1995; 6:152-156/
- Fleiss JL. Confidence intervals for the odds ratio in case-control studies: The state of the art. **J Chron Dis** 1979; 32: 69-77.
- Fleiss JL. **Statistical methods for rates and proportions**. (Second Edition). New York: John Wiley and Sons, 1981: 212-225.
- Floderus B, Barlow L, Mack TM. Recall bias in subjective reports of familial cancer. **Epidemiology** 1990; 4: 318-321.
- Forsyth BW, Horwitz RI, Acampora D, et al. New epidemiologic evidence confirming that bias does not explain the aspirin/Reye's syndrome association. **JAMA** 1989; 261: 2517-2524.
- Fowler FJ, Mangione TW. The value of interview training and supervision. **Public Opinion Quarterly** 1970; 34: 621-624.
- Freeman LC, Romney AK, Freeman SC. Cognitive structure and informant accuracy. **American Anthropologist** 1987; 89: 310-325.
- Friedenreich CM. (Thesis) **Recall Bias in the Association of Diet and Breast Cancer**. Toronto: University of Toronto, 1990.
- Friedenreich CM, Howe GR, Miller AB. An investigation of recall bias in the reporting of past food intake among breast cancer cases and controls. **Annals of Epidemiology** 1991a; 1: 439-453.
- Friedenreich CM, Howe GR, Miller AB. The effect of recall bias on the association of calorie providing nutrients and breast cancer. **Epidemiology** 1991b; 2: 424-429.
- Friedenreich CM, Howe GR, Miller AB. Recall bias in association of micronutrient intake and breast cancer. **J Clin Epidemiol** 1993; 46: 1009-1017.
- Friedenreich CM. (Editorial) Improving long-term recall in epidemiologic studies. **Epidemiology** 1994; 5: 1-4.
- Frith CD et al. Effects of ECT and depression on various aspects of memory. **Br J Psychiat** 1983; 142: 173-181.
- Fugate Woods N, Most A, Kramer Derry G. Estimating perimenstrual distress: A comparison of two methods. **Research in Nursing and Health** 1982; 5: 81-91.
- Furnham A. Response bias, social desirability and dissimulation. **Person Individ Diff** 1986; 7: 385-400.
- Gammon MD, Thompson WD. Polycystic ovaries and the risk of breast cancer. **Am J Epidemiol** 1991; 134: 818-824.

- Gauld A, Stephenson GM. Some experiments relating to Bartlett's theory of remembering. **British Journal of Psychology** 1967; 58: 39-49.
- Gehlbach SH. Study design: The case-control approach. In **Interpreting the Medical Literature**. Lexington, Mass: DC Heath, 1982, pp.39-54.
- Gentner D, Loftus EF. Integration of verbal and visual information as evidenced by distortions in picture memory. **American Journal of Psychology** 1979; 92: 366-375.
- Gergen KJ, Back GW. Communication in the interview and the disengaged respondent. **Public Opinion Quarterly** 1969; 33: 385-398.
- Giovannucci E, Stampfer MJ, Colditz GA, et al. A comparison of prospective and retrospective assessments of diet in the study of breast cancer. **Am J Epidemiol** 1993; 137: 502-511.
- Gladen B, Rogan WJ. Misclassification and the design of environmental studies. **Am J Epidemiol** 1980; 109: 607-616.
- Glass R, Johnson B, Vessey M. Accuracy of recall of histories of oral contraceptive use. **Br J Prev Soc Med** 1974; 28: 273-275.
- Glenn NG. Aging, disengagement and opinionation. **Public Opinion Quarterly**. 1969; 1: 17-33.
- Goldberg JD. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. **J Am Statist Assoc** 1975; 70: 561-570.
- Goldberg MS, Semiatycki J, Gerin M. Inter-rater agreement in assessing occupational exposure in a case-control study. **Br J Indust Med** 1986; 43: 667-676.
- Goldberg RJ, Pastides R, Ellison C, et al. Uses of the case-control and cohort epidemiological approaches in pediatric practice and research. **Ped Res** 1985; 19: 787-790.
- Goodman MT, Nomura AMY, Wilkens LR, Kolonel LN. Agreement between interview information and physician records of history of menopausal estrogen use. **Am J Epidemiol** 1990; 131: 815-825.
- Gordis L. Assuring the quality of questionnaire data in epidemiologic research. **Am J Epidemiol** 1979; 109: 21-24.
- Gove W, Geerken M. Response bias in surveys of mental health: An empirical investigation. **American Journal of Sociology** 1977; 82: 1289-1317.
- Graham S, Marshall J, Mettlin C, et al. Diet in the epidemiology of breast cancer. **Am J Epidemiol** 1982; 116: 68-75.

- Graham S, Hellmann R, Marshall J, et al. Nutritional epidemiology of postmenopausal breast cancer in western New York. *Am J Epidemiol* 1991; 134:552-565.
- Granroth G. Defects of the central nervous system in Finland: III- Diseases and drugs in pregnancy. *Early Hum Dev* 1978; 2: 147-162.
- Gray-Donald K, Kramer MS. Causality inference in observational versus experimental studies. *Am J Epidemiol* 1988; 127: 885-892.
- Greenacre MJ. **Theory and Applications of Correspondence Analysis**. London: Academic Press, 1984.
- Greenacre MJ. **Correspondence Analysis in Practice**. London: Academic Press, 1993.
- Greenacre MJ, Blasius J. **Correspondence Analysis in the Social Sciences: Recent Developments and Applications**. London: Academic Press, 1994.
- Greenberg E, Rosner B, Hennekens C, et al. An investigation of bias in a study of nuclear shipyard workers. *Am J Epidemiol* 1985; 121: 301-308.
- Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol* 1977; 106: 184-187.
- Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980; 112: 564-569.
- Greenland S, Criqui MH. Are case-control studies more vulnerable to response bias? *Am J Epidemiol* 1981; 114: 175-177.
- Greenland S, Morgenstern H, Thomas DC. Considerations in determining matching criteria and stratum sizes for case-control studies. *Int J Epidemiol* 1981; 10: 389-392.
- Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985; 122: 495-506.
- Greenland S, Robins JM. Identifiability, exchangeability and epidemiological confounding. *Int J Epidemiol* 1986; 15: 412-418.
- Greenland S. Statistical uncertainty due to misclassification: Implications for validation substudies. *J Clin Epidemiol* 1988a; 41: 1167-1174.
- Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. *Statistics in Medicine* 1988b; 7: 745-747.
- Greenland S, Morgenstern H. Classification schemes for epidemiologic research designs. *J Clin Epidemiol* 1988; 41: 715-716.
- Gregorio DI, Marshall JR, Zielezny M. Fluctuation in odds ratios due to variance differences in case-control studies. *Am J Epidemiol* 1985; 121: 767-774.

- Griching WL, Caltabiano ML. Amount and direction of bias in survey interviewing. **Austr Psychol** 1986; 21: 69-78.
- Gross PA. Collection of data documentation risk factors: Safeguards in conducting case-control studies. **Am J Med** 1984; 76: 28-33.
- Groves RM, Kahn RL. **Surveys by Telephone: A National Comparison with Interviews**. New York: Academic Press, 1979.
- Groves RM. **Survey Errors and Survey Costs**. New York: John Wiley and Sons, 1989.
- Gullen WH, Bearman JE, Johnson EA. Effects of misclassification in epidemiologic studies. **Public Health Reports** 1968; 83: 915-918.
- Hammar N, Norell SE. Retrospective versus original information on diet among cases of colorectal cancer and controls. **Int J Epidemiol** 1991; 20: 621-627.
- Harlow SD, Linet MS. Agreement between questionnaire data and medical records: The evidence for accuracy of recall. **Am J Epidemiol** 1989; 129: 233-248.
- Harper D. Misclassification in epidemiologic surveys. **AJPH** 1964; 54: 1882-1886.
- Harris NV, Weiss NS, Francis AM, Polissar L. Breast cancer in relation to patterns of oral contraceptive use. **Am J Epidemiol** 1982; 116: 643-651.
- Harris RJ, Monaco GE. Psychology of pragmatic implication: Information processing between the lines. **Journal of Experimental Psychology** 1978; 107: 1-22.
- Hasher L, Griffin M. Reconstructive and reproductive processes in memory. **Journal of Experimental Psychology** 1978; 4: 318-330.
- Hastie R, Carlston D (eds). **Person Memory: the Cognitive Basis of Social Perception**. Hillsdale NJ: L. Erlbaum, 1980.
- Hauser P. (Comments on Coleman's paper) In R. Bierstedt (ed), **A Design for Sociology: Scope, Objectives and Methods**. Philadelphia: American Academy of Political and Social Science, 1969.
- Hayden GF, Kramer MS, Horwitz RI. The case-control study: A practical review for the clinician. **JAMA** 1982; 247: 326-331.
- Heidam LZ, Olsen J. Self-reported data on spontaneous abortion compared with data obtained by computer linkage with the hospital registry. **Scand J Soc Med** 1985; 13: 159-163.
- Helmrich SP, Shapiro S, Rosenberg L, Kaufman DW, et al. Risk factors for breast cancer. **Am J Epidemiol** 1983; 117: 35-45.
- Henneckens CH, Buring JE. **Epidemiology in Medicine**. Boston: Little, Brown and Company, 1987.

- Hermann N. Retrospective information from questionnaires. Comparability of primary respondents and their next-of-kin. **Am J Epidemiol** 1985; 121: 937-947.
- Hertz-Picciotto I, Swan SH, Neutra RN, et al. Spontaneous abortions in relation to consumption of tap water: An application of methods from survival analysis to a pregnancy follow-up study. **Am J Epidemiol** 1989; 130: 79-93.
- Hertz-Picciotto, I. Re: "Agreement between questionnaire data and medical records: The evidence for accuracy of recall". (Letter) **Am J Epidemiol** 1991; 133: 408-409.
- Hertz-Picciotto I, Swan SH, Neutra RR. Reporting bias and mode of interview in a study of adverse pregnancy outcomes and water consumption. **Epidemiology** 1992; 3: 104-112.
- Hertzman C, Mayes M, Singer J, et al. Upper Ottawa Street landfill site health study. **Environ Health Perspect** 1987; 75: 173-195.
- Herzog AR, Rodgers WL. Age and response rates to interview sample surveys. **Journal of Gerontology: Social Sciences** 1988; 43: S200-205.
- Herzog AR, Rodgers WL. Age differences in memory performance and memory ratings in a sample survey. **Psychology of Aging** 1989; 65: 88-90.
- Hewitt D, Sanders B, Stewart A. Oxford Survey of Childhood Cancers: Progress Report IV - Reliability of Data Reported by Case and Control Mothers. **Mon Bull Minist Hlth** 1966; 25: 80-85.
- Hewson D, Bennett A. Childbirth research data: Medical records or women's reports? **Am J Epidemiol** 1987; 125: 484-491.
- Hill AB. Observation and experiment. **N Engl J Med** 1953; 248: 995.
- Hislop TG, Lamb CW, Ng VTY. Differential misclassification bias and dietary recall for the distant past using a food frequency questionnaire. **Nutrition and Cancer** 1990; 13: 223-233.
- Hoekelman RA, Kelly J, Zimmer AW. The reliability of maternal recall: Mothers' remembrance of their infant's health and illness. **Clinical Pediatrics** 1976; 15: 261-265.
- Hoffman DL, Franke GR. Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research. **Journal of Marketing Research** 1986; 23: 213-227.
- Hogue CJ. Low birth-weight subsequent to induced abortion: A historical prospective study in Skopje Yugoslavia. **Am J Obstet Gynecol** 1975; 123: 675-681.
- Hogue CJ. The effect of common exposures on reproductive outcomes. **Teratogenesis, Carcinogenesis and Mutagenesis** 1984; 4: 45-57.

- Hopwood DC, Guidotti TL. Recall bias in exposed subjects following a toxic exposure incident. **Archives of Environmental Health** 1988; 43: 234-237.
- Hornsby PP, Wilcox AJ. Validity of questionnaire information on frequency of coitus. **Am J Epidemiol** 1989; 130: 94-99.
- Horwitz O, Lysgaard-Hansen B. Medical observation and bias. **Am J Epidemiol** 1975; 101: 391-399.
- Horwitz RI, Feinstein AR. Post trohoc ergo propter trohoc: problems, conflicting results and criteria for scientific standards in retrospective 'case-control' research. **Clin Res** 1976; 24: 248A.
- Horwitz RI, Feinstein AR. Alternate analytic methods for case-control studies of estrogens and endometrial cancer. **N Engl J Med** 1978; 299: 1089-1094.
- Horwitz RI, Feinstein AR. Methodologic standards and contradictory results in case-control research. **Am J Med** 1979; 66: 556-564.
- Horwitz RI, Feinstein AR, Stremlau JR. Alternative data sources and discrepant results in case-control studies of estrogens and endometrial cancer. **Am J Epidemiol** 1980; 111: 389-394.
- Horwitz RI, Feinstein AR. The application of therapeutic trial principles to improve the design of epidemiologic research: A case-control study suggesting that anticoagulants reduce mortality in patients with myocardial infarction. **J Chron Dis** 1981; 34: 575-583.
- Horwitz RI, Yu EC. Assessing the reliability of epidemiologic data obtained from medical records. **J Chron Dis** 1984; 37: 825-831.
- Horwitz RI. Comparison of epidemiologic data from multiple sources. **J Chron Dis** 1986; 39: 889-896.
- Horwitz RI. The experimental paradigm and observational studies of cause-effect relationships in clinical medicine. **J Chron Dis** 1987; 40: 91-99.
- Howe GR. The use of polytomous dual response data to increase power in case-control studies: An application to the association between dietary fat and breast cancer. **J Chron Dis** 1985; 38: 663-670.
- Howe HL, Hoff MB. Breast self-examination and breast cancer: A note on postdisease reporting bias. **Cancer Detection and Prevention** 1983; 6: 473-483.
- Hsieh CH, Walter S. The effect of non-differential exposure misclassification on estimates of the attributable and prevented fraction. **Statistics in Medicine** 1988; 7: 1073-1085.
- Hulika IM. Age differences in retention as a function of interference. **Journal of Gerontology** 1967; 22: 180-184.

