

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI[®]
800-521-0600

**ANALYSIS OF THE COMBINED PRIMARY DATA FROM CASE-
CONTROL STUDIES OF RESIDENTIAL RADON AND LUNG CANCER:
A PILOT STUDY OF THREE NORTH-AMERICAN SITES**

Vanessa Spurll Catalan

Department of Epidemiology and Biostatistics

McGill University, Montreal

May 1998

A Thesis

**presented to Faculty of Graduate Studies and Research
in partial fulfilment of the requirements of the degree of
Doctor of Philosophy**

© Vanessa Spurll Catalan 1998



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-44380-9

Canada

ABSTRACT

INTRODUCTION: Radon gas is a naturally occurring, radioactive contaminant of indoor air. One of only 49 agents classified as a 'known human carcinogens', radon was identified in 1988 by the US Environmental Protection Agency and assistant surgeon general of the US Public Health Service as a leading cause of lung cancer, second only to smoking. Based on extrapolations from occupational epidemiologic studies in underground miners, as many as 10-14 % of lung cancer deaths may be attributable to radon exposure. Such extrapolations involve many sources of uncertainty, however, some of these relating to dosimetric differences between mines and homes. Unfortunately, the individual results of the eight major completed studies investigating residential exposures have been equivocal. Inadequate statistical power is a substantial problem in their interpretation, particularly in the context of gaps and errors in exposure measurement that result in an underestimate of a relative effect already anticipated to be small. To mitigate this problem, an analysis of the combined primary data from the three completed North American studies has been undertaken.

METHODS: Individual-level data were received from Winnipeg, Canada, and New Jersey and Missouri, USA and compiled into a common format. Gaps in exposure history were imputed using the observed control mean, a method which, by virtue of its Berkson error structure, does not create bias under a linear model of relative risk. Radon exposure was expressed both in cumulative working level months (WLM) and as average intensity in Becquerels per cubic metre, with separate weighting factors applied to radon exposure in each of three time windows (5 - 15 years, 15 - 25 years, and 25 - 50 years) prior to subject ascertainment. Parallel logistic regression analyses, employing over seventy distinct models and four alternative approaches for the control of tobacco smoke exposure, were undertaken to identify a common model for the prediction of lung cancer rate ratio which would adjust for all matching variables and risk factors (including potential confounders) determined to be important in at least one site. Next, a combined analysis utilized a random-effects model

to obtain the most realistic assessment to date of the rate ratio of lung cancer as a function of residential radon exposure. Sensitivity of the results to exposure time window definition and statistical model form (exponential logistic versus additive excess relative risk) was also investigated.

RESULTS: Independent predictors of lung cancer included intensity of cigarette smoking (whose effect varied by recency of smoking and by source of information), gender, ethnicity, family income, and education. Age at ascertainment (a matching variable in all sites), age at initiation of smoking (a significant predictor only in New Jersey), and environmental tobacco smoke exposure were also retained in the final common model. Estimated arithmetic mean (standard error) cumulative exposures among population controls for the period 5 to 50 years prior to subject ascertainment were 4.6 (2.6), 10.2 (6.3), and 16.7 (9.9) weighted WLMs in New Jersey, Missouri, and Winnipeg, respectively. During this same exposure period, the estimated rate ratio [and 95% confidence interval] was 0.94 [0.52, 1.72] per increment of 10 weighted WLMs radon exposure, and 0.98 [0.84, 1.14] per increment of 100 Bq/m³ weighted average radon concentration. No or negligible differences were observed in the estimated radon effect between sites. Specific recommendations were made for the implementation of the Phase II combining effort that will follow this pilot analysis.

CONCLUSIONS: At the observed levels of radon exposure, the log-linear and linear excess relative risk models provided indistinguishable fits to the data. Modification of the weighting factors, which effectively truncated the exposure time window to 5 - 15 or 5 - 25 years prior to ascertainment, did not affect radon effect estimates. Statistical and other sources of uncertainty, although substantially reduced compared with the individual studies, preclude a definitive conclusion with respect to the magnitude of effect of residential radon on lung cancer. Improved precision to estimate lung cancer risk as a function of residential radon exposure will be achieved with the incorporation of data from four North American studies for which primary data are still being collected.

RÉSUMÉ

INTRODUCTION: Le gaz radon est un contaminant radioactif naturel de l'air intérieur. Il fait partie de la liste des 49 agents cancérigènes humains reconnus, et a été identifié en 1988 par la "US Environmental Protection Agency and assistant surgeon general of the US Public Health Service" comme la principale cause du cancer du poumon après le tabagisme. Des extrapolations faites sur la base de recherche en santé au travail sur les mineurs de fond révèlent que 10 à 15% des décès par cancer du poumon pourraient être attribuables à l'exposition au radon. Ces extrapolations souffrent cependant d'un grand nombre d'incertitudes, dont certaines sont liées aux différences dosimétriques entre les mines et les résidences. Malheureusement, les résultats individuels des huit études majeures achevées à date sur les expositions résidentielles sont équivoques. Leur interprétation souffre de pouvoirs statistiques inadéquats résultant en grande partie de données manquantes et d'erreurs dans les mesures d'exposition. Ceci entraîne la sous-estimation d'un risque relatif qu'on anticipe déjà faible. Pour palier à ce problème, l'auteur a entrepris une analyse des données de base combinées pour les trois études nord-américaines achevées à date.

MÉTHODES: Les données individuelles reçues de Winnipeg (Canada), du New Jersey (É-U) et du Missouri (É-U) ont été compilées dans un format commun. Les données manquantes sur l'historique d'exposition ont été remplacées en utilisant la moyenne observée du groupe témoin. Cette méthode, en vertu de sa structure d'erreur de type Berkson, ne crée pas d'erreur systématique dans le cadre d'un modèle linéaire de risque relatif. L'exposition au radon a été exprimée à la fois en "Working Level Months (WLM)" cumulés et en intensité moyenne (Bq/m^3). Des facteurs de pondération séparés ont été appliqués aux expositions au radon dans chacune des trois fenêtres de temps suivantes (5-15 ans, 15-25 ans et 25-50 ans) avant la date du début de l'observation. Des analyses parallèles de régression logistique employant plus de 70 modèles distincts et quatre approches différentes pour le contrôle de l'exposition à la fumée du tabac ont été menées afin d'identifier un modèle commun pour la

prédiction du ratio des taux de cancer du poumon. Ce modèle tient compte de toutes les variables d'appariement et de tous les facteurs de risque dont l'importance a été démontrée pour au moins un des sites. Par la suite, une analyse combinée employant un modèle d'effets aléatoires a été réalisée pour obtenir l'évaluation la plus réaliste à date du ratio de taux de cancer du poumon en fonction de l'exposition au radon résidentiel. La sensibilité des résultats en fonction des fenêtres du temps d'exposition et de la forme du modèle statistique a aussi été étudiée.

RÉSULTATS: Les variables prédictives indépendantes du cancer du poumon incluent le nombre de cigarettes fumées par jour (dont l'effet estimé varie avec le temps depuis que le sujet a cessé de fumer et avec la source d'information), le sexe, l'ethnie, le revenu familial et le niveau d'éducation. L'âge à la date du début de l'observation (variable d'appariement pour tous les sites), l'âge d'initiation à la cigarette (variable de prédiction seulement au New Jersey) et l'exposition passive à la fumée du tabac ont aussi été retenus dans le modèle commun final. Les moyennes arithmétiques estimées des expositions cumulatives pour la période de 5 à 50 ans avant la date de début de l'observation du sujet sont respectivement égales à 4.6, 10.2, et 16.7 WLMs pour le New Jersey, le Missouri, et Winnipeg. Les déviations standards correspondantes sont 2.6, 6.3, et 9.9. Durant cette même période d'exposition, le ratio de taux estimé est égal à 0.94 pour chaque augmentation de 10 WLMs pondérés d'exposition au radon et 0.98 pour chaque augmentation de 100 Bq/m^3 . Les intervalles de confiance correspondants sont respectivement [0.52, 1.72] et [0.84, 1.14]. Les différences entre les effets estimés du radon à différents sites sont nulles ou négligeables. Des recommandations spécifiques ont été faites pour la mise en place de la phase II de l'effort de combinaison qui suivra cette analyse pilote.

CONCLUSIONS: Au niveaux observés d'exposition au radon, les modèles log-linéaires et linéaires d'excès du risque relatif donnent des résultats semblables. La modification des facteurs de pondération, qui tronquent efficacement la fenêtre de durée d'exposition à 5 - 15

ou 5 - 25 ans avant la date du début de l'observation, n'affecte pas les risques dûs au radon. Les sources d'erreurs statistiques ou autres, bien que réduites par rapport aux études individuelles, empêchent de conclure définitivement quant à la grandeur de l'effet du radon résidentiel sur le cancer du poumon. La précision de l'estimation du risque de cancer du poumon en fonction de l'exposition au radon résidentiel sera améliorée par l'incorporation des données des quatre études nord-américaines en cours.

STATEMENT OF ORIGINALITY

This thesis contains no material which has been previously accepted for the award of any degree or diploma in any university. To the best of my knowledge, this thesis contains no material previously published or written by another person, except where references are made in the context of the thesis.

This thesis is an endeavour in quantitative research synthesis involving the re-analysis of the combined primary data from three case-control studies which examined the relationship between lung cancer and residential exposure to radon. The topic of this thesis was suggested to me by Dr. Daniel Krewski, and the primary investigators of the three included studies provided individual-level data in electronic format. The specific objectives of the study were developed by me. I devised and executed a plan for the analysis of the combined data based on my own critical review of the literature concerning radon and lung cancer, and my knowledge concerning the principles of epidemiology (and published literature relating thereto). My thesis supervisory committee (Drs. James A. Hanley, Daniel Krewski, Lawrence Joseph, and John C. Bailar III) have contributed suggestions, comments, and corrections on the methods that I have used to carry out my analyses and on the documentation of my work.

ACKNOWLEDGMENTS

My supervisory committee brought together the most gifted academics from whom I have had the privilege of learning: Drs. James A. Hanley and Dan Krewski (co-supervisors), Dr. Lawrence Joseph and Dr. John C. Bailar III. If I have grasped but a portion of their shared insight, my academic experience has been immeasurably enriched. To each and every one I express my sincerest gratitude, not only for their generosity of time and energy, but for their special gifts. Dr. Hanley's limitless curiosity and patience, meticulous attention to detail, and keen sense of perspective are a continuing source of awe and inspiration. Dr. Krewski's diplomacy and consistent positivity fueled this effort. The shared talent of Drs. Bailar and Joseph for providing gentle yet incisive critique is one I will always admire and aspire to develop. Dr. Joseph's encouragement buoyed my spirit at times of greatest need.

To the principal investigators Dr. Jan Zielinski, Dr. Ernie Létourneau, Dr. Sandy Wilcox, Dr. Michael Alavanja, and Eric Berger, who provided their data and answered my questions, I express my sincere thanks.

Thanks go also to the staff in the Department of Epidemiology and Biostatistics who shared their space and friendship, and made me feel at home.

I gratefully acknowledge the Eco-Research Fellowship Program of the Tri-Council Secretariate, without whose financial support this work would not have been possible.

To my late father, Geoffrey Spurrll, it is to you that I owe my academic drive and dedication.

To my husband, Dr. Lionel JJ Catalan, I express my deepest gratitude for your presence of mind and warmth of heart.

TABLE OF CONTENTS

ABSTRACT	i
RÉSUMÉ	iii
STATEMENT OF ORIGINALITY	vi
ACKNOWLEDGEMENTS	vii
LIST OF TABLES	xii
LIST OF FIGURES	xv

1 INTRODUCTION

1.1 BACKGROUND	1-01
1.2 RATIONALE	1-04
1.3 PURPOSE	1-05
1.4 OBJECTIVES	1-06
1.4.1 Improve Comparability of Site-Specific Risk Estimates	1-07
1.4.2 Reduce Uncertainty in Current Radon Risk Estimates	1-08
1.4.3 Explore Key Methodological Issues Through Sensitivity Analyses	1-10

2 CRITICAL REVIEW OF THE LITERATURE

2.1 BACKGROUND AND SCOPE OF THE REVIEW	2-01
2.2 OVERVIEW OF EIGHT PUBLISHED CASE-CONTROL STUDIES	2-03
2.2.1 Shenyang China: Blot et al. 1990	2-03
2.2.2 New Jersey USA: Schoenberg et al. 1990	2-03
2.2.3 South Finland: Ruosteenoja 1991	2-04
2.2.4 Stockholm Sweden: Pershagen et al. 1992	2-05
2.2.5 Winnipeg Canada: Létourneau et al. 1994	2-05
2.2.6 Missouri USA: Alavanja et al. 1994	2-06
2.2.7 Sweden: Pershagen et al. 1994	2-06
2.2.8 Finland: Auvinen 1996	2-07
2.3 POTENTIAL FOR SELECTION BIAS	2-08