- Hulka BS, Hogue CJ, Greenberg BG. (Reviews and Commentary) Methodological issues in epidemiologic studies of endometrial cancer and exogenous estrogen. **Am J Epidemiol** 1978; 107: 267-276.
- Hulka BS, Kupper LL, Cassell JC, et al. Medication use and misuse: Physician-patient discrepancies. **J Chron Dis** 1979; 28: 7-21.
- Hulka BS, Grimson RC, Greenberg BG, et al. "Alternative" controls in a case-control study of endometrial cancer and exogenous estrogens. **Am J Epidemiol** 1980; 112: 376-387.
- Huttly SR, Barros FC, Victoria CG, Beira JV, et al. Do mothers overestimate breast feeding duration? An example of recall bias from a study in Southern Brazil. **Am J Epidemiol** 1990; 132: 572-575.
- Hyman HH. **Interviewing in Social Research**. Chicago: University Chicago Press, 1954.
- Ibrahim MA, Spitzer WO. The case-control study: The problem and the prospect. **J Chron Dis** 1979; 32: 139-144.
- Ibrahim MA (ed). **The case-control study: Consensus and controversy**. Oxford: Pergamon Press, 1979.
- Imperiale TF, Horwitz RI. Scientific standards and the design of case-control research. **Biomed and Pharmacother** 1989; 43: 187-196.
- Jabine TB, Straf ML, Tanur JM, Tourangeau R (eds). **Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines**. Washington: Academy Press, 1984.
- Jaeger CM, Pennock JL. An analysis of consistency of response in household surveys. **Am Statistical Assoc J** 1961; 320-327.
- Jain M, Cook GM, Davis FG, et al. A case-control study of diet and colorectal cancer. **Int J Cancer** 1980; 26: 757-768.
- Jain M, Howe GR, Harrison L, Miller AB. A study of repeatability of dietary data over a seven year period. **Am J Epidemiol** 1989; 129: 422-429.
- Janerich DT, Glebatis D, Flink E, Hoff MB. Case-control studies on the effect of sex steroids on women and their offspring. **J Chron Dis** 1979; 32: 83-88.
- Janerich DT, Polednak AP. Epidemiology of birth defects. **Epidemiologic Review** 1983; 5: 16-37.
- Jemain AA. (Thesis) **Mathematical models for evaluating and planning screening programmes and for correcting biases resulting from case-control studies based on screening data**. Reading: University of Reading, 1988.
- Jennings BM, Rogers S. Managing measurement error. **Nursing Research** 1989; 38: 186-187.

- Jick H, Vessey MP. Case-control studies in the evaluation of drug-induced illness. **Am J Epidemiol** 1978; 107: 1-7.
- Jick H. (Letter) Recall "error" in interview studies of past drug use. **Am J Public Health** 1982; 72: 405.
- Jobe JB, Mingay DJ. Cognitive research improves questionnaires. **AJPH** 1989; 79: 1053-1055.
- Joffe M. Validity of exposure data derived from a structured questionnaire. **Am J Epidemiol** 1992; 135: 564-570.
- Jones EE, Nisbett RE. The actor and the observer: Divergent perceptions of the causes of behavior. In EE Jones (ed), **Attribution: Perceiving the Causes of Behavior**. Morristown, NJ: General Learning Press, 1972.
- Kahn RL, Cannell CF. **The Dynamics of Interviewing: Theory, Technique and Cases**. New York: John Wiley and Sons, 1957.
- Kahneman D, Tversky A. Subjective probability: A judgment of representativeness. **Cognitive Psychology** 1972; 3: 430-454.
- Kahneman D, Tversky A. On the psychology of prediction. **Psychological Review** 1973; 80: 237-251.
- Kalish LA. Relationship of body size with breast cancer. **J Clin Oncol** 1984; 2: 287-293.
- Kalton G, Schuman H. The effect of the question on survey responses: A review. **J R Statist Soc Assoc** 1982; 145: 42-73.
- Katz D. Do interviewers bias poll results? **Public Opinion Quarterly** 1942; 6: 248-268.
- Kelsey JL. A review of the epidemiology of human breast cancer. **Epidemiologic Reviews** 1979; 1: 74-109.
- Kelsey JL, Hildreth NG. **Breast and Gynecologic Cancer Epidemiology**. Boca Raton, Florida: CRC Press, 1983.
- Kelsey JL, Thompson WE, Evans AS. **Methods in Observational Epidemiology**. New York: Oxford University Press, 1986.
- Kelsey JL, Berkowitz GS. Breast cancer epidemiology. **Cancer Research** 1988; 48: 5615-5623.
- Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. **Epidemiologic Reviews** 1991; 15: 36-47.
- Kelsey JL, Horn-Ross PL. Breast cancer: Magnitude of the problem and descriptive epidemiology. **Epidemiologic Reviews** 1993; 15: 7-16.
- Kelsey JL. Breast cancer epidemiology: Summary and future directions. **Epidemiologic Reviews** 1993; 15: 256-263.

- Keys A, Kihlberg JK. Effect of misclassification on estimated relative prevalence of a characteristic: Part I - Two populations infallibly distinguished. Part II - Errors in two variables. **AJPH** 1963; 53: 1656-1665.
- Kintsch W, Van Dijk TA. Toward a model of text comprehension and production. **Psychological Review** 1978; 85: 363-394.
- Kish L. Studies of interviewer variance for attitudinal variables. **J Am Statist Assoc** 1962; 57: 92-115.
- Kleinbaum DG, Kupper LL, Morgenstern H. **Epidemiologic Research: Principles and Quantitative Methods**. Belmont, CA: Lifetime Learning Publications, 1982.
- Klemetti A, Saxen L. Prospective versus retrospective approach in the search for environmental causes of malformations. **Am J Public Health** 1967; 57: 2071-2075.
- Kline J, Stein Z, Susser M, Warburton D. Induced abortion and spontaneous: No connection? **Am J Epidemiol** 1978; 107: 290-310.
- Kopec JA, Esdaile JM. Bias in case-control studies. A review. **J Epidemiol Comm Health** 1990; 44: 179-186.
- Kramer MS, Feinstein AR. Clinical biostatistics LIV - The biostatistics of concordance. **Clin Pharmacol Ther** 1981; 29: 111-123.
- Kramer MS, Boivin J-F. Toward an "unconfounded" classification of epidemiologic research design. **J Chron Dis** 1987; 40: 683-688.
- Kramer MS, Boivin JF. (Response) The importance of directionality in epidemiologic research design. **J Clin Epidemiol** 1988; 41: 717-718.
- Kramer MS. **Clinical Epidemiology and Biostatistics: A Primer for Clinical Investigators and Decision Makers**. Heidelberg: Springer-Verlag, 1988.
- Kramer MS, Boivin J-F. Directionality, timing, and sample selection in epidemiologic research design. **J Clin Epidemiol** 1989; 42: 827-828.
- Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response-order effects in survey measurement. **Public Opinion Quarterly** 1987; 51: 201-219.
- Kune S, Kune GA, Watson LF. Observations on the reliability and validity of the design and diet history methods in the Melbourne Colorectal Cancer Study. **Nutrition and Cancer** 1987; 9: 5-20.
- Kupper LL, McMichael AJ, Spirtas R. A hybrid epidemiologic study design useful in estimating relative risk. **J Am Statist Assoc** 1975; 70: 524-528.
- Lachenbruch PA, Reinsch S, MacRae PG, Tobis JS. Adjusting for recall bias with the proportional hazards model. **Meth Inform Med** 1991; 30: 108-110.

- Landis JR, Koch GG. The measurement of observer agreement for categorical data. **Biometrics** 1977; 33: 159-174.
- Last JM. **A dictionary of epidemiology**. (Third Edition). New York: Oxford University Press, 1995.
- Laurent A. Effects of question length on reporting behavior in the survey interview. **J Am Statist Assoc** 1972; 67: 298-305.
- Layde P, Edmonds LD, Erickson JD. Does recall bias account for the association of maternal hyperthermia and fetal neural tube defects? **Am J Epidemiol** 1979; 110: 354.
- Lebart L, Morineau A, Warwick KM. **Multivariate Descriptive Analysis: Correspondence Analysis and Related Techniques for Large Matrices**. New York: John Wiley and Sons, 1984.
- Lebarthe DR. Methodologic variation in case-control studies of reserpine and breast cancer. **J Chron Dis** 1979; 32: 95-104.
- Leck I. Backwards and forwards in search of teratogens. **Early Hum Dev** 1978; 2: 203-205.
- Leclerc A, Luce D, Lert J, Chastang F, Longeay P. (Letter) Correspondence analysis and logistic modelling: Complimentary use in the analysis of a health survey among nurses. **Statistics in Medicine** 1988; 7: 983-993.
- Lessler JT, Sirken MG. Laboratory-based research on the cognitive aspects of survey methodology: The goals and methods of the National Center for Health Statistics Study. **Millbank Memorial Fund Quarterly-Health and Society** 1985; 63: 63-81.
- Levin SM. Problems and pitfalls in conducting epidemiological research in the area of reproductive toxicology. **Am J of Industrial Medicine** 1983; 4: 349-364.
- Ley P. Studies of recall in medical settings. **Human Learning** 1982; 1: 223-233.
- Lichtenstein S, Slovic P, Fischhoff B, Layman L, Combs B. Judged frequency of lethal events. **Journal of Experimental Psychology: Human Learning and Memory** 1978; 4: 551-578.
- Lichtenstein S, Mulrow C, Elwood P. Guidelines for reading case-control studies. **J Chron Dis** 1987; 40-9:893-903.
- Lilienfeld AM, Graham S. Validity of determining circumcision status by questionnaire as related to epidemiological studies of cancer of the cervix. **JNCI** 1958; 21: 713-720.
- Lilienfeld AM, Lilienfeld DE. A century of case-control studies: Progress? **J Chron Dis** 1979; 32: 5-13.

- Lindefors-Harris BM, EKG, Adami H-O, et al. Response bias in a case-control study: Analysis utilizing comparative data concerning legal abortions from two independent Swedish studies. **Am J Epidemiol** 1991; 134: 1003-1008.
- Lindsted KD, Kuzma JW. Long-term (24-year) recall reliability in cancer cases and controls using a 21-item food frequency questionnaire. **Nutr Cancer** 1989; 12: 135-149.
- Lindsted KD, Kuzma JW. Reliability of eight-year diet recall in cancer cases and controls. **Epidemiology** 1990; 1: 392-401.
- Linnet MS, Brookmeyer R. Use of cancer controls in case-control cancer studies. **Am J Epidemiol** 1987; 125: 1-11.
- Linnet MS, Harlow SD, McLaughlin JK, McCaffrey LD. A comparison of interview data and medical records for previous medical conditions and surgery. **J Clin Epidemiol** 1989; 42: 1207-1213.
- Linton M. Transformations of memory in everyday life. In U Neisser (ed), **Memory Observed**. San Francisco: WH Freeman, 1982.
- Lippman A, Mackenzie SG. What is "recall bias" and does it exist? In M Marois (ed), **Progress in Clinical and Biological Research, Volume 163. Prevention of Physical and Mental Congenital Defects, Part C: Basic and Medical Science, Education and Future Strategies**. New York: Alan R. Liss; 1985: 206-209.
- Locker D, Wiggins R, Sittampalam Y, Patrick DL. Estimating the prevalence of disability in the community: The influence of sample design and response bias. **J of Epid and Community Health** 1981; 35: 208-212.
- Loftus EF, Loftus GR. On the permanence of stored information in the human brain. **American Psychologist** 1980; 35: 409-420.
- London SJ, Colditz GA, Stampfer MJ, et al. Prospective study of smoking and the risk of breast cancer. **JNCI** 1986; 76: 833-838.
- London SJ, Colditz GA, Stampfer MJ, et al. Prospective study of relative weight, height, and risk of breast cancer. **JAMA** 1989; 262: 2853-2858.
- London SJ, Colditz GA, Stampfer MJ, et al. Lactation and risk of breast cancer in a cohort of US Women. **Am J Epidemiol** 1990; 132: 17-26.
- Longnecker MP, Newcomb PA, Mittendorf R, et al. The reliability of self-reported alcohol consumption in the remote past. **Epidemiology** 1992; 3: 535-539.
- Lord CG. Schema and images as memory aids: Two modes of processing social information. **Journal of Personality and Social Psychology** 1980; 38: 257-269.
- Love RR, Evans AM, Josten DM. The accuracy of patient reports of a family history of cancer. **J Chron Dis** 1985; 38: 289-293.