2.4	POTENTIAL FOR INFORMATION BIAS	2-09
2.4.1	Mis-classification of Disease Status	2-09
2.4.2	Exposure Measurement Error	2-10
2.5	ANALYTIC ISSUES	2-32
2.5.1	Model Form	2-32
2.5.2	Confounding and Effect Modification	2-34
2.6	COHERENCE OF THE EVIDENCE	2-45
2.6.1	Biologic Plausibility and Evidence of Carcinogenicity	2-45
2.6.2	Main Limitations of The Reviewed Studies	2-45
2.6.3	Consistency of The Results	2-49
2.6.4	Conclusions	2-55

3 METHODS

3.1	WORKING DEFINITIONS AND APPROACH	3-01
3.2	SELECTION OF STUDIES FOR PILOT ANALYSIS	3-04
3.3	STUDY POPULATIONS	3-05
3.3.2	Overview of the Included Studies	3-05
3.3.3	Selection of Subjects For Each Study	3-07
3.4	RADON EXPOSURE INFORMATION	3-18
3.5	DATA PREPARATION	3-20
3.5.2	Data Verification and Formatting	3-21
3.5.3	Imputation of Missing Data	3-24
3.6	DEVELOPMENT OF A COMMON LOGISTIC MODEL	3-29
3.6.2	Matching	3-29
3.6.3	Analysis Strategy	3-30
3.6.4	Treatment of Covariates	3-32
3.6.5	Radon Exposure Estimation	3-43
3.6.6	Summary of First Order Candidates for Regression	3-54
3.6.7	Effect Modification	3-56
3.7	COMBINED LOGISTIC ANALYSES	3-58
3.7.1	Fixed Versus Random Effects	3-58
3.7.3	Adequacy of Using Common Coefficients for the Covariates	3-63

3.8	SENSITIVITY ANALYSES	3-64
3.8.1	Time Window Weighting	3-64
3.8.2	Model Form	3-66
 4	 RESULTS	 4-01
4.1	COVARIATE DISTRIBUTIONS AMONGST INTERVIEWED SUBJECTS	4-01
4.2	RADON EXPOSURE HISTORIES	4-11
4.2.1	Home Identification and Radon Monitoring	4-11
4.2.2	Home Radon Levels	4-16
4.2.3	Imputation of Gaps in Radon Exposure History	4-18
4.2.4	Radon Exposure in the 5 -50 Years Prior to Ascertainment ..	4-20
4.3	RELATIONSHIPS AMONGST RADON EXPOSURE, OUTCOME, AND COVARIATES	4-24
4.3.1	Active and Passive Tobacco Smoke Exposure	4-33
4.3.2	Gender	4-34
4.4	BASIC INFORMATION RELATING ACTIVE SMOKING AND LUNG CANCER	4-34
4.5	DEVELOPMENT OF A COMMON LOGISTIC MODEL	4-46
4.5.1	Overview	4-46
4.5.2	Best-Fitting Models Within Site	4-47
4.5.3	Restriction of Subjects	4-51
4.5.4	Stratified Analyses	4-53
4.5.5	The Final Common Logistic Model	4-57
4.6	RESULTS OF THE FINAL ANALYSES	4-60
4.6.1	Predictors of Lung Cancer	4-60
4.6.2	Rate Ratio of Lung Cancer As a Function of Radon in the Default (5 - 50 Year) Exposure Time Window	4-67
4.6.3	Sensitivity of the Pooled Results to Time Window Weighting	4-71
4.6.4	Sensitivity of the Pooled Results to Model Form	4-73
 5	 DISCUSSION	 5-01

5.1	NOVEL ASPECTS OF THE ANALYSIS	5-01
5.2	FINDINGS REGARDING RADON EXPOSURE	5-03
5.2.1	Estimation of Radon Exposure	5-03
5.2.2	Main Results	5-05
5.2.3	Consistency of the Results with Estimates Based on Miner Cohorts	5-06
5.3	META-ANALYSES VERSUS ANALYSES OF COMBINED PRIMARY DATA	5-07
5.3.1	Comparability of Results	5-07
5.3.2	Advantages of Primary Over Secondary Data	5-10
5.3.3	Main Limitations of Review Methods	5-12
5.4	UNCERTAINTIES SPECIFIC TO RESIDENTIAL RADON	5-14
5.5	RECOMMENDATIONS FOR THE PHASE II ANALYSIS OF NORTH AMERICAN CASE-CONTROL STUDIES OF RESIDENTIAL RADON AND LUNG CANCER	5-19
6	SUMMARY AND CONCLUSIONS	6-01
6.1	SUMMARY	6-01
6.2	CONCLUSIONS	6-03
7	REFERENCES	7-01
APPENDIX A Detailed Tables Comparing Design Features of Eight Published Case-Control Studies		
APPENDIX B Common Data Format		
APPENDIX C Development of a Common Logistic Regression Model		
APPENDIX D Covariates Not Associated With Radon Exposure		

LIST OF TABLES

Table 2.1	Sources of Error in the Estimation of Cumulative Indoor Radon-Progeny Exposure	2-15
Table 2.2	Physical and Biologic Factors Influencing Exposure-Dose Relationships	2-27
Table 2.3	Summary of Main Study Limitations	2-46
Table 2.4	Reference Exposure Categories and Associated Results	2-52
Table 2.5	Estimates of Radon Effect Based on A Continuous Measure of Exposure	2-53
Table 3.1	Approaches to Quantitative Research Synthesis	3-01
Table 3.2	Overview of the Included Studies	3-06
Table 3.3	Population Selection	3-16
Table 3.4	Radon Exposure Information	3-19
Table 3.5	Canadian Home Occupancy Data	3-49
Table 3.6	Time Window Weighting for the Default 5 - 50 Year Radon Analysis	3-54
Table 3.7	Candidates for Regression	3-56
Table 3.8	Time Window (and Annual) Weights For Estimation of Cumulative Radon Exposure Under Three Scenarios	3-65
Table 3.9	Time Window (and Annual) Weights For Estimation of <u>Average Radon Concentration</u> Under Three Scenarios	3-66
Table 4.1	Covariate Distributions Amongst Included and Excluded Subjects: New Jersey USA	4-02
Table 4.2	Covariate Distributions Amongst Included and Excluded Subjects: Winnipeg CA	4-04