- Lund E. (Letter) Re: "Potential for bias in case-control studies of oral contraceptives and breast cancer. *Am J Epidemiol* 1988; 127: 1067-1068.
- Lund E. The validity of different control groups in a case-control study. Oral contraceptive use and breast cancer in young women. *J Clin Epidemiol* 1989; 42: 987-993.
- Lyon JL, Egger ML, Robison LM, French TK, Gao R. Misclassification of exposure in a case-control study: The effect of different types of exposure and different proxy respondents in a study of pancreatic cancer. *Epidemiology* 1992; 3: 223-231.
- Mackenzie SG, Lippman A. An investigation of report bias in a case-control study of pregnancy outcome. *Am J Epidemiol* 1989; 129: 65-75.
- Mackenzie, SG. (Thesis) **Report Bias: Negative Evidence from a Case-Referent Study of Pregnancy Outcome**. Montreal: McGill University, 1986.
- Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987; 126: 161-169.
- MacMahon B, Cole P, Brown J. Etiology of human breast cancer: A review. *JNCI* 1973; 50: 21-42.
- Madow WG. Net differences in interview data on chronic conditions and information derived from medical records. **Vital and Health Statistics**. Series 2, No. 57. DHEW Pub. No. (HSM) 73-1331. National Center for Health Statistics. Washington. U.S. Government Printing Officer, 1973.
- Mandler JM. Categorical and schematic organization in memory. in Puff CP (ed.). **Memory Organization and Structure**. New York: Academic Press, 1979.
- Mann JL, Vessey MP, Jones R, Young D. (Letter) The case-control study and retrospective controls. *Br Med J* 1979; 2: 1507-1508.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *JNCI* 1959; 22: 719-748.
- Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973; 29: 479-486.
- Markus GB. Stability and change in political attitudes: Observed, recalled and explained. *Political Behavior* 1986; 8: 21-44.
- Markus H, Zajonc RB. The cognitive perspective in social psychology. In G Lindzey, E Aronson (eds), **Handbook of Social Psychology**. (Third Edition). New York: Random House, 1985, pp. 137-230.
- Marshall JR, Priore R, Haughey B, et al. Spouse subject interviews and the reliability of diet studies. *Am J Epidemiol* 1980; 112: 675-683.
- Marshall JR, Priore R, Graham S, Brasure J. On the distortion of risk estimates in multiple exposure level case-control studies. *Am J Epidemiol* 1981; 113: 464-473.

- Marshall JR. Random and patterned misclassification and bias in case-control studies. **Am J Epidemiol** 1982; 116: 734-5.
- Marshall JR, Graham S. Use of dual responses to increase validity of case-control studies. **J Chron Dis** 1984; 37: 125-136.
- Marshall JR. The use of dual or multiple reports in epidemiologic studies. **Statistics in Medicine** 1989; 8: 1041-1049.
- Marshall JR. Misclassification of exposure in case-control studies: Assessment by quality indices. **Epidemiology** 1994; 5: 309-314.
- Matanoski GM. Issues in the measurement of exposure. In L Gordis (ed), **Epidemiology and Health Risk Assessment**. NY: Oxford University Press, 1988, pp.107-119.
- Mathiowetz N, Groves RM. The effects of respondent rules on health survey reports. **AJPH** 1985; 75: 639-644.
- Mausner JS, Kramer S. **Mausner and Bahn - Epidemiology: An Introductory Text**. Philadelphia: W.B. Saunders, 1985.
- Mayes LC, Horwitz RI, Feinstein AR. A collection of 56 topics with contradictory results in case-control research. **Int J Epidemiol** 1988; 17: 680-685.
- McDowall J. Recall of pleasant and unpleasant words in depressed subjects. **J Abnorm Psychol** 1984; 93: 401-407.
- McFarlane MJ, Feinstein AR, Horwitz RI. Diethylstilbestrol and clear cell vaginal carcinoma - Reappraisal of the epidemiologic evidence. **Am J of Med** 1986; 81: 855-863.
- McKeown-Eyssen GE, Thomas DC. Sample size determination in case-control studies: The influence of the distribution of exposure. **J Chron Dis** 1985; 38: 559-568.
- McKinlay SM. The design and analysis of the observational study - A review. **J Am Statist Assoc** 1975; 70: 503-523.
- Meirik O, Lund E, Adami H-O, et al. Oral contraceptive use and breast cancer in young women. **Lancet** 1986; 2: 650-654.
- Miettinen OS. Matching and design efficiency in retrospective studies. **Am J Epidemiol** 1970; 91: 111-118.
- Miettinen OS. Estimability and estimation in case-referent studies. **Am J Epidemiol** 1976; 103: 226-235.
- Miettinen OS. (Comments and Dissents) The case-control study: Valid selection of subjects. **J Chron Dis** 1985; 38: 543-548.
- Miettinen OS. **Theoretical Epidemiology: Principles of Occurrence Research in Medicine**. New York: John Wiley and Sons, 1985, pp.79-82.

- Miettinen OS. "Directionality" in epidemiologic research. **J Clin Epidemiol** 1989; 42: 829-833.
- Miller AB, Bulbrook RD. UICC Multidisciplinary project on breast cancer: The epidemiology, aetiology and prevention of breast cancer. **Int J Cancer** 1986; 37: 173-177.
- Minsky M. A framework for representing knowledge. In PH Winston (ed), **The Psychology of Computer Vision**. New York: McGraw-Hill, 1975.
- Mitchell AA, Rosenberg L, Shapiro S, Slone D, Birth defects related to Bendectin use in pregnancy. I - Oral clefts and cardiac defects. **JAMA** 1981; 245: 2311-2314.
- Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. **Am J Epidemiol** 1986; 123: 670-676.
- Morabia A, Moore M, Wynder EL. Reproducibility of food frequency measurements and inferences from a case-control study. **Epidemiology** 1990; 1: 305-310.
- Morabia A, Flandre P. Misclassification bias related to definition of menopausal status in case-control studies of breast cancer. **Int J Epidemiol** 1992; 21: 222-228.
- Moran A, Engelman L, Stephen E, Fitzgerald G. Correspondence Analysis. In **BMDP Statistical Software Manual, Volume 2**, WJ Dixon, (ed). Los Angeles, CA: BMDP Statistical Software, Inc, 1990.
- Mork T. Some problems related to the use of mail questionnaires. **J Chron Dis** 1970; 23: 399-404.
- Moss L, Goldstein H. **The Recall Method in Social Surveys**. London: University of London Institute of Education, Studies in Education, 1975.
- Mote VL, Anderson RL. An investigation of the effect of misclassification on the properties of chi-square tests in the analysis of categorical data. **Biometrika** 1965; 53: 95-109.
- Moussa MA. Allocation designs in cohort and case-control studies. **Statistics in Medicine** 1986; 5: 319-326.
- Moussa MAA, Ouda BA. Correspondence analysis of contingency tables. **Computer Methods and programs in Biomedicine**. 1988; 27: 111-119.
- Munro BH, Page EB. **Statistical Methods for Health Care Research**. Philadelphia: J.B. Lippincott Company, 1993.
- National Cancer Institute of Canada. **Canadian Cancer Statistics 1995**. Toronto: 1995.

National Center for Health Statistics: Comparison of hospital reporting in three survey procedures. **Vital and Health Statistics**. PHS Pub No. 1000-Series 2, No. 8. Public Health Service. Washington. U.S. Government Printing Office, March 1965.

National Center for Health Statistics: Health interview responses compared with medical records. **Vital and Health Statistics**. PHS Pub. No. 1000-Series 2-No. 7. Public Health Service. Washington. U.S. Government Printing Office, July 1965.

National Center for Health Statistics: Reporting of hospitalization in the Health Interview Survey. **Vital and Health Statistics**. PHS Pub. No. 1000-Series 2-No. 6. Public Health Service. Washington. U.S. Government Printing Office, July 1965.

National Center for Health Statistics: Comparison of hospitalization reporting in three survey procedures. **Vital and Health Statistics**. PHS Pub. No. 1000-Series 2-No. 8. Public Health Service. Washington. U.S. Government Printing Office, July 1965.

National Center for Health Statistics: Interview data on chronic conditions compared with information derived from medical records. **Vital and Health Statistics**. PHS Pub. No. 1000-Series 2-No. 23. Public Health Service. Washington. U.S. Government Printing Office, May 1967.

National Center for Health Statistics: The influence of interviewer and respondent psychological and behavioral variables on the reporting in household interviews. **Vital and Health Statistics**. PHS Pub. No. 1000-Series 2-No. 26. Public Health Service. Washington. U.S. Government Printing Office, March 1968.

National Center for Health Statistics: Reporting of health events in household interviews: Effects of reinforcement, question length, and reinterviews. **Vital and Health Statistics**. PHS Pub. No. 1000-Series 2-No. 45. DHEW Pub No. (HSM) 72-1028. Health Services and Mental Health Administration. Washington. U.S. Government Printing Office, March 1972.

National Information Standards Organization. **The American Standard for Bibliographic References** (Index Medicus). Gaithersburg, MD: National Institute of Standards and Technology, 1977.

Neisser U (ed). **Real Events Remembered: Ecological Approaches to the Study of Memory**. Cambridge: Cambridge University Press, 1988.

Neter J, Waksberg J. A survey of response errors in expenditures data from household interviews. **J Am Statist Assoc** 1964a; 18: 18-55.

Neter J, Waksberg J. A study of response errors in experimental data from household interviews. **J Am Statist Assoc** 1964b; 59: 18-55.

Neugebauer, R. Reliability of 24 hours recall of seizure reports. **Neuroepidemiology** 1989; 8: 228-233.

- Neugebauer R, Ng S. Differential recall as a source of bias in epidemiologic research. **Clinical Epidemiology** 1990; 43: 1337-1341.
- Neutra RR, Lipscomb J, Satin K, Shusterman D. Hypotheses to explain the higher symptom rates observed around hazardous waste sites. **Environmental Health Perspectives** 1991; 94: 31-38.
- Neutra RR, Swan SH, Hertz-Picciotto I, et al. Potential sources of bias and confounding in environmental epidemiologic studies of pregnancy outcomes. **Epidemiology** 1992; 3: 134-142.
- Newell DJ. Errors in the interpretation of errors in epidemiology. **AJPH** 1962; 52: 1925-1928.
- Newell DJ. Misclassification in 2x2 tables. **Biometrics** 1963; 187-190.
- Nisbett RE, Ross L. **Human Inference: Strategies and Shortcomings of Social Judgment**. Englewoods Cliffs, NJ: Prentice-Hall, 1980.
- Nunally JC. **Psychometric Theory**. New York: McGraw-Hill, 1978.
- O'Muircheartaigh CA. Response errors in attitudinal sample survey. **Quality and Quantity** 1976; 10: 97-115.
- O'Neil M. Estimating the non-response bias due to refusals in telephone surveys. **Public Opinion Quarterly** 1979, pp.218-232.
- Oates RK, Forrest D. Reliability of mothers' reports of birth data. **Aust Paediatr J** 1984; 20: 185-186.
- Oksenberg L, Cannell CF. Some of the factors underlying the validity of response in self-reports. **International Statistical Institute Bulletin** 1977; 48: 342-346.
- Owens J, Bower GH, Black JB. The "soap opera" effect in story recall. **Memory and Cognition** 1979; 7: 185-191.
- Paffenbarger RS, Kampert JB, Chang H-G. Characteristics that predict risk of breast cancer after menopause. **Am J Epidemiol** 1980; 112: 258-268.
- Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. **Am J Epidemiol** 1982; 116: 114-122.
- Paganini-Hill A, Krailo MD, Pike MC. Age at natural menopause and breast cancer risk: The effect of errors in recall. **Am J Epidemiol** 1984; 119: 81-85.
- Paganini-Hill A, Chao A. Accuracy of recall of hip fracture, heart attack and cancer: a comparison of postal survey data and medical records. **Am J Epidemiol** 1993; 138: 101-106.
- Palmer JR, Rosenberg L, Clarke EA, Stolley PD, et al. Breast cancer and cigarette smoking: A hypothesis. **Am J Epidemiol** 1991; 134: 1-13.
- Papaiounou AN. **The etiology of human breast cancer: Endocrine, genetic, viral, immunologic and other considerations**. NY: Springer-Verlag, 1974.

- Pearce N. Incidence density matching with a simple SAS computer program. **Int J Epidemiol** 1989; 18: 981-984.
- Pearson RJC. Significance of retrospective studies. **Fed Proc** 1979; 38: 1880-1882.
- Peat JK, Salome CM, Toelle BG, Bauman A, et al. Reliability of a respiratory history questionnaire and effect of mode of administration on classification of asthma in children. **Chest** 1992; 102: 153-157.
- Petrakis NL, Ernster VL, King MC. Breast Cancer Risk Factors. In **Cancer Epidemiology and Prevention** (D Schottenfeld, J Fraunrence, eds.). Philadelphia: WB Saunders, 1982, pp. 855-870.
- Phillips DL, Clancy KJ. Some effects of social desirability in survey studies. **American Journal of Sociology** 1972; 77: 921-940.
- Pichert JW, Anderson RC. Taking different perspectives on a story. **Journal of Educational Psychology** 1977; 69: 309-315.
- Pike MC, Hill AP, Smith PG. Bias and efficiency in logistic analyses of stratified case-control studies. **Int J Epidemiol** 1980; 9: 89-95.
- Pike MC, Henderson BE, Casagrande JT, Rosario I, Gray GE. Oral contraceptives and early abortion as risk factors for breast cancer in young women. **Br J Cancer** 1981; 43:72-76.
- Pillemer DB. Flashbulb memories of the assassination attempt on President Ronald Reagan. **International Journal of Cognitive Psychology** 1984; 16: 63-80.
- Prentice RL, Pepe M, Self SG. Dietary fat and breast cancer: A quantitative assessment of the epidemiological literature. **Cancer Research** 1989; 49:3147-3156.
- Preston-Martin S, Bernstein L, Andrews Maldonado A, et al. Comparison of information from patient interviews and dental charts. **Am J Epidemiol** 1985; 121: 430-437.
- Pron GE, Burch JD, Howe GR, Miller AB. The reliability of passive smoke histories reported in a case-control study of lung cancer. **Am J Epidemiol** 1988; 127: 257-263.
- Pron, GE. (Thesis) **Reliability and validity of proxy-reported information in a case-control study of lung cancer**. Toronto: University of Toronto, 1988.
- Raphael KG. Recall bias: A proposal for assessment and control. **Int J Epidemiol** 1987; 16: 167-169.
- Raphael KG. (Letter) Differential recall as a source of bias. **J Clin Epidemiol** 1991; 44: 1281-1283.