Table 4.3	Covariate Distributions Amongst Included and Excluded Subjects: Missouri USA	4-07
Table 4.4	Central Tendency and Inter-quartile Range of the Distribution of Radon Concentration (in Bq/m³) Amongst Monitored Control Homes	4-16
Table 4.5	Proportion of Imputed Annual Radon Measures Within Three Exposure Time Windows	4-19
Table 4.6	Characterization of Active Smoking Amongst Cigarette-Smoking Subjects	4-35
Table 4.7	Adjusted Odds Ratios for Lung Cancer By Pack Years of Cigarette Smoking	4-38
Table 4.8	Summary of Odds Ratios for Lung Cancer By Cigarette-Years of Smoking Estimated by Siemiatycki et al., (1994 and 1995)	4-40
Table 4.9	Adjusted Odds Ratios for Lung Cancer By Pack Years of Cigarette Smoking and Restricted to Subjects Who Did Not Smoke in the 13 Years Prior to Ascertainment	4-42
Table 4.10	Adjusted[†] Odds Ratios for Lung Cancer By <u>Recency</u>[°] of Cigarette Smoking	4-44
Table 4.11	Summary of Best-Fitting Logistic Models By Site	4-50
Table 4.12	Covariate Control in Original and Stratified Common Analyses	4-55
Table 4.13	Covariates Included in the Final Common Logistic Model	4-58
Table 4.14	Adjusted Regression Coefficients For Smoking-Related Covariates in the Final Common Logistic Model	4-61
Table 4.15	Adjusted Rate Ratios [95% CI] Associated with Selected Increments of Radon Exposure Based on Exponential and Linear Models	4-75
Table 5.1	Sources of Uncertainty in the Assessment of Lung Cancer Risk as a Function of Residential Radon Exposure	5-15
Table A.1	Subject Selection	A-02

Table A.2	Proxy Data	A-05
Table A.3	Radon Exposure Information	A-07
Table A.4	Levels of Exposure	A-09
Table A.5	Prevention and Control of Confounding	A-10
Table A.6	Effect Modification	A-11
Table A.7	Smoking-Specific Analyses	A-12
Table A.8	Treatment of Age	A-15
Table A.9	Histological Analyses	A-17
Table B.1	Record 1: Covariates	B-01
Table B.2	Record 2: Radon Exposure	B-03
Table C.1	Logistic Models of Types A, B & C	C-03
Table C.2	Logistic Models of Types D and E	C-06

LIST OF FIGURES

Figure 1.1	The Uranium-238 Decay Chain and Radon-222 Decay Series	1-02
Figure 2.1	Simplified Diagram of Ideal Radon Exposure Data	2-12
Figure 2.2	Errors in Radon Exposure Estimation	2-13
Figure 2.3	Odds Ratios Estimated from Eight Case-Control Studies	2-51
Figure 3.1	Subject Inclusion/Exclusion in New Jersey	3-10
Figure 3.2	Subject Inclusion/Exclusion in Winnipeg	3-12
Figure 3.3	Subject Inclusion/Exclusion in Missouri	3-14
Figure 3.4	Common Data	3-22
Figure 3.5	Completeness of Radon Exposure Data	3-46
Figure 3.6	Estimated Home Occupancy in Missouri USA	3-51
Figure 4.1	Number of Homes Identified and Recorded	4-13
Figure 4.2	Number of Homes in Which Radon Monitoring was Completed	4-15
Figure 4.3	Frequency Distribution of Radon Concentration Amongst Monitored Control Homes	4-17
Figure 4.4	Cumulative Frequency Distribution of Cumulative Radon Exposure in the 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months	4-21
Figure 4.5	Cumulative Frequency Distribution of Average Radon Concentration in the 5 - 50 Years Prior to Ascertainment in Weighted Becquerels per Cubic Metre	4-22
Figure 4.6	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Smoking Recency and Outcome	4-25

Figure 4.7	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Smoking <u>Intensity</u> (in Packs Per Day) and Outcome	4-26
Figure 4.8	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by <u>Pack-Years</u> of Smoking and Outcome	4-27
Figure 4.9	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Smoking <u>Recency</u> and Outcome <u>Assuming Non-Gender-Differential Home Occupancy</u>	4-28
Figure 4.10	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Smoking <u>Intensity</u> (in Packs Per Day) and Outcome <u>Assuming Non-Gender-Differential Home Occupancy</u>	4-29
Figure 4.11	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by <u>Pack-Years</u> of Smoking and Outcome <u>Assuming Non-Gender-Differential Home Occupancy</u>	4-30
Figure 4.12	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by <u>Gender</u> and Outcome	4-31
Figure 4.13	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by <u>Gender</u> and Outcome <u>Assuming Non-Gender-Differential Home Occupancy</u>	4-32
Figure 4.14	Adjusted Odds Ratios [95% CI] Associated with Pack-Years of Smoking	4-39
Figure 4.15	Adjusted Odds Ratios [95% CI] Associated with <u>Recency</u> of Smoking	4-45
Figure 4.16	Adjusted Odds Ratios [95% CI] Associated with an Increment of 10 Weighted WLMs Radon Exposure Within Age and Smoking Strata .	4-56
Figure 4.17	Crude and Adjusted Odds Ratios [95% CI] Associated with an Increment of 10 Weighted WLMs and 100 Weighted Bq/m ³ Radon	

	Exposure	4-59
Figure 4.18	Adjusted Odds Ratios and 95% Confidence Intervals for Covariates Other Than Smoking	4-64
Figure 4.19	Adjusted Odds Ratios [95% CI] Associated with an Increment of 10 Weighted WLMs Radon Exposure in the 5 - 50 Years Prior to Ascertainment: Site-Specific and Pooled Results	4-69
Figure 4.20	Adjusted Odds Ratios [95% CI] Associated with an Increment of <u>100 Weighted Bq/m³</u> Radon Exposure in the 5 - 50 Years Prior to Ascertainment: Site-Specific and Pooled Results	4-70
Figure 4.21	Sensitivity of the Pooled Results to Time Window Definition: Adjusted Odds Ratios [95% CI] Associated with Increments of 10 Weighted WLMs and 100 Weighted Bq/m³ of Radon Exposure ..	4-72
Figure 4.22	Sensitivity of the Pooled Results to <u>Model Form</u>: Adjusted Rate Ratios for Radon Exposure in the 5 - 50 Years Prior to Ascertainment	4-74
Figure D.1	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Environmental Tobacco Smoke Exposure (in Pack-Years) and Outcome	D-01
Figure D.2	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by <u>Education</u> and Outcome	D-02
Figure D.3	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by <u>Income</u> and Outcome	D-03
Figure D.4	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by <u>Age at Ascertainment</u> and Outcome	D-04
Figure D.5	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by <u>Ethnicity</u> and Outcome	D-05
Figure D.6	Median and Inter-quartile Range of Radon Exposure 5 - 50 Years	

Prior to Ascertainment in Weighted Working Level Months by
Source of Information and Outcome D-06

Figure D.7 Median and Inter-quartile Range of Radon Exposure 5 - 50 Years
Prior to Ascertainment in Weighted Working Level Months by
Year of Ascertainment and Outcome D-07

1 INTRODUCTION

1.1 BACKGROUND

Radon gas is the largest single source of human exposure to ionizing radiation (Bodansky 1987; NRC 1990). In the United States, the average effective radiation dose from environmental radon is estimated to constitute 55% of all sources (both natural and anthropogenic) of radiation exposure. As illustrated in Figure 1.1, radon¹ is formed as an intermediate product in the radioactive decay of uranium-238 (Bodansky 1987, Robkin 1987). Uranium-238 is ubiquitous in the earth's crust, being present at varying levels in virtually all forms of rocks and soil. The only gaseous intermediary in the uranium-238 decay chain, radon-222 readily diffuses out of the upper layers of the soil and enters homes where it accumulates and remains present as a naturally occurring contaminant of indoor air (Lubin et al. 1994a).

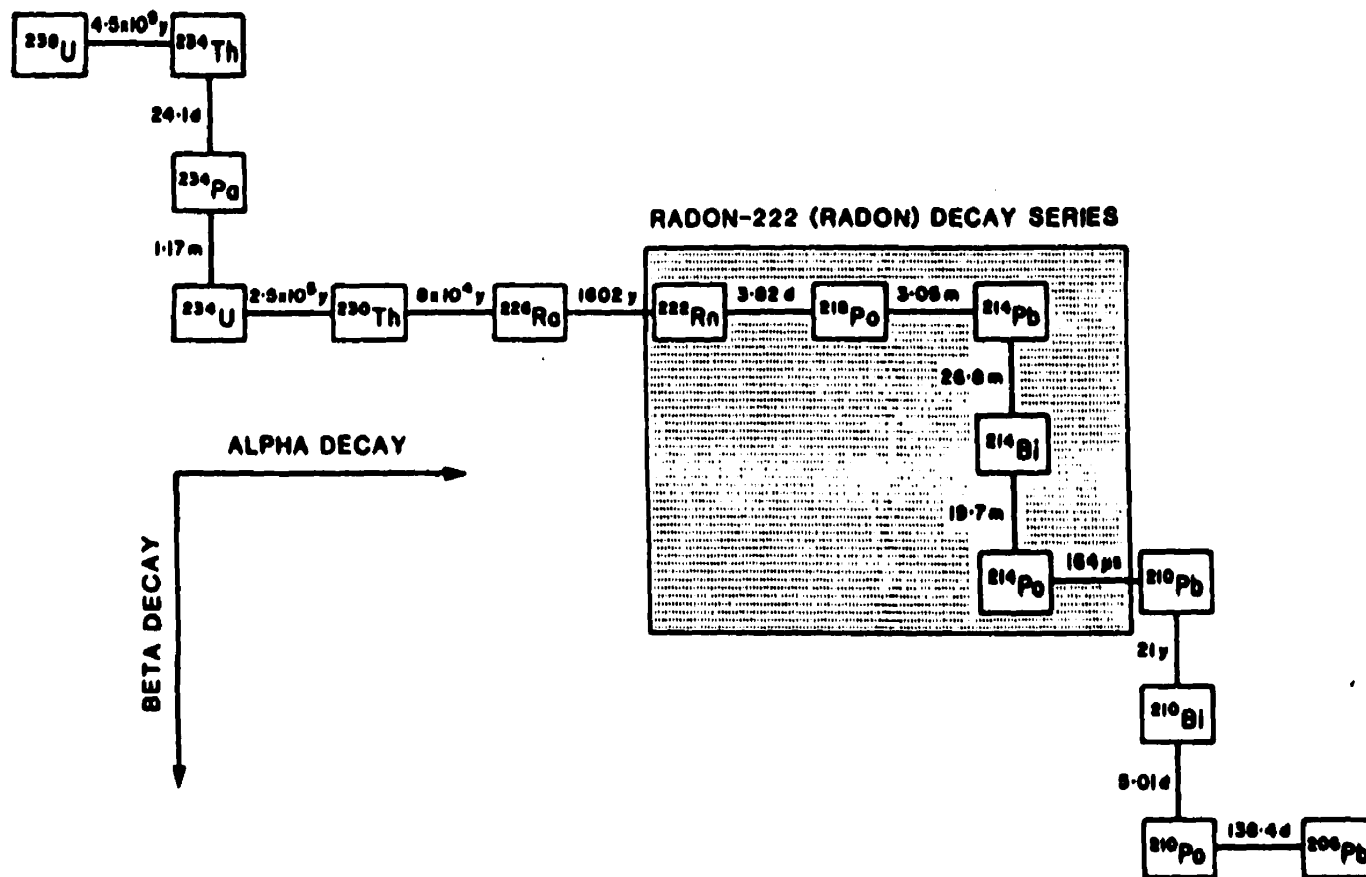
An unreactive noble gas, radon breathed into the lungs is mostly exhaled; only a small fraction decays in the lung or is transferred to the blood (NRC 1988; Robkin 1987). Radon's most important hazard arises from the fact that its decay products (also known as 'daughters' or 'progeny') are solid particles which when themselves inhaled, deposit on the interior surfaces of the lung. As confirmed in experimental studies in rats (Chameaud et al., 1974,

1

Radon designates the gaseous element with atomic number 86. The term 'radon' in its common usage and in the remainder of the text refers to radon-222, the specific isotope of radon with molecular weight 222. Most radon risk assessments (IARC 1988; NCRP 1984; NRC 1988; NRC 1994; Pathak 1989; WHO 1985) have restricted attention to this compound to the exclusion of two other isotopes. Radon-219, also known as 'actinon', is a decay product of uranium-235; radon-220, also known as 'thoron' is a decay product of thorium-232. The contributions of radon-219 and radon-220 to total radon activity have been neglected because of their relative rarity compared to radon-222, and their much shorter half lives (Bodansky 1987, NRC 1988). The half lives of the radon isotopes with molecular weight 219, 220, and 222, respectively, are 4 seconds, 56 seconds, and 3.86 days. The much shorter half lives of radon-219 and radon-220 compared to that of radon-222, results in the much greater tendency of these rarer isotopes to decay to solid progeny before the gas can escape the ground.