- Raphael KG, Cloitre M, Dohrenwend BP. Problems of recall and misclassification with checklist methods of measuring stressful life events. **Health Psychol** 1991; 10: 62-74.
- Raphael KG, Cloitre M. Does mood-congruence or causal research govern recall bias? A test of life events recall. **J Clin Epidemiol** 1994; 47: 555-564.
- Reiser BJ, Black JB, Abelson RP. Knowledge structures in the organization and retrieval of autobiographical memories. **Cognitive Psychology** 1985; 17: 89-137.
- Rice SA. Contagious bias in the interview: A methodological note. **American Journal of Sociology** 1992; 35: 420-423.
- Richardson J, Gropper MS. Learning during recall trials. **Psychological Reports** 1964; 15: 551-560.
- Robins J, Pike M. The validity of case-control studies with nonrandom selection of controls. **Epidemiology** 1990; 1: 273-284.
- Rodgers WL, Herzog AR. Interviewing older adults: The accuracy of factual information. **Journal of Gerontology** 1987; 42: 387-394.
- Roeleveld NR, Kiemeny L, Schattenberg G, Peer P. Information bias in a case-referent study on mental retardation and parental occupation: Colleagues as dual respondents. **Epidemiology** 1990; 1: 292-297.
- Rohan TE, Potter JD. Retrospective assessment of dietary intake. **Am J Epidemiol** 1984; 120: 876-887.
- Roht LH, Vernon SW, Weir FW, et al. Community exposure to hazardous waste disposal sites: Assessing reporting bias. **Am J Epidemiol** 1985; 122: 418-433.
- Rorer LG. The great response-style myth. **Psychological Bulletin** 1965; 63: 129-156.
- Rosenbaum PR. Discussing hidden bias in observational studies. **Annals of Int Med** 1991; 115: 901-905.
- Rosenberg MJ, Layde PM, Ory HW, et al. Agreement between women's histories of oral contraceptive use and physician records. **Int J Epidemiol** 1983; 12: 84-87.
- Rosenberg MJ, Palmer JR, Miller DR, Clarke EA, Shapiro S. A case-control study of alcoholic beverage consumption and breast cancer. **Am J Epidemiol** 1990; 131: 6-14.
- Rosenthal R. On the social psychology of psychological experiment: The experimenter's hypotheses as unintended determinant of experimental results. **American Scientist** 1962; 51: 268-283.
- Rosenthal R. **Experimenter Effects in Behavioral Research**. New York: Appleton-Century-Crofts, 1966.

- Rosenthal R, Rosnow RL. (eds). **Artifact in Behavioral Research**. New York: Academic Press, 1969.
- Rosner B. **Fundamentals of Biostatistics**. (Third Edition). Boston: PWS-Kent Publishing Company, 1990.
- Ross MW. Retrospective distortion in homosexual research. **Archives of Sexual Behavior** 1980; 9: 523-531.
- Ross RK, Paganini-Hill A, Gerkins VR, Mack TM, Pfeffer R, et al. A case-control study of menopausal estrogen therapy and breast cancer. **JAMA** 1980; 243: 1635-1639.
- Rossi PH, Wright JD, Anderson AB (Editors). **Handbook of Survey Research**. New York: Academic Press, 1983.
- Rothman KJ. **Modern Epidemiology**. Boston: Little, Brown and Company, 1986.
- Rubin DC, Kozin M. Vivid memories. **Cognition** 1984; 16: 80-95.
- Rubin DC, Baddeley AD. Telescoping is not time compression: A model of the dating of autobiographical events. **Memory and Cognition**. 1989; 17(6):653-661.
- Sackett DL. Bias in analytic research. **J Chron Dis** 1979; 32: 51-63.
- Salmi LR, Wacholder S. Dual responses to increase validity in case-control studies. **Rev Epidem et Sante Publ** 1990; 38: 41-46.
- Sartwell PE. Retrospective studies: a review for the clinician. **Ann Int Med** 1974; 81: 381-386.
- Sartwell PE. (Comment) **J Chron Dis** 1979; 32:42-43.
- Schacter DC, Pless IB, Bruck M. Self-report of family histories of learning difficulties. **Can J Psychiatry** 1992; 37: 29-32.
- Schank R, Abelson R. **Scripts, Plans, Goals and Understanding**. Hillsdale, NJ: Erlbaum, 1977.
- Schechter MT, Miller AB, Howe GR. Cigarette smoking and breast cancer: A case-control study of screening program participants. **Am J Epidemiol** 1985; 121: 479-487.
- Schlesselman JJ. **Case-Control Studies: Design, Conduct and Analysis**. New York: Oxford University Press; 1982.
- Schlesselman JJ. Dissent: Valid selection of subjects in case-control studies. **J Chron Dis** 1985; 38: 549-550.
- Schneiderman MA, Levin DL. Parallels, convergences and departures in case-control studies and clinical trials. **Cancer Res** 1973; 33: 1498-1503.

- Schuman H, Presser S. Question wording as an independent variable in survey analysis. **Sociological Methods and Research** 1977; 6: 151-170.
- Schuman H, Presser S, Ludwig J. Context effects on survey responses to questions about abortion. **Public Opinion Quarterly** 1981; 45: 216-223.
- Schuman H, Presser S. **Questions and Answers in Attitude Surveys**. New York: Academic Press, 1981.
- Schuman H, Kalton G, Ludwig J. Context and contiguity in survey questionnaires. **Public Opinion Quarterly** 1983; 47: 112-115.
- Schwartz GG. (Letter) Recall bias in epidemiologic studies. **J Clin Epidemiol** 1990; 43-12:1431-1433.
- Seigel DG, Podgor MJ, Remaley NA. Acceptable values of kappa for comparison of two groups. **Am J Epidemiol** 1992; 135: 571-578.
- Selltiz C, Wrightsman LS, Cook SW. **Research Methods in Social Relations**. (Third Edition). New York: Holt, Rinehart and Winston, 1976.
- Shapiro S, Strax P, Venet L, Fink R. The search for risk factors in Breast Cancer. **Am J Publ Health** 1968; 58: 820-835.
- Shea S, Stein AD, Lantigua R, et al. Reliability of the behavioral risk factor survey in a triethnic population. **Am J Epidemiol** 1991; 133: 489-500.
- Shy CM, Kleinbaum DG, Morgenstern H. The effect of misclassification of exposure status in epidemiological studies of air pollution health effects. **Bull NY Acad Med** 1978; 54: 1155-1165.
- Skegg DCG. Potential for bias in case-control studies of oral contraceptives and breast cancer. **Am J Epidemiol** 1988; 127: 205-212.
- Smith AD. Age differences in encoding, storage and retrieval. In Poon LW et al. (eds). **New Directions in Memory and Aging: Proceedings of the George A. Talland Memorial Conference**. Hillsdale NJ: Erlbaum, 1980.
- Smith AH, Pearce NE, Callas PW. Cancer case-control studies with other cancers as controls. **Int J Epidemiol** 1988; 17: 298-306.
- Smith MW. The case-control or retrospective study in retrospect. **J Clin Pharmacol** 1981; 21: 269-274.
- Smith TW. Non-attitudes: A review and evaluation. In CF Turner, E Martin (eds), **Surveying Subjective Phenomenon**. New York: Russell Sage Company, 1984, pp. 215-255.
- Spanier GB. The use of recall data in survey research on human sexual behavior. **Social Biology** 1976; 23: 244-253.
- Spengler RF, Clarke EA, Woolever CA, et al. Exogenous estrogens and endometrial cancer: A case-control study and assessment of potential biases. **Am J Epidemiol** 1981; 114: 497-506.

- Spiro RJ. Remembering information from text: 'State of Schema' reconstruction hypothesis. In RC Anderson, RJ Shapiro, WE Montague (eds), **Schooling and the Acquisition of Knowledge**. Hillsdale, NJ: Erlbaum, 1977.
- Spiro RJ. Prior knowledge and story processing: Integration, selection and variation. **Poetics** 1980a; 9: 313-327.
- Spiro RJ. Accomodative reconstruction in prose recall. **Journal of Verbal Learning and Verbal Behavior** 1980b; 19: 84-95.
- Spitzer WO. Ideas and words: Two dimensions for debates on case controlling. **J Chron Dis** 1985; 38: 541-2.
- SPSS Inc (Norusis, MJ). **SPSS Advanced Statistics Student Guide**, Release 4. Chicago, IL: SPSS Inc, 1990.
- SPSS Inc (Norusis, MJ). **SPSS Advanced Statistics User's Guide**, Release 4. Chicago, IL: SPSS Inc, 1990.
- SPSS Inc (Norusis, MJ). **SPSS Base System User's Guide**, Release 4. Chicago, IL: SPSS Inc, 1990.
- SPSS Inc. **SPSS Reference Guide**, Release 4. Chicago, IL: SPSS Inc, 1990.
- Stavracky KM, Emmons S. Breast cancer in premenopausal and postmenopausal women. **JNCI** 1974; 53: 647-654.
- Stavracky KM, Clarke AE. Hospital or population controls? An unanswered question. **J Chron Dis** 1983; 36: 301-307.
- Stolley PD, Tonascia JA, Sartwell PE, Tockman MS, Tonascia S, et al. Agreement rates between the oral contraceptive users and prescribers in relation to drug histories. **Am J Epidemiol** 1978; 107: 226-235.
- Stott DH. Some psychosomatic aspects of causality in reproduction. **J Psychosom Res** 1958; 3: 42-45.
- Striker LS. Acquiescence and social desirability response styles, item characteristics, and conformity. **Psychological Reports** 1976; 12: 319-341.
- Stryker S, Statham A. Symbolic interactionism and role theory. In G Lindzey, E Aronson (eds), **Handbook of Social Psychology**. (Third Edition). New York: Random House, 1985, pp. 311-379.
- Sudman S, Bradburn NM. Effects of time and memory factors on response in surveys. **J Am Statist Assoc** 1973; 68: 805-815.
- Sudman S, Bradburn NM. **Response Effects in Surveys: A Review and Synthesis**. Chicago: Aldine, 1974.
- Sudman S, Bradburn NM, Blair E, Stocking C. Modest expectations: The effects of interviewers' prior expectations on responses. **Sociological Methods and Research** 1977; 6: 171-182.

- Sudman S. Reducing reponse errors in surveys. **Statistician** 1980; 29: 237-273.
- Sun M. Federal VDT study finally wins approval. **Science** 1986; 232: 1594-1595.
- Swan SH, Shaw GM. (Letter) Re: "Reporting accuracy among mothers of malformed and nonmalformed infants. **Am J Epidemiol** 1990; 133: 935-937.
- Swan SH, Shaw GM, Schulman J. Reporting and selection bias in case-control studies of congenital malformations. **Epidemiology** 1992; 3: 356-363.
- Thomas DB. Epidemiologic and Related Studies of Breast Cancer Etiology. In AM Lilienfeld (ed), **Reviews in Cancer Epidemiology, Vol V. IARC Sci Publ. No. 88.** Lyon: IARC, 1987.
- Thompson WD. On the comparison of effects. **AJPH** 1987; 77: 256-258.
- Thompson WJ. Kappa and attenuation of the odds ratio. **Epidemiology** 1990; 1: 357-369.
- Thompson WJ. Nonrandom yet unbiased. **Epidemiology** 1990; 1: 262-264.
- Thorndyke PW. Cognitive structures in comprehension and memory of narrative discourse. **Cognitive Psychology** 1977; 9: 77-110.
- Tilley BC, Barnes AB, Bergstrahl E, et al. A comparison of pregnancy history recall and medical records. Implications for retrospective studies. **Am J Epidemiol** 1985; 121: 269-281.
- Tourangeau R, Lessler J, Salter W. **Cognitive Aspects of Questionnaire Design: Part C (Technical Report.** Chicago: National Opinion Research Center, 1986).
- Tretli S, Lund-Larsen PG, Foss OP. Reliability of questionnaire information on cardiovascular disease and diabetes: Cardiovascular disease study in Finnmark county. **J Epidemiol Community Health** 1982; 36: 269-273.
- Tulving E, Psotka J. Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. **Journal of Experimental Psychology** 1971; 87: 1-8.
- Tulving E. Episodic and semantic memory. In E Tulving, W Donaldson (eds), **Organization of Memory.** New York: Academic Press, 1972.
- Tulving E, Thomson DM. Encoding specificity and retrieval processes in episodic memory. **Psychological Review** 1973; 80: 352-373.
- Tversky A, Kahneman D. Belief in the law of small numbers. **Psychological Bulletin** 1971; 76: 105-110.
- Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. **Cognitive Psychology** 1973; 5:207-232.

- Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. **Science** 1974; 185: 1124-1131.
- Ucros CG. Mood state-dependent memory: A meta-analysis. **Cognit Emot** 1989; 3: 139-167.
- UK National Case-Control Study Group. Oral contraceptive use and breast cancer risk in young women. **Lancet** 1989; 973-982.
- Underwood BJ. Attributes of memory. **Psychological Review** 1969; 76: 559-573.
- Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. **Am J Epidemiol** 1991; 124: 895-907.
- Van Leeuwen FE, De Vet HCW, Hayes RB, et al. As assessment of the relative validity of retrospective interviewing for measuring dietary intake. **Am J Epidemiol** 1983; 118: 752-758.
- Velema JP, Blettner M. Re: "Agreement between questionnaire data and medical records: The evidence for accuracy of recall". (Letter) **Am J Epidemiol** 1990; 131: 1100-1101.
- Velema JP, Blettner M, Restrepo M, Munoz N. The evaluation of agreement by means of log-linear models: Proxy interviews on reproductive history among floriculture workers in Colombia. **Epidemiology** 1991; 2: 107-115.
- Verkerk PH. The impact of alcohol misclassification on the relationship between alcohol and pregnancy outcome. **Int J Epidemiol** 1992; 21: S23-S37.
- Vihko R, Apter D. Endogenous steroids in the pathophysiology of breast cancer. **Critical Reviews in Oncology/Hematology** 1989; 9: 1-16.
- Vobecky JS, Vobecky J, Froda S. The reliability of maternal memory in a retrospective assessment of nutritional status. **J Clin Epidemiol** 1988; 41: 261-265.
- Wacholder S, Dosemici M, Lubin J. Blind assessment of exposure does not always prevent differential misclassification. **Am J Epidemiol** 1991; 130:433-437.
- Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. **Am J Epidemiol** 1992; 135: 1019-1028.
- Wacholder S. When measurement errors correlate with truth: Surprising effects of nondifferential misclassification. **Epidemiology** 1995; 6: 157-161.
- Wagenaar WA. My memory: A study of autobiographical memory over six years. **Cognitive Psychology** 1986; 18: 225-252.
- Walker AM, Velema JP, Robins JM. Analysis of case-control data derived in part from proxy respondents. **Am J Epidemiol** 1988; 127: 905-914.
- Wallis C. A puzzling plague. **Time** January 14,1991; 42-46.

- Warner KE. Possible increases in the underreporting of cigarette consumption. **J Am Statist Assoc** 1978; 73: 314-318.
- Warren-Burhenne LJ, Hislop TG, Burhenne HJ. The BC Mammography screening program: Evaluation of the first 15 months. **AJR** 1992; 158: 45-49.
- Weinberg CR, Umbach DM, Greenland S. When will nondifferential misclassification of an exposure preserve the direction of a trend? **Am J Epidemiol** 1994; 140: 565-571.
- Weinstock MA, Colditz GA, Willett WC, Walter C, et al. Recall (report) bias and reliability in the retrospective assessment of melanoma risk. **Am J Epidemiol** 1991; 133: 240-245.
- Weiss CH. Validity of welfare mothers' interview responses. **Public Opinion Quarterly** 1968; 32: 622-633.
- Weiss KM, Chakraborty R, Majumder PP. Problems in the assessment of relative risk of chronic disease among biological relatives of affected individuals. **J Chron Dis** 1982; 35: 539-551.
- Werler MM, Nelson K, Pober BR, Holmes LB. Recall accuracy among mothers of malformed and normal infants. **Teratology** 1986; 33: 30C.
- Werler MM, Pober BR, Nelson K, Holmes LB. Reporting accuracy among mothers of malformed and nonmalformed infants. **Am J Epidemiol** 1989; 129: 415-421.
- Werler MM, Nelson K, Pober BR, Holmes LB. Author's reply to "Reporting accuracy among mothers of malformed and nonmalformed infants". **Am J Epidemiol** 1990; 131: 937-938.
- West DW, Schuman KL, Lyon JL, Robison LM, Allred R. Differences in risk estimations from a hospital and a population-based case-control study. **Int J Epidemiol** 1984; 13: 235-239.
- Whitten WB, Leonard JM. Directed search through autobiographical memory. **Memory and Cognition** 1981; 9: 566-579.
- Wickelgren WA. Multitrac Strengths Theory. In DA Norman (ed), **Models of Human Memory**. New York: Academic Press, 1972.
- Wigle DT, Mao Y, Semenciw R, Morrison HI. Cancer patients in Canada. **CMAJ** 1986; 134:231-235.
- Wilcox AJ, Horney LF. Accuracy of spontaneous abortion recall. **Am J Epidemiol** 1984; 120: 727-733.
- Willett WC. An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. **Statistics in Medicine** 1989; 8: 1031-1040.
- Williams MD, Hollan JD. The process of retrieval from very long-term memory. **Cognitive Science** 1981; 5: 87-119.

- Wolkind S, Coleman EZ. Adult psychiatric disorder and childhood experiences: The validity of retrospective data. **Br J Psychiat** 1983; 143: 188-191.
- Woods NF, Most A, Dery GK. Estimating perimenstrual distress: A comparison of two methods. **Res Nurs Health** 1982; 5: 81-91.
- Wright JG, Feinstein AR. A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. **J Clin Epidemiol** 1992; 45: 1201-1218.
- Wynder EL, Licklider SD. The question of circumcision. **Cancer** 1960; 13: 442-445.
- Wynder EL. Investigator bias and interviewer bias: The problem of reporting systematic error in epidemiology. **J Clin Epidemiol** 1994; 47: 825-827.
- Yekovich FB, Thorndyke PW. An evaluation of alternative functional models of narrative schemata. **Journal of Verbal Learning and Verbal Behavior** 1981; 20: 454-469.
- Zierler S, Rothman KJ. Congenital heart disease in relation to maternal use of bendectin and other drugs in early pregnancy. **N Engl J Med** 1985; 313: 347-352.

Appendix 1: Validity Scale Development - Selection of Exposure Variables and  
Weighting Factors: Letter of Information to Potential Subjects

Appendix 2: The Validity Scale Questionnaire Used to Evaluate the Plausibility  
and Perceived Etiologic Importance of Several Exposure Variables  
and Conditions Being Considered For Possible Inclusion in  
Questionnaire Two

METHODS TO COLLECT  
MEDICAL AND EXPOSURE HISTORIES

Thank you for agreeing to participate in the first part of the study on the "Methods to Collect Exposure and Medical Histories". Please read the following instructions before you begin to answer the attached questionnaire.

Below is a list of factors (various exposures and health conditions) that have been considered as possible causes of cancer. Consider each factor carefully, and select the response that BEST describes "HOW IMPORTANT" you feel the factor is as a RISK FACTOR responsible for the development of breast cancer in women. You are to indicate if you feel the presence of the factor would be a high risk, moderate risk, low risk, or no risk in the development of breast cancer.

To indicate your choice, please circle the number under the heading which BEST describes how you feel about the importance statement. Do NOT skip any item. If you change your mind, please cross out your first choice completely, and then circle the appropriate choice.

RESPONSE STATEMENTS

	VERY IMPORTANT   HIGH RISK  1	MODERATELY IMPORTANT   MODERATE RISK  2	SOMEWHAT IMPORTANT   LOW RISK  3	NOT IMPORTANT   NO RISK  4
1. A woman's RACIAL background (ie., white);	1	2	3	4
2. A previous history of traumatic injury to the chest/breast area;	1	2	3	4
3. A woman's MARITAL status: (ie., NEVER MARRIED);	1	2	3	4
4. The overall size of a woman's breasts;	1	2	3	4
5. The surgical removal of both ovaries;	1	2	3	4
6. The regular use of artificial sweeteners (aspartame/saccharin) in foods and drinks;	1	2	3	4
7. Beginning menopause at an early age (before 45 years of age);	1	2	3	4

## RESPONSE STATEMENTS

	VERY IMPORTANT   HIGH RISK 1	MODERATELY IMPORTANT   MODERATE RISK 2	SOMEWHAT IMPORTANT   LOW RISK 3	NOT IMPORTANT   NO RISK 4
8. Cigarette smoking (ie., nicotine exposure);	1	2	3	4
9. A woman's body build after menopause begins - (ie., excess weight/obesity);	1	2	3	4
10. A diet high in saturated fats/cholesterol;	1	2	3	4
11. A previous history of fibrocystic or benign breast disease;	1	2	3	4
12. A history of irregular & painful periods;	1	2	3	4
13. A woman having her first baby after age 30;	1	2	3	4
14. Excessive coffee consumption (ie., greater than 5 cups per day);	1	2	3	4
15. Having a large number of children (greater than four pregnancies);	1	2	3	4
16. Chronic breast infection (mastitis) during the period when breast-feeding;	1	2	3	4
17. Starting ones menstrual cycle (monthly periods) at an early age ( before age 13 years);	1	2	3	4
18. Regular use of alcohol (hard liquor, wine, or beer);	1	2	3	4
19. Exposure of a woman's mother to diethylstilbestrol (DES)- a drug used during pregnancy to prevent a miscarriage;	1	2	3	4

RESPONSE STATEMENTS

	VERY IMPORTANT   HIGH RISK 1	MODERATELY IMPORTANT   MODERATE RISK 2	SOMEWHAT IMPORTANT   LOW RISK 3	NOT IMPORTANT   NO RISK 4
20. Exposure to chronic viral infections such as herpes, hepatitis B or HIV-(AIDS);	1	2	3	4
21. Experiencing prolonged periods of extreme personal stress;	1	2	3	4
22. The presence of silicon breast implants (for breast enlargement);	1	2	3	4
23. The use of estrogen containing drugs after menopause (such as premarin, climacteron or hormone injections);	1	2	3	4
24. Longterm, chronic exposure of breasts to ultra-violet radiation during sunbathing;	1	2	3	4
25. Exposure to chemicals used in the manufacture of paints, paint removers, solvents, varnishes, wood stains;	1	2	3	4
26. Exposure to urea formaldehyde which is used in home insulation;	1	2	3	4
27. Environmental or occupational exposure to asbestos products including asbestos insulation, brake lining and building materials or fire retardants;	1	2	3	4
28. Exposure to chemical weed killers (phenoxy herbicides);	1	2	3	4
29. Exposure to household bleaches, cleaning agents & solvents containing such chemicals as ammonia, chlorine, acetone or benzene;	1	2	3	4

RESPONSE STATEMENTS

	VERY IMPORTANT   HIGH RISK 1	MODERATELY IMPORTANT   MODERATE RISK 2	SOMEWHAT IMPORTANT   LOW RISK 3	NOT IMPORTANT   NO RISK 4
30. The use of spray underarm deodorants;	1	2	3	4
31. Consumption of food additives (iron, MSG, chemical preservatives, high sodium etc);	1	2	3	4
32. The use of oral contraceptives (birth control pills);	1	2	3	4
33. The use of various drugs:				
a. reserpine (Serapsil) for control of high blood pressure;	1	2	3	4
b. analgesics (pain killers) such as aspirin, acetaminophen (Tylenol), Ibuprofen (Motrin);	1	2	3	4
c. Thiazides (water pills)- furosemide (Lasix), hydrochlorothiazide (Hydrodiuril);	1	2	3	4
d. trihexahydrol	1	2	3	4
e. Captopril (Capoten) for high blood pressure	1	2	3	4
f. Propranolol hydrochloride (Inderal)- a drug used for control of high blood pressure;	1	2	3	4
g. elavil (a mood elevator)	1	2	3	4
h. steroids / (cortisone);	1	2	3	4
34. A history of an underactive thyroid and the need for thyroxin replacement therapy;	1	2	3	4
35. A history of breast cancer in a mother or sister;	1	2	3	4

## RESPONSE STATEMENTS

<b>VERY IMPORTANT</b>	<b>MODERATELY IMPORTANT</b>	<b>SOMEWHAT IMPORTANT</b>	<b>NOT IMPORTANT</b>
<b>HIGH RISK</b>	<b>MODERATE RISK</b>	<b>LOW RISK</b>	<b>NO RISK</b>
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

36. Eating foods containing nitrites and nitrates (cured & smoked meats and cheeses);

**1**

2

3

4

### OPTIONAL QUESTION

37. List any other factors or conditions NOT mentioned previously that you feel play a significant role in the development of breast cancer:

[illegible]

Appendix 3: The Enrolment Screening Questionnaire Administered to Mammography Clients by the Screening Mammography Program of British Columbia. (This Questionnaire is referred to in this thesis as Questionnaire One (Q1). Pre-diagnostic exposure data were generated from responses to Q1 items).



I.D.: \_\_\_\_\_

Name: \_\_\_\_\_

This information is important for the overall evaluation of the Screening Mammography Program. All information will be held in strictest confidence.

2. What is your present marital status?
- |  |                                    |                                     |
|--|------------------------------------|-------------------------------------|
| never married <input type="checkbox"/> | married <input type="checkbox"/>   | common law <input type="checkbox"/> |
| widowed <input type="checkbox"/>       | separated <input type="checkbox"/> | divorced <input type="checkbox"/>   |

3. What is the highest level of education that you have reached?

If elementary/secondary, please indicate grade: \_\_\_\_\_

If post secondary, please indicate program:

certificate/diploma ☐ college/university ☐ some bachelor degree ☐ graduate degree ☐

4. What is your racial origin?

White ☐ Chinese ☐ Japanese ☐ Native Indian ☐ East Indian ☐

Other ☐ please specify \_\_\_\_\_

5. What is your present occupation? \_\_\_\_\_

6. Do you practice breast self-examination (BSE)? Yes ☐ No ☐

If yes: How many times per year? \_\_\_\_\_

7. Have you had a breast physical examination by your doctor in the last 12 months? Yes ☐ No ☐

8. Did your doctor refer you to the Program? Yes ☐ No ☐

If no: Where did you hear about the program? (please check the most persuasive one)

TV ☐ Radio ☐ Magazine/Newspaper ☐ Poster/Pamphlet ☐ Friend ☐ Other ☐

9. Have you ever had a mammogram (breast x-ray examination)? Yes ☐ No ☐

If yes: How many mammograms? \_\_\_\_\_ Date of most recent mammogram \_\_\_\_\_ / \_\_\_\_\_  
month year

10. Have you ever had any breast needle aspiration (a needle to obtain breast tissue or fluid)?

Yes ☐ No ☐

If yes: How many? \_\_\_\_\_ Age at first aspiration: \_\_\_\_\_

11. Age at first menstrual period: \_\_\_\_\_

Are you still having menstrual periods? Yes ☐ No ☐

If no: Age when menstrual periods stopped: \_\_\_\_\_

Did you have a hysterectomy at that time? Yes ☐ No ☐

Did you have both ovaries totally removed? Yes ☐ No ☐

12. Have you ever been pregnant? Yes ☐ No ☐

If yes: How many deliveries? \_\_\_\_\_ Age at first delivery? \_\_\_\_\_

13. Excluding during pregnancy, have you ever had painful or tender breasts? Yes ☐ No ☐

If yes: Which of the following symptoms did you usually have? Pain alone ☐ Tenderness alone ☐ Tenderness with pain ☐