Figure 1.1

The Uranium-238 Decay Chain and Radon-222 Decay Series²



1981), the carcinogenic potential of radon derives principally from the emission of alpha particles from radon's short-lived progeny, (polonium-218 and polonium-214), during their own subsequent radioactive decay (Bale 1980).

Owing to their relatively large size, alpha particles have relatively low penetrative capacity³ (Hobbs and McClellan 1986; Holum 1987; Robkin 1987). The protective barrier of the skin renders external alpha radiation harmless (Cole 1993). If alpha particles are emitted internally, however, as for example by solid radon progeny lodged in the lung, the sensitive basal cells lining the bronchial tree may sustain DNA damage. Cellular mutagenesis studies, experimental research in several animal species, and numerous epidemiologic investigations in underground miners, have established radon as a 'known human carcinogen' (IARC 1988)⁴.

Lung cancer risk was first associated with residential radon exposure in 1984 when a nuclear engineer leaving work at a Nuclear Power Plant in Pottstown PA set off the checkpoint radiation alarm (Abate 1992). The source of the radiation was ultimately traced to his home, which demonstrated radon readings as high as those typically found in uranium mines. Epidemiologic studies had already established a link between radon exposure and lung cancer among underground hard-rock miners (Axelson and Sundell 1978; Axelson 1980; Damber and Larsson 1982; De Villers and Windish 1964; De Villers et al. 1971; Edling 1982; Edling and Axelson 1983; Hornung and Samuels 1981; IARC 1988; Jorgenson 1973; Kunz et al. 1979; Morrison et al. 1981; Muller et al. 1983; Piechowski et al. 1981; Sevc et al. 1971; Snihs 1974; Whittemore and McMillan 1983). Prompted by widespread publicity about this new radon risk, the United States (US) Environmental Protection Agency (EPA) began to extrapolate risk estimates based on studies among miners to generate expected numbers of cancers from residential exposures.

³ The typical range of alpha particles in tissues is approximately 5 mm (Robkin 1987).

⁴ See Jostes 1996 for a review of radon's genetic, cytogenetic, and carcinogenic effects.

In 1986 the EPA and assistant surgeon general of the Public Health Service identified residential radon as a leading cause of lung cancer, second only to smoking (Samet & Nero 1989; Smigel 1988). Since then, residential exposure to radon gas has been estimated to result in more cancer deaths than any other source of environmental pollution in the US (Guimond and Page, 1992). Projections from risk models based on a combined analysis of 11 miner cohorts (Lubin 1994; Lubin et al. 1994a) have estimated that 11-13 percent of all lung cancer deaths per year among residents of single family dwellings in the US (about 10% of all lung cancer deaths) may be attributable to indoor radon exposure. In the US, this amounts to 15,000 lung cancer deaths per year, with an approximate range of 6000 to 36000 deaths.

1.2 RATIONALE

The extrapolation of risk estimates based on the experience of miners to the domestic context is subject to numerous sources of uncertainty (Lubin et al. 1994a; Lubin 1994; Puskin 1992; Samet 1992). The most prominent of these include: a) the paucity of accurate information on the smoking habits of miner populations, such information having therefore been excluded from the main models of Lubin et al. (1994a); b) confounding by concomitant exposures to other lung carcinogens present in some mining environments [including: inorganic arsenic (Taylor et al. 1989; Xiang-Zhen et al. 1993), silica (Amandus et al. 1991), chromium (Lubin et al. 1994a), nickel (Lubin et al. 1994a), diesel and blasting fumes, (IARC 1989), and mycotoxins (Bartsch et al. 1996; Lo et al. 1996; Venitt and Biggs 1996; Scram et al. 1993)] which may have resulted in an over-estimate of the level of risk attributable to radon; c) differences between working miners and the general population in duration and intensity of exposure to radon progeny⁵ and in breathing patterns, which are likely to result in a greater risk to miners (NRC 1991); d) differences between the physical states of radon and its

⁵ Although ambient concentrations of radon in homes can reach levels similar to those in mines, the mean home concentrations are much lower.

progeny in mine and home environments, which result in an actual absorbed dose per WLM⁶ exposure perhaps 30% lower in homes than in mines (NRC 1991); and e) the absence of data on the effects in females and children and on long durations of exposure (NRC 1988; NRC 1991). Such uncertainties, and the assumption of a linear no-threshold risk model generated controversy that the US EPA's had over-estimated radon-related risk (Cole 1993; Horgan 1994; Neuberger 1991; Neuberger 1992). Lubin et al. (1994a) cautioned, however, that because of the skewed distribution of radon concentrations in homes, up to 75% of the estimated radon-attributed lung cancers occur in residences where radon levels are below the EPA action level of 4 pCi/L (equivalent to 148 Bq/m³). This level is considerably less than the current Canadian action guideline of 400 Bq/m³ (Krewski et al. 1989).

Since radon exposure is ubiquitous, precise estimates of attributable risk are needed to guide sensible public policy for the prevention of excess cases of lung cancer which may arise from even a slightly elevated relative risk. Extensive worldwide epidemiologic research has recently been undertaken to provide estimates of radon risk from residential exposure, overcoming the uncertainties inherent in the miner data. Unfortunately, the results of those studies published to date have been equivocal. Inadequate statistical power is a substantial problem in their interpretation (Lubin 1995a), particularly in the context of unavoidable "random" (non-differential) errors in exposure measurement that may result in an underestimate of a relative effect already anticipated to be small. There is also a possibility of bias resulting from inappropriate treatment of missing exposure data.