Was it related to your periods? Yes ☐ No ☐

Approximate age when it started: \_\_\_\_\_

Has it stopped? Yes ☐ No ☐

If yes: Approximate age when it stopped: \_\_\_\_\_

14. Could you usually tell by the way your body feels (or felt) that your period was coming?

No ☐ Sometimes ☐ Always ☐ Uncertain ☐

15. Did you usually have breast swelling before your periods? Yes ☐ No ☐

If yes: Approximate age when it started: \_\_\_\_\_

Has it stopped? Yes ☐ No ☐

If yes: Approximate age when it stopped: \_\_\_\_\_

16. Have you ever taken oral contraceptives (birth control pills)? Yes ☐ No ☐

If yes: Age at first taking oral contraceptives: \_\_\_\_\_

Do you still take them? Yes ☐ No ☐

If no: Approximate age when you stopped: \_\_\_\_\_

Total number of years you have taken them: \_\_\_\_\_

17. Have you ever taken other estrogen containing drugs (such as praemarin, climacteron or hormone injections)?

Yes ☐ No ☐

If yes: Age at first taking estrogens: \_\_\_\_\_

Do you still take them? Yes ☐ No ☐

If no: Approximate age when you stopped: \_\_\_\_\_

Total number of years you have taken them: \_\_\_\_\_

18. Did your mother have breast cancer? Yes ☐ No ☐

If yes: Was it diagnosed before age 50? Yes ☐ No ☐ Unknown ☐

Did it involve both breasts? Yes ☐ No ☐ Unknown ☐

19. How many (full) sisters do/did you have? \_\_\_\_\_

Did any of these sisters have breast cancer? Yes ☐ No ☐

If yes: How many? \_\_\_\_\_

How many had it diagnosed before age 50? \_\_\_\_\_

How many had it involving both breasts? \_\_\_\_\_

20. Have you ever smoked? Yes ☐ No ☐

If yes: How many cigarettes per day do/did you smoke on average? \_\_\_\_\_

Approximate age when you first started smoking? \_\_\_\_\_

Do you still smoke? Yes ☐ No ☐

If no: Approximate age when you stopped: \_\_\_\_\_

Total number of years you have smoked (approximately): \_\_\_\_\_

21. Have you ever drunk alcoholic beverages more than once or twice a month? Yes ☐ No ☐

If yes: How many drinks per month do/did you usually consume? \_\_\_\_\_

Approximate age when you first started? \_\_\_\_\_

Do you still drink alcoholic beverages? Yes ☐ No ☐

If no: Approximate age when you stopped: \_\_\_\_\_

What do/did you usually drink? Beer ☐ Wine ☐ Liquor ☐

22. Approximate height: \_\_\_\_\_ cm. or \_\_\_\_\_ ft./in.

23. Approximate weight: \_\_\_\_\_ kg. or \_\_\_\_\_ lbs.

Appendix 4: Letter of Introduction Sent to Potential Subjects from the Executive Director of the Screening Mammography Program of British Columbia. (This letter introduces the 'purpose' of the research study as well as the researcher who would be calling the potential subjects to determine their willingness to participate).

Appendix 5: Letter of Information Accompanying the Study Questionnaire  
(Questionnaire Two)

Appendix 6: The Study Questionnaire - Questionnaire Two (Q2). (This questionnaire was used to collect retrospective (post-diagnostic) exposure reports as well as subject responses to the included validity scale items.

## STUDY QUESTIONNAIRE

"A STUDY OF THE METHODS TO COLLECT MEDICAL AND EXPOSURE HISTORIES "

STUDY IDENTIFICATION NUMBER: \_\_\_\_\_ NAME: \_\_\_\_\_

The following information is important to medical researchers who are studying the factors which may or may not be responsible for the development of breast cancer in women. It is requested that you answer ALL questions as fully and as accurately as possible. The information that you give will be kept in strictest confidence.

1. What is your present marital status?

never married \_\_\_\_\_

married \_\_\_\_\_

widowed \_\_\_\_\_

separated \_\_\_\_\_

common law \_\_\_\_\_

divorced \_\_\_\_\_

2. What is your racial origin?

White \_\_\_\_\_

Chinese \_\_\_\_\_

Japanese \_\_\_\_\_

Native Indian \_\_\_\_\_

East indian \_\_\_\_\_

Other \_\_\_\_\_ (please specify) \_\_\_\_\_

3. Were you born in Canada? Yes \_\_\_\_\_ No \_\_\_\_\_

If YES, go to QUESTION 4

If NO, please specify the country in which you were born? \_\_\_\_\_

How old were you when you immigrated to Canada? \_\_\_\_\_

4. What is your present occupation? \_\_\_\_\_

If you are now retired, please specify the type of work you were employed in prior to your retirement? \_\_\_\_\_

5. What is the highest level of education that you have reached?

If elementary/secondary, please indicate grade: \_\_\_\_\_

If post-secondary, please indicate program and course of study:

certificate/diploma \_\_\_\_\_; course of study \_\_\_\_\_

some college/university \_\_\_\_\_; course of study \_\_\_\_\_

bachelor degree \_\_\_\_\_; course of study \_\_\_\_\_

graduate degree \_\_\_\_\_; course of study \_\_\_\_\_

6. What is your date of birth? \_\_\_\_\_  
day/month/year

7. Have you had a physical exam by your family doctor in the last year? Yes \_\_\_\_\_ No \_\_\_\_\_

8. Have you had a PAP smear in the last year? Yes \_\_\_\_\_ No \_\_\_\_\_

9. Do you practice breast self-examination (BSE)? Yes \_\_\_\_\_ No \_\_\_\_\_

If NO, go to QUESTION 10

If YES: How many times per year did you practice breast self-examination? \_\_\_\_\_

10. How many mammograms (breast x-ray examinations) have you had? \_\_\_\_\_

11. What was the date of your most recent mammogram? \_\_\_\_\_  
month/year

12. Have you ever had any breast needle aspiration (a needle to obtain breast tissue or fluid)? Yes \_\_\_\_\_ No \_\_\_\_\_

If NO, go to QUESTION 13

If YES: How many needle aspirations have you had? \_\_\_\_\_

Age at first aspiration? \_\_\_\_\_

13. Age at first menstrual period? \_\_\_\_\_

14. Are you still having menstrual periods? Yes \_\_\_\_\_ No \_\_\_\_\_

If YES, go to QUESTION 15

If NO: Age when your menstrual periods stopped? \_\_\_\_\_

Did you have a hysterectomy at the time? Yes ☐ No ☐

If YES: Did you have both ovaries totally removed? Yes ☐ No ☐

15. Before the age of 40 years, did you experience dysmenorrhea (painful and/or irregular menstrual periods)? Yes ☐ No ☐

16. Could you usually tell by the way your body feels (or felt) that your menstrual period was coming?

No ☐ Sometimes ☐ Always ☐ Uncertain ☐

17. Did you usually have breast swelling before your menstrual periods? Yes ☐ No ☐

If NO, go to QUESTION 18

If YES: Approximate age when breast swelling started:

Has the breast swelling stopped? Yes ☐ No ☐

If YES: Approximate age when breast swelling stopped:

18. Have you experienced traumatic (painful) injury/injuries to the breast or chest area? Yes ☐ No ☐

If NO, go to QUESTION 19

If YES: Briefly describe your injury/injuries describing when the injury occurred, what the injury was, and what medical treatment was required (if any)?

---



---



---



---

19. Excluding during pregnancy, have you ever had painful or tender breasts? Yes ☐ No ☐

If NO, go to QUESTION 20

If YES: Which of the following SYMPTOMS did you usually have?

pain alone ☐

tenderness alone ☐

tenderness with pain ☐

Was the pain and/or tenderness related to your menstrual periods? Yes ☐ No ☐

Approximate age when your breast pain/tenderness started?

Has your breast pain/tenderness stopped? Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 20

If YES: Approximate age when your breast pain/tenderness stopped: \_\_\_\_

20. Have you ever been pregnant? Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 23

If YES: How many deliveries? \_\_\_\_

Age at first delivery? \_\_\_\_

21. Did you breastfeed one (or more) of your children?

Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 23

If YES: On average, how many months did you breastfeed your children? \_\_\_\_ (months)

22. Did you ever experience a longlasting breast infection (mastitis) during any of the periods when you were breastfeeding?

No, never \_\_\_\_

Yes, once \_\_\_\_

Yes, two or more times \_\_\_\_

23. Have you ever taken oral contraceptives (birth control pills)?

Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 24

If YES: Do you still take birth control pills? Yes \_\_\_\_ No \_\_\_\_

If NO: Approximate age when you stopped birth control pills: \_\_\_\_

What is the total number of years that you have taken oral contraceptives (birth control pills)? \_\_\_\_

24. Have you ever taken other estrogen containing drugs (such as premarin, climacteron, or hormone injections)?

Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 25

If YES: What was your age when you started taking estrogens? \_\_\_\_

Do you still take estrogens? Yes ☐ No ☐

If you are no longer taking estrogens, please specify your approximate age when you stopped taking estrogens:  (years)

The total number of years that you took estrogens is:

25. Before YOUR birth, was your mother exposed to diethylstilbesterol (DES) - a drug that was used to prevent miscarriage?  
Yes ☐ No ☐

26. Have you been told by your doctor that you have an underactive thyroid gland? Yes ☐ No ☐

If NO, go to QUESTION 27

If YES: Do you take a thyroid stimulating drug such as ELTROXIN?  
Yes ☐ No ☐

If YES: What was the name of the drug taken?

The total number of years that you took this drug:

27. Have you ever experienced periods of great stress which interfered with your usual daily activities?

No, never ☐

Yes, once or twice ☐

Yes, more than two times ☐

28. Did your mother have breast cancer? Yes ☐ No ☐

If NO, go to QUESTION 29

If YES: Was it diagnosed before age 50?  
Yes ☐ No ☐ Unknown ☐

Did the breast cancer involve both breasts?  
Yes ☐ No ☐ Unknown ☐

29. How many (full) sisters do/did you have?

If you have no sisters, go to QUESTION 30

If you have sisters, please answer the following questions:

Did any of your sisters have breast cancer? Yes ☐ No ☐

If NO, go to QUESTION 30

If YES: How many (full) sisters had breast cancer?

How many of your sisters had breast cancer diagnosed before age 50? \_\_\_\_

How many had breast cancer involving both breasts? \_\_\_\_

THE NEXT QUESTION ASKS ABOUT YOUR EXPOSURES AT WORK OR AT HOME TO VARIOUS CHEMICALS

Have you been in contact with the following:

30. a. chemical weed killers (phenoxy herbicides)?  
Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 30b

If YES: How would you describe your frequency of contact with the chemical weed killers (phenoxy herbicides)?

- (1) frequent - the greater part of each working day (more than 20 hours per week) \_\_\_\_
- (2) regular contact between 5-19 hours per week \_\_\_\_
- (3) occasional contact between 1-4 hours per week \_\_\_\_
- (4) only rarely - less than 1 hour per week \_\_\_\_

30. b. chemicals used in the manufacture of paints, paint removers, paint solvents, varnishes and wood stains?  
Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 30c

If YES: How would you describe the frequency of your contact over the years?

- (1) frequent - the greater part of each working day (more than 20 hours per week) \_\_\_\_
- (2) regular contact between 5-19 hours per week \_\_\_\_
- (3) occasional contact between 1-4 hours per week \_\_\_\_
- (4) only rarely - less than 1 hour per week \_\_\_\_

30. c. urea formaldehyde which is used to insulate buildings, homes and apartment complexes? Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 30d

If YES: Approximately how many years (total) did you live or work in buildings insulated with urea formaldehyde? \_\_\_\_ (years)

30. d. asbestos products, including insulation, brake linings, oven mitts, ironing board covers and building materials?  
Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 31

If YES: How would you describe the frequency of your contact over the years?

- (1) frequent - the greater part of each working day  
(more than 20 hours per week) \_\_\_\_
- (2) regular contact between 5-19 hours per week \_\_\_\_
- (3) occasional contact between 1-4 hours per week \_\_\_\_
- (4) only rarely - less than 1 hour per week \_\_\_\_

THE NEXT QUESTION ASKS ABOUT YOUR EXPOSURE TO VARIOUS DRUGS
---

31. Have you taken any of the following drugs for six months or longer?

- a. reserpine (Serapsil) - a drug used in the control of high blood pressure? Yes \_\_\_\_ No \_\_\_\_
- b. elavil (amitrypteline) - a drug to improve your mood? Yes \_\_\_\_ No \_\_\_\_
- c. trihexahydrol (Mellal) - a tranquillizer? Yes \_\_\_\_ No \_\_\_\_
- d. steroids (prednisone, cortisone, medrol, betaderm)? Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 32

If YES: What was the name of the steroid taken? \_\_\_\_\_

<p>THE NEXT SET OF QUESTIONS SEEK INFORMATION ABOUT LIFESTYLE AND DIETARY FACTORS</p>
---

32. Have you ever smoked? Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 33

If YES: How many cigarettes per day do/did you smoke on average?  
\_\_\_\_ cigarettes per day

How old were you when you first started smoking? \_\_\_\_

Do you still smoke? Yes \_\_\_\_ No \_\_\_\_

If you still smoke, go to QUESTION 33

If you are now a non-smoker:

How old were you when you stopped smoking? \_\_\_\_

How many years (in total) were you a smoker? \_\_\_\_

33. Have you ever drunk alcoholic beverages more than once or twice  
a month? Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 34

If YES: How many drinks per month do/did you usually consume? \_\_\_\_

How old were you when you started drinking? \_\_\_\_

What do/did you usually drink?

Beer \_\_\_\_ Wine \_\_\_\_ Liquor \_\_\_\_

Do you still drink alcoholic beverages? Yes \_\_\_\_ No \_\_\_\_

If NO: How old were you when you stopped drinking? \_\_\_\_

34. Do/Did you sunbathe during the summer months, and/or on winter  
vacations? Yes \_\_\_\_ No \_\_\_\_

If NO, go to QUESTION 35

If YES: please answer the next two questions about sunbathing:

How often do/did you sunbathe?

a. occasionally (less than 5 hours per week) \_\_\_\_

b. regularly (5 to 19 hours per week) \_\_\_\_

c. frequently (greater than 20 hours per week) \_\_\_\_

35. I would describe my average weekly diet to be: (check one only)
- a. high in fats and cholesterol \_\_\_\_\_
  - b. normal/average with respect to the consumption of fats and cholesterol \_\_\_\_\_
  - c. low in fats and cholesterol \_\_\_\_\_
36. Do you regularly consume 6 or more servings per week of smoked or cured meats such as European sausages, salamis, bacon, ham or pork hocks? Yes \_\_\_\_\_ No \_\_\_\_\_
37. Do you consume food products containing MSG (monosodium glutamate) - a food additive used for tenderizing, flavor enhancement, or for preserving freshness? Yes \_\_\_\_\_ No \_\_\_\_\_
38. How often do you consume food items or drinks that contain artificial sweeteners?
- a. never \_\_\_\_\_
  - b. occasionally (one or two servings per week) \_\_\_\_\_
  - c. regularly (greater than three servings per week) \_\_\_\_\_
39. Approximate height: \_\_\_\_\_ cm. or \_\_\_\_\_ feet/inches
40. Approximate weight: \_\_\_\_\_ kg. or \_\_\_\_\_ lbs (pounds)

THANK YOU FOR TAKING TIME TO ANSWER THESE QUESTIONS. PLEASE CHECK THAT YOU HAVE ANSWERED EACH QUESTION BEFORE PLACING THE COMPLETED QUESTIONNAIRE IN THE ENVELOPE PROVIDED

## Appendix 7: Reminder Letter Sent to Non-Respondents

Appendix 8: Codebook Developed for Questionnaire One and Two Data  
Processing

## CODEBOOK

## RECORD 1 - DEMOGRAPHIC AND OTHER PERSONAL INFORMATION

<u>ITEM</u>	<u>VARIABLE CODE</u>	<u>COLUMNS</u>
CLINIC	CLINIC	1-2
01 Vancouver		
02 Surrey		
03 Kamloops		
04 Victoria		
05 Kelowna		
06 Burnaby		
07 Prince George		
STUDY IDENTIFICATION NO	ID	3-9
DIAGNOSTIC GROUP	GROUP	10
1 Cases (Breast Cancer)		
2 Control Group 1 - (Abnormal Mammogram: No Breast Cancer)		
3 Control Group 2 - (Normal Mammogram: Healthy Subjects)		
PARTICIPATION STATUS	PART	11
0 Non-Participant		
1 Participant		
REASON FOR NON-PARTICIPATION	NPRATNL	12
0 Not Applicable - Participant		
1 Language - Unable to Complete Questionnaire Independently		
2 Age - Inability to Remember Exposure Information		
3 Poor Health - Self or Other Family Member		
4 Refusal - Not Interested		
5 Refusal - No Reason Given		
6 Cannot Locate/Moved		
7 Initial Agreement But Lost- to-Followup		
8 Cannot Complete Questionnaire - Going Away on Vacation		
9 Dead		
TIME BETWEEN ADMINISTRATION OF QUESTIONNAIRE ONE AND TWO (months)	TIMEQ1Q2	13-14
[Date of Initial Visit to Clinic - Approximate Date when Q2 returned]		
00 Not Applicable - Non-Participant		

## MARITAL STATUS

1MARITAL

15

2MARITAL

16

1-Questionnaire 1

2-Questionnaire 2

- 0 Never Married
- 1 Married
- 2 Common-Law
- 3 Widowed
- 4 Separated
- 5 Divorced
- 6 Not Stated/Missing Data

## RACIAL ORIGIN

1RACIAL

17

2RACIAL

18

- 0 White
- 1 Chinese
- 2 Japanese
- 3 Native Indian
- 4 East Indian
- 5 Korean
- 6 Persian
- 7 Black
- 8 Not Stated
- 9 Filipino

## COUNTRY OF BIRTH

BIRTHLOC

19-20

- 01 Canada
- 02 USA
- 03 British Isles
- 04 Scandinavia
- 05 Western Europe
- 06 Eastern Europe
- 07 Hong Kong
- 08 Japan
- 09 Mainland China
- 10 Mexico/Central America
- 11 South America
- 12 Caribbean
- 13 Africa
- 14 Pacific Rim - ANZA
- 15 Korea
- 16 India

## PRESENT OCCUPATION

OCCUPNOW

23-24

- 03 Housewife
- 04 Student
- 06 Unemployed/Retired
- 11 Manager/Administrator
- 21 Natural Sciences, engineers,  
mathematicians
- 23 Social Sciences
- 25 Religion
- 27 Teaching and Related

31 Medicine and Health  
 33 Artistic, Literary and  
 Recreational  
 41 Clerical  
 51 Sales  
 61 Service  
 71 Farming and Horticulture  
 82 Food, Beverage and Related  
 Processing  
 91 Motor Transport Operations  
 88 Other Occupations Not Specified  
 99 Unknown/Not Stated

PREVIOUS OCCUPATION OCCUPPRE 25-26

03 Housewife  
 04 Student  
 06 Unemployed/Retired  
 11 Manager/Administrator  
 21 Natural Sciences, engineers,  
 mathematicians  
 23 Social Sciences  
 25 Religion  
 27 Teaching and Related  
 31 Medicine and Health  
 33 Artistic, Literary and  
 Recreational  
 41 Clerical  
 51 Sales  
 61 Service  
 71 Farming and Horticulture  
 82 Food, Beverage and Related  
 Processing  
 91 Motor Transport Operations  
 88 Other Occupations Not Specified  
 99 Unknown/Not Stated

PRESENT WORK HISTORY WORKSTAT 27

0 Presently Retired  
 1 Presently Employed  
 3 Housewife  
 4 Student  
 6 Unemployed

EDUCATION 1EDUC 28-29  
 2EDUC 30-31

00-13 Some Elementary  
 (Highest Grade)  
 20 Post-Secondary (Certificate/  
 Diploma)  
 30 Post-Secondary (Some College/  
 University)  
 40 Post-Secondary (Bachelor's  
 Degree)

## EDUCATION (cont'd)

50 Post-Secondary (Graduate Degree)

60 Not Stated/Missing Data

AGE (Years) AGE 32-33

HEALTH SEEKING BEHAVIOUR/ HSBMOTIV 34

MEASURE OF SUBJECT MOTIVATION

0 No Health Seeking Behaviour (HSB)/Low Motivation

1 Moderate HSB/Moderate Motivation

2 Strong HSB/Strong Motivation

BREAST SELF EXAMINATION 1BSE 35

2BSE

0 Does Not Practice Breast Self Examination (BSE)

1 Practices BSE

9 Not Stated/Missing Data

FREQUENCY OF BSE 1FREQBSE 37-39

2FREQBSE

000 Nil/Does Not Practice BSE

001-365 (Valid Values)

999 Not Stated/Missing Data

MAMMOGRAM FREQUENCY 1MAMFREQ 43-44

2MAMFREQ 45-46

00 Nil

01-98 Yes

99 Unknown/Not Stated

DATE OF MOST RECENT MAMMOGRAM 1DATEMAM 47-50

2DATEMAM 51-54

(mm/yr)

0000 Not Applicable/Nil Mammogram

9999 Not Stated/Missing Data

Leave Columns 55 to 80 Blank - Record One Complete

RECORD 2 AND 4 - RESPONSES TO QUESTIONS RELATED TO MEDICAL,  
REPRODUCTIVE AND INDIVIDUAL EXPOSURE  
QUESTIONS

FREQUENCY OF NEEDLE ASPIRATIONS 1FREQASP 01-02

00 Nil

2FREQASP

01-98 Yes

99 Unknown/Not Stated/Missing Data

AGE AT FIRST ASPIRATION (Years)	1AGE1ASP 2AGE2ASP	03-04
88 Not Applicable/ Nil Aspirations		
99 Unknown/Not Stated/Missing Data		
Valid Values - Age at Menarche to Current Age		
AGE AT MENSTRUATION	1MENARCH 2MENARCH	05-06
99 Not Stated/Missing Data		
Valid Values - 08-39 years		
MENSTRUAL/MENOPAUSE HISTORY	1STAMENO 2STAMENO	07
0 Menopausal		
1 Still Menstruating		
9 Not Stated/Missing Data		
AGE AT MENOPAUSE (Years)	1AGEMENO 2AGEMENO	08-09
88 Not Applicable/Still Menstruating		
99 Unknown/Not Stated/ Missing Data		
TYPE OF MENOPAUSE	1TYPMENO 2TYPMENO	10
1 Natural		
2 Surgical Hysterectomy With Bilateral Oophrectomy		
3 Surgical Hysterectomy Both Ovaries Intact		
4 Not Applicable/Still Menstruating		
9 Unknown/Not Stated/Missing Data		
KNOWLEDGE THAT PERIOD WAS COMING/PRESENCE OF SOMATIC CHANGES BEFORE MENSTRUAL CYCLE	1KNOPER 2KNOPER	11
0 No/Nil Body Changes Associated With Commencement of Cycle		
1 Sometimes		
2 Always		
3 Uncertain		
9 Not Stated/Missing Data		
BREAST SWELLING PERIOD PERIOD	1BRSWPER 2BRSWPER	12
1 Yes		
2 No		
9 Unknown/Not Stated/Missing Data		

AGE WHEN SWELLING STARTED (Years)	1AGESWST 2AGESWST	13-14
88 Not Applicable/No Breast Swelling		
99 Unknown/Not Stated/Missing Data		
 PRESENT HISTORY OF BREAST SWELLING	 1PRSTASW 2PRSTASW	 15
0 Still Experiences Breast Swelling		
1 Breast Swelling Has Stopped		
8 Not Applicable/Nil History of Breast Swelling		
9 Unknown/Not Stated/Missing Data		
 AGE WHEN BREAST SWELLING STOPPED (Years)	 1AGESWST 2AGESWST	 16-17
88 Not Applicable/ No Breast Swelling		
99 Unknown/Not Stated/Missing Data		
 HISTORY OF BREAST PAIN/TENDERNESS	 1HXBPTEN 2HXBPTEN	 18
0 No History of Breast Pain and Tenderness		
1 History/Presence of Breast Pain and Tenderness		
9 Unknown/Not Stated/Missing Data		
 SYMPTOMS ASSOCIATED WITH BREAST PAIN AND TENDERNESS	 1SYMPTOM 2SYMPTOM	 19
0 Pain Alone		
1 Tenderness Alone		
2 Tenderness With Pain		
8 Not Applicable/No History of Breast Pain and Tenderness		
9 Unknown/Not Stated/Missing Data		
 BREAST PAIN AND TENDERNESS PERIOD RELATED	 1PTENPER 2PTENPER	 20
0 Pain/Tenderness Not Period Related		
1 Pain/Tenderness Period Related		
8 Not Applicable/ No Pain/ Tenderness		
9 Unknown/Not Stated/Missing Data		
 AGE WHEN PAIN AND TENDERNESS STARTED (Years)	 1AGEPTST 2AGEPTST	 21-22
88 Not Applicable/No History of Breast Pain and Tenderness		
99 Unknown/Not Stated/Missing Data		

PRESENT STATUS OF BREAST PAIN AND TENDERNESS	1PRSTAPT 2PRSTAPT	23
0 Pain and /tenderness has not Stopped/Persists		
1 Pain and Tenderness Has Stopped		
8 Not Applicable/No History of Breast Pain and Tenderness		
9 Unknown/Not Stated/Missing Data		
AGE WHEN PAIN AND TENDERNESS STOPPED	1AGPTSTP 2AGPTSTP	24-25
(Years)		
55 Not Applicable/Still Having Breast Pain/Tenderness		
88 Not Applicable/No History of Breast Pain and Tenderness		
99 Unknown/Not Stated/Missing Data		
PREGNANCY HISTORY	1HXPREG 2HXPREG	26
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
NUMBER OF PREGNANCIES	1NUMPREG 2NUMPREG	27-28
88 Not Applicable/Nulliparous		
99 Unknown/Not Stated/Missing Data		
AGE AT FIRST FULL TERM PREGNANCY	1AGEFFTP 2AGEFFTP	29-30
(Years)		
88 Not Applicable/Nulliparous/ Never Pregnant		
99 Unknown/Not Stated/Missing Data		
ORAL CONTRACEPTIVE HISTORY	1HXBCP 2HXBCP	31
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
AGE WHEN ORAL CONTRACEPTIVES	1AGBCPST 2AGBCPST	32-33
(Years)		
88 Not Applicable/Nil Use		
99 Unknown/Not Stated/Missing Data		
PRESENT STATUS ORAL CONTRACEPTION	1PRSTBCP 2PRSTBCP	34
0 Not Applicable/Never Used		
1 Still Using		
2 Stopped Using		
9 Unknown/Not Stated/Missing Data		