1.3 PURPOSE

In order to reduce uncertainty regarding the risk associated with residential exposure to radon

⁶ Working Level Month (WLM) is the unit of exposure to radon progeny employed in the mining studies. One WLM is equivalent to exposure to 1 working level (WL) for 170 hours (a typical miner's working month). One WL is any combination of radon progeny in 1L of air that ultimately releases 1.3×10^5 MeV of alpha energy during decay. The SI unit for the rate of decay is Bq/m³; assuming 50% equilibrium between radon gas and its progeny, 1 WL is approximately equal to 37 Bq/m³ (IARC 1988).

gas, a combined re-analysis of the raw data from several residentially-based case-control studies was proposed. The use of studies conducted in a residential setting, including a preponderance of females, surmounts each of the uncertainties identified above regarding extrapolation of estimates based on mining studies. The accumulation of a large number of subjects will increase the precision in the final estimate of risk over that of any single study. The use of a common and unbiased method for handling missing exposure data will eliminate uncertainty created by the use of differing methods in the primary studies. Sensitivity analyses will address uncertainty relating to model form.

An international meeting of radon researchers sponsored by the US Department of Energy (US DOE) and Commission of European Communities (CEC) in 1995 resulted in the agreement of North American Principal Investigators to contribute their primary data to a collaborative pooling effort to be coordinated through officials at Health Canada (US DOE/CEC 1995). This thesis constitutes the first and pilot phase of a two stage effort to combine the North American data. The purpose of this thesis is to use the individual level data from the three North American case-control studies already published to provide the most precise estimate to date of the lung cancer risk associated with residential radon exposure while addressing key methodologic issues. The insights based on the pilot analysis will be applied to the phase II analysis, which will incorporate data from four additional investigations in which primary data are still being collected.

A similar pooling endeavour is ongoing in Europe, and lags approximately one year behind the North American effort. The author has had the privilege of sharing preliminary methodologic insights with these researchers at the first European planning workshop.

1.4 OBJECTIVES

The thesis has three main objectives. First, site-specific analyses will increase the comparability of individual study results through a standardized approach to: reference

exposures, covariate control, and treatment of missing variables. The standardization of the analysis will allow an assessment of the consistency of the site-specific results, considered by Clayton (1991) as the over-riding goal of collaborative re-analyses. Second, an analysis of the combined data, where appropriate, will reduce uncertainty, improving on past estimates of risk associated with residential radon through improved statistical precision and generalizability of the findings (Thacker 1988), and appropriate statistical caution; Third: investigation of the sensitivity of results to time window definition and model form (exponential logistic versus linear) will provide insights for application to the Phase II combination of data from all seven North American case-control studies of residential radon and lung cancer. These three advantages are described in more detail in the remainder of this section.

1.4.1 IMPROVE COMPARABILITY OF SITE-SPECIFIC RISK ESTIMATES THROUGH A STANDARDIZED APPROACH TO THE ANALYSIS

1.4.1.1 Approach to Missing Data

The original radon studies employed different strategies in dealing with missing radon exposure data, and these have differing implications for both bias in the estimate of radon effects and the precision of the estimates. The use of a standard approach across data sets will reduce uncertainty from this source.

1.4.1.2 Comparability of Reference Exposure

Since odds ratios (OR) depend on the baseline exposure category to which they are related, and the studies completed to date have all used different baseline categories, the results of these studies are not directly comparable. A pooled analysis using the original exposure data will avoid this problem. An estimate of the OR per unit of radon exposure will be derived within each site, and these estimates combined.

1.4.1.3 Common Constellation of Covariates

A set of candidates for potential inclusion as confounders and effect modifiers will be determined *a priori*; the use of *a priori* covariates is preferred over the use of automated model selection procedures (Greenland & Robins 1985; Miettinen & Cook 1981). Variables will be retained if they : 1) were matching variables in any site; 2) are strong *a priori* risk factors for lung cancer; or, 3) appear to be important predictors of lung cancer in any site. The final model will include this common constellation of covariates, ensuring comparability of the site-specific results.

1.4.2 REDUCE UNCERTAINTY IN CURRENT RADON RISK ESTIMATES

1.4.2.1 Improved Statistical Precision

By increasing the overall number of cases and controls over that available in any single residential study, pooling the estimates from the three studies will increase precision in the estimated rate ratios. The additional data will also help in the investigation of possible effect modifiers (smoking, age, gender, etc.).

1.4.2.2 Appropriate Statistical Caution

While standard ‘fixed effects’ models assume only random variation within studies in estimating what is assumed to be a single fixed effect, ‘random effects’ methods explicitly model random variation both within and between studies. Random effects models assume that the studies selected are representative of some hypothetical population of studies and that the effects estimated in different studies are randomly positioned about some central value (Thompson and Pocock 1991).

Variability in effects between studies can arise either because of: a) inherent difficulties in

estimating a biologically constant effect; b) uncontrolled confounding or other sources of bias; or c) inherently different effects, arising from unidentified or non-standardized 'effect modifiers', if the distribution of such modifiers varies between populations. In the latter two cases, variability is not truly 'random'. Potential confounders and effect modifiers will be thoroughly considered during model development.

Lubin et al.'s (1994a) combined analysis of 11 miner cohorts utilized a 'fixed effects' analysis to combine cohort-specific estimates in spite of significant heterogeneity in these estimates across cohorts; these cohort-specific estimates did not appear to take into account known effect modifiers or to adequately control for possible confounding by smoking, or by mycotoxins which may be more prevalent in some mines than others.

When heterogeneity is present and explicitly accounted for using a 'random effects' model, the confidence interval will be wider than it would have been in a standard 'fixed-effects' analysis (Thompson and Pocock 1991). Random effects models thus introduce a degree of statistical caution, and probably, realism, into the results (Laird and Mosteller 1990; Thompson and Pocock 1991; Whitehead and Whitehead 1991).

The present undertaking will utilize a simplified approach to random effects analysis known as the two stage approach, recently applied to the data from miners (Wang et al. 1996), to explicitly account for possible heterogeneity in radon associated lung cancer risk across the case-control study sites. The results of the two stage random effects analysis will be compared to that of a fixed effects analyses.

1.4.2.3 Increased Generalizability of the Findings

External validity, or generalizability, is the extent to which exposure-outcome relationship observed within the studied population are also true in a wider external population. The external validity of the lung cancer risks from exposure in underground mines is limited by

many sources of uncertainty which threaten their generalizability to residential exposures (NRC 1991). A combination of data from residential radon studies conducted in different North-American sites will provide the most direct and precise estimate to date of radon-induced lung cancer risk among the general population. Residential studies have included women in their target populations and included exposures pertaining to early periods of life - prominent gaps in the miner data-sets. The use of a combining method which explicitly models inter-study variance in the radon effect and a model which adequately accounts for other lung cancer risk factors will improve confidence in their generalizability of the derived risk estimates to the North American population at large.

1.4.3 EXPLORE KEY METHODOLOGICAL ISSUES THROUGH SENSITIVITY ANALYSES

1.4.3.1 Sensitivity of Results to Exposure Time Window Definition

Inappropriate assumptions about the empirical induction period attenuates estimates of exposure effect (Rothman 1981). Although data from underground miners has been used extensively to estimate the relative importance of exposure periods prior to a given age, such occupational exposures were necessarily limited to the time window corresponding to adult employment age; it is unwise to extrapolate such inferences to a wider lifespan. The residential studies provide data up to 50 years prior to subject ascertainment. The relative weighting of radon exposures within this period will be varied through sensitivity analyses to determine the most appropriate of three scenarios defined *a priori*.

1.4.4 Sensitivity Of Results To Model Form

Although recent analyses of miner cohort studies used an additive model of excess relative risk (ERR), all of the case-control residential studies completed to date have assumed a multiplicative logistic model, as is standard in the multivariate analysis of case-control data.

Both are members of the family of generalized relative risk models (Moolgavkar and Venzon 1987). Whereas the logistic model assumes an exponential (or log-linear) increase in the rate ratio of lung cancer as a function of exposure, the ERR model assumes a linear increase. Two residential studies, one in Europe (Pershagen et al. 1992), and one in North-America (Schoenberg et al 1990), used an additive ERR model in addition to their main logistic regression analyses. Although the range of exposures and magnitude of effect under consideration should lead to small differences in the predictions of the two types of models, the present analysis will include a comparison of the two types of models in each study site.

The sensitivity analyses outlined above will not only reduce uncertainty in the estimation of radon risk from the present combined analysis, but provide a more firm methodologic basis for the final combination of data from seven North American studies to be undertaken following data collection in four additional sites.

2 CRITICAL REVIEW OF THE LITERATURE

2.1 BACKGROUND AND SCOPE OF THE REVIEW

In 1989, the International Workshop on Residential Radiation Epidemiology endorsed the undertaking of case-control studies to determine the risk of lung cancer associated with exposure to radon gas and progeny in homes (US DOE/Office of Energy Research 1989). This followed inconsistent results from the numerous ecologic studies (see review by Stidley & Samet 1993). Further ecologic studies were specifically discouraged in view of the inherent limitations of such studies for testing causal hypotheses (US DOE/Office of Energy Research 1989; see also Greenland & Robins 1994). Since that time, eight large-scale case-control studies with directly measured home radon exposures have been published, including almost 5,000 incident cases of lung cancer.

Lubin et al., (1990) catalogued a number of potentially limiting factors in the design of epidemiologic studies of indoor radon exposure; in a commentary following the publication of the first seven such studies, Lubin (1994) described their individual results as inconclusive. An analysis of the combined primary data from three of the female-restricted studies (in New Jersey, USA; Shenyang, China; and Stockholm, Sweden) was published by Lubin and his colleagues (1994a) shortly thereafter. Assuming a linear model of excess relative risk, these authors estimated an overall excess relative risk¹ [and 95% confidence interval] per pCi/l (equivalent to 37 Bq/m³) of 0.00 [-0.05, 0.07], not inconsistent with that extrapolated from

¹ ERR = RR - 1; RR denotes relative risk, which, under certain conditions, such as the study of a rare disease (as is lung cancer) or with incidence-density sampling, estimates the incidence density ratio (IDR).

the analysis of 11 miner cohorts (Lubin et al., 1994b)². Lubin and Boice subsequently conducted a meta-analysis of all eight residential studies in 1997³. The meta-analysis employed a log-linear model of exposure-response and random effects methods for combining the data from each site, to estimate an overall relative risk (RR) per 150 Bq/m³ of 1.14 (95% CI [1.0, 1.3]). The causal plausibility of these results was supported by a statistically significant trend (two-sided $p=0.03$) of increasing response with exposure (Samet 1997). Moreover, the results of the meta-analysis were consistent with those of Lubin et al.'s (1994) re-analysis of the data from 11 miner cohorts. Significant but unexplained heterogeneity was observed between the residential studies in the estimates of risk, however. Lubin and Boice could only acknowledge the potential for confounding and exposure misclassification, key issues which cannot be investigated in a meta-analysis that employs only summary level data. It was concluded that the meta-analysis provided support for the general validity of the studies of underground miners, but that the latter studies still provided the best source of data for the assessment of indoor radon risk, pending completion of a pooled analysis of the individual-level data from the residential studies. Further efforts to combine case-control studies would be required to quantify the carcinogenic potential of indoor exposure to radon.

This chapter undertakes a comparative review of the eight original case control studies according to current epidemiologic theory, to reconcile the results of these studies and to *qualitatively* estimate the impact of the major sources of uncertainty in the estimation of lung cancer risk resulting from residential exposure to radon. The principal design features of each study will first be described. This will be followed by a critical and synthetic epidemiologic analysis that will address in order: 1) potential for selection bias; 2) potential

² The results of Lubin et al.'s (1994a) re-analysis were both a validation and an extension of the 1988 analysis by the BEIR IV (Biological Effects of Ionizing Radiation) committee of the National Academy of Sciences, which combined data from four miner cohorts (NRC 1988).

³ Since the meta-analysis had not been published at the time of writing this review, it is not considered in-depth.

for information bias; 3) limitations in the analysis, including exposure-response modelling and appropriate treatment of covariates (confounders and effect modifiers), and, 4) coherence of the evidence in terms of biologic plausibility, main study limitations, and overall consistency. Detailed tables comparing the reviewed studies along the foregoing dimensions are included in Appendix A for ease of detachment and reference.

2.2 OVERVIEW OF THE EIGHT CASE-CONTROL STUDIES

2.2.1 SHENYANG CHINA: BLOT ET AL., 1990

An industrial city in the northeast part of the People's Republic of China, Shenyang has one of the world's highest rates of lung cancer in women. Blot et al.'s study investigated the reasons for this. The investigation included 308 registry-identified cases of lung cancer in women aged 30 - 69 years, diagnosed between September 1985 and September 1987, and 356 population controls of similar ages. As a proportion of those 'eligible', the analysis included 79% of cases and 91% of controls. One year radon measurements were attempted in only one home per subject, in which minimal duration of residence was 5 years. Covariates included active smoking (dichotomous), spousal smoking (dichotomous), air pollution (dichotomous), and age and education in unspecified categories. Indoor air pollution, described as "high for western standards" is an important confounder particular to this study; radon