AGE WHEN ORAL CONTRACEPTIVES STOPPED (Years) 00 Not Applicable/Never Used 88 Not Applicable/Still Using 99 Unknown/Not Stated/Missing Data	1AGOCSTP 2AGOCSTP	35-36
DURATION ORAL CONTRACEPTIVE USE 00 Not Applicable/Never Used 99 Unknown/Not Stated/Missing Data	1DURBCP 2DURBCP	37-38
HISTORY OF ESTROGEN THERAPY 0 No 1 Yes 9 Unknown/Not Stated/Missing Data	1USEEST 2USEEST	39
AGE WHEN ESTROGEN THERAPY STARTED (Years) 00 Not Applicable/Never Used 99 Unknown/Not Stated/Missing Data	1AGSTEST 2AGSTEST	40-41
PRESENT STATUS ESTROGEN THERAPY 0 Not Applicable/Never Used 1 Still on Estrogen Therapy 2 No Longer on Estrogen Therapy 9 Unknown/Not Stated/Missing Data	1PRSTEST 2PRSTEST	42
AGE WHEN ESTROGEN THERAPY STOPPED (Years) 00 Not Applicable/Never Used 01 Not Applicable/Still Using 99 Unknown/Not Stated/Missing Data	1AGESSTP 2AGESSTP	43-44
DURATION ESTROGEN THERAPY 00 Not Applicable/Never Used 99 Unknown/Not Stated/Missing Data	1DUREST 2DUREST	45-46
MATERNAL HISTORY BREAST CANCER 0 No 1 Yes 9 Unknown/Not Stated/Missing Data	1MAHXBCA 2MAHXBCA	47
DIAGNOSIS MATERNAL CANCER BEFORE AGE 50 YEARS 0 No 1 Yes 8 Not Applicable/No Maternal Hx 9 Unknown/Not Stated/Missing Data	1DXBEF50 2DXBEF50	48

HISTORY BILATERAL BREAST CANCER	1MABILAT 2MABILAT	49
0 No/Unilateral Breast Cancer		
1 Yes/Bilateral Breast Cancer		
8 Not Applicable/No Maternal Hx		
9 Unknown/Not Stated/Missing Data		
NUMBER OF SISTERS	1NUMSIS 2NUMSIS	50-51
99 Unknown/Not Stated/Missing Data		
FAMILIAL HISTORY IN SISTERS	1HXSIBCA 2HXSIBCA	52
0 No		
1 Yes		
8 Not Applicable/No Sisters		
9 Unknown/Not Stated/Missing Data		
NUMBER OF SISTERS WITH BREAST CANCER	1NUMSICA	53-54
00 Not Applicable/No Hx Breast Cancer Among Sisters		
88 Not Applicable/No Sisters		
99 Unknown/Not Stated/Missing Data		
SISTERS DIAGNOSED BEFORE 50 YEARS	1NUMBE50 2NUMBE50	55-56
00 Not Applicable/No Hx Breast Cancer Among Sisters		
88 Not Applicable/No Sisters		
99 Unknown/Not Stated/Missing Data		
SISTERS WITH BILATERAL BREAST CANCER	1NUBILAT 2NUBILAT	57-58
00 Not Applicable/No Breast Cancer Among Sisters		
88 Not Applicable/No Sisters		
99 Unknown/Not Stated/Missing Data		
SMOKING HISTORY	1HXSMOK 2HXSMOK	59
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DAILY CIGARETTE CONSUMPTION	1NUMCIGS 2NUMCIGS	60-61
00 Not Applicable/Nil/Does Not Smoke		
99 Unknown/Not Stated/Missing Data		

AGE WHEN STARTED SMOKING (Years)	1AGCIGST 2AGCIGST	62-63
00 Not Applicable/Does Not Smoke		
99 Unknown/Not Stated/Missing Data		
PRESENT SMOKING HISTORY	1STSMOHX 2STSMOHX	64
0 No Longer Smokes/Quit		
1 Yes/Still Smokes		
8 Not Applicable/Never Smoked		
9 Unknown/Not Stated/Missing Data		
AGE WHEN SMOKING STOPPED (Years)	1AGSMSTP 2AGSMSTP	65-66
00 Not Applicable/Never Smoked		
88 Not Applicable/Still Smoking		
99 Unknown/Not Stated/Missing Data		
DURATION - SMOKING HISTORY	1DURSMOK 2DURSMOK	67-68
00 Not Applicable/Never Smoked		
99 Unknown/Not Stated/Missing Data		
HISTORY OF ALCOHOL CONSUMPTION	1ALCOLHX 2ALCOLHX	69
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DAILY ALCOHOL CONSUMPTION	1DRINKS 2DRINKS	70-72
000 Not Applicable/Non-Drinker		
999 Unknown/Not Stated/Missing Data		
AGE WHEN STARTED DRINKING (Years)	1AGSTALC 2AGSTALC	73-74
00 Not Applicable/Non-Drinker		
99 Unknown/Not Stated/Missing Data		
CONSUMPTION - BEER	1BEER 2BEER	75
0 No		
1 Yes		
8 Not Applicable/Non-Drinker		
9 Unknown/Not Stated/Missing Data		
CONSUMPTION - WINE	1WINE 2WINE	76
0 No		
1 Yes		
8 Not Applicable/Non-Drinker		
9 Unknown/Not Stated/Missing Data		

CONSUMPTION - LIQUOR	1LIQUOR 2LIQUOR	77
0 No		
1 Yes		
8 Not Applicable/Non-Drinker		
9 Unknown/Not Stated/Missing Data		
PRESENT STATUS ALCOHOL CONSUMPTION	1PRSTALC 2PRSTALC	78
0 No Longer Drinks		
1 Still Drinks		
8 Not Applicable/Non-Drinker		
9 Unknown/Not Stated/Missing Data		
AGE WHEN STOPPED DRINKING	1AGALSTP 2AGALSTP	79-80
00 Not Applicable/Non-Drinker		
88 Not Applicable/Still Drinks		
99 Unknown/Not Stated/Missing Data		
RECORD 3 AND 5 - RESPONSES TO QUESTIONS RELATED TO MEDICAL, REPRODUCTIVE AND INDIVIDUAL EXPOSURE QUESTIONS		
DURATION OF ALCOHOL CONSUMPTION	1DURALC 2DURALC	01-02
00 Not Applicable/Non-Drinker		
88 Not Applicable/Still Drinks		
99 Unknown/Not Stated/Missing Data		
HEIGHT	1HEIGHT 2HEIGHT	03-05
999 Unknown/Not Stated/Missing Data		
Valid Values - 050-250 cm		
WEIGHT	1WEIGHT 2WEIGHT	06-08
999 Unknown/Not Stated/Missing Data		
Valid Values - 030-150 cm		
QUETELET'S INDEX	1QINDEX 2QINDEX	09-10
2		
[weight (kg)/height (m )]		
99 No Information To Calculate		

## RECORD 6 - RESPONSES TO VALIDITY SCALE EXPOSURE ITEMS

HISTORY OF DYSMENORRHEA BEFORE	HXDYSB40	01
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
TRAUMATIC INJURY(IES) TO BREAST/ CHEST AREA	HXTRMINJ	02
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
BREASTFEEDING HISTORY	HXBRFEED	03
0 No		
1 Yes		
8 Not Applicable/No Children		
9 Unknown/Not Stated/Missing Data		
DURATION OF BREASTFEEDING (months)	DURBFEEED	04-05
00 Not Applicable/Did Not Breastfeed		
88 Not Applicable/No Children		
99 Missing Data/Cannot Calculate		
HISTORY OF MASTITIS	HXMAST	06
0 History of Breastfeeding But No Mastitis		
1 Yes		
5 Not Applicable/Did Not Breastfeed		
8 Not Applicable/No Children		
9 Unknown/Not Stated/Missing Data		
ORAL CONTRACEPTION HISTORY	HXBCP	07
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DURATION - ORAL CONTRACEPTION	DURBCP	08-10
000 Not Applicable/Never Used		
999 Unknown/Not Stated/Missing Data		
HISTORY OF ESTROGEN THERAPY	USEEST	11
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DURATION ESTROGEN THERAPY (months)	DUREST	12-14
000 Not Applicable/Never Used		
999 Unknown/Not Stated/Missing Data		
Valid Values - 001-998		

MATERNAL EXPOSURE TO DES - DIETHYSTILBESETROL 0 No 1 Yes 9 Unknown/Not Stated/Missing Data	MATDES	15
HISTORY OF UNDERACTIVE THYROID 0 No 1 Yes 9 Unknown/Not Stated/Missing Data	UNACTTHY	16
USE OF THYROID MEDICATION 0 No - Condition Existed But No Medication taken 1 Yes - Medication Taken 8 Not Applicable/No Condition/No Medication Required 9 Unknown/Not Stated/Missing Data	THYMED	17
DURATION THYROID MEDICATION 000 Nil - Condition Existed But No Medication Required or Taken 999 Unknown/Not Stated/Missing Data	DUTHYMED	18-20
HISTORY OF DYSFUNCTIONAL STRESS 0 No 1 Yes 9 Unknown/Not Stated/Missing Data	DYSTRESS	21
EXPOSURE TO CHEMICAL WEED KILLERS 0 No 1 Yes 9 Unknown/Not Stated/Missing Data	CHEMWEED	22
FREQUENCY OF EXPOSURE TO CHEMICAL WEED KILLERS 0 Nil 1 Rarely 2 Occasional 3 Regular 4 Frequent 9 Unknown/Not Stated/Missing Data	FREQCWK	23
CHEMICAL EXPOSURE - PAINTS/ SOLVENTS 0 No 1 Yes 9 Unknown/Not Stated/Missing Data	PAINSOL	24

FREQUENCY OF EXPOSURE TO PAINTS/ SOLVENTS	FREQPS	25
0 Nil		
1 Rarely		
2 Occasional		
3 Regular		
4 Frequent		
9 Unknown/Not Stated/Missing Data		
EXPOSURE TO UREA FORMALDEHYDE	UREAFORM	26
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DURATION UREA FORMALDEHYDE EXPOSURE (Years)	DURUREA	27-28
00 Not Applicable/No Exposure		
99 Unknown/Not Stated/Missing Data		
ASBESTOS EXPOSURE	ASBESTOS	29
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
FREQUENCY OF EXPOSURE TO ASBESTOS	FREQASB	30
0 Nil		
1 Rarely		
2 Occasional		
3 Regular		
4 Frequent		
9 Unknown/Not Stated/Missing Data		
DRUG EXPOSURE - RESERPINE	RESERP	31
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DRUG EXPOSURE - ELAVIL	ELAVIL	32
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DRUG EXPOSURE - MELLAL	MELLAL	33
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DRUG EXPOSURE - STEROIDS	STEROID	34
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		

SMOKING HISTORY	HXSMOK	35
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DAILY CIGARETTE CONSUMPTION	NUMCIG	36-38
000 Not Applicable/Non-Smoker		
999 Unknown/Not Stated/Missing Data		
DURATION SMOKING HISTORY	DURSMOK	39-40
00 Not Applicable/Non-Smoker		
99 Unknown/Not Stated/Missing Data		
HISTORY OF ALCOHOL CONSUMPTION	ALCOLHX	41
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
CONSUMPTION - DRINKS PER MONTH	DRINKS	42-44
000 Not Applicable/Non-Drinker		
999 Unknown/Not Stated/Missing Data		
DURATION ALCOHOL CONSUMPTION	DURALCOL	45-46
00 Not Applicable/Non-Drinker		
99 Unknown/Not Stated/Missing Data		
SUNBATHING HISTORY	SUNBATHE	47
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
FREQUENCY OF SUNBATHING	FREQSB	48
0 Nil		
1 Occasional		
2 Regular		
3 Frequent		
9 Unknown/Not Stated/Missing Data		
OVEREXPOSURE TO ULTRAVIOLET	OVEREUV	49
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
DIETARY FAT CONSUMPTION	DIETFAT	50
1 Low		
2 Normal		
3 High		
CONSUMPTION OF NITRATES	NITRATES	51
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		

CONSUMPTION MONOSODIUM GLUTAMATE	MSG	52
0 No		
1 Yes		
9 Unknown/Not Stated/Missing Data		
CONSUMPTION - ARTIFICIAL SWEETNERS	ARTSWEET	53
0 Never		
1 Occasional		
2 Regular		

## Appendix 9: Power Calculations

Reference: Schlesselman, JJ. Case-Control Studies: Design Conduct and Analysis. New York: Oxford University Press, 1982, pp. 150-152.

### Definitions of Terms Used:

$p_0$  - the relative frequency of exposure among the controls = 0.25

R - the odds ratio detectable = 1.6

alpha ( $\alpha$ ) = 0.05 (two-sided);  $Z_\alpha = 1.96$

beta ( $\beta$ ) = 0.15

n = number of cases recruited for the study = 238

c = 2 (number of controls per case)

**The power associated with 238 cases and two controls per case is calculated as follows:**

$$p_1 = p_0 R / 1 + p_0 (R - 1) = (.25)(1.6) / 1 + (.25)(1.6 - 1) = .348$$

$$(p_1 - p_0) = (.348 - .25) = .098$$

$$p = 1/2 (p_0 + p_1) = 1/2 (.25 + .348) = .299$$

$$q = (1 - p) = (1 - .299) = .701$$

$$\bar{p}' = (p_1 + c p_0) / (1 + c) = .348 + 2(.25) / (1 + 2) = .283$$

$$\bar{q}' = (1 - \bar{p}') = (1 - .283) = .717$$

$$\hat{z}_p = [ n (p_1 - p_0)^2 / (1 + 1/c) \bar{p}' \bar{q}' ]^{1/2} - z_\alpha$$

$$= [ 238 (.098) / (1.5) (.283) (.717) ]^{1/2} - 1.96$$

$$= 0.74$$

$$\text{Power} = (1 - \beta) = P(Z \leq \hat{z}_p) = 76\%$$

Therefore, with 238 cases, and two controls per case, the study power to detect a relative risk of 1.6 with a prevalence of exposure in the controls equal to 0.25 is 76%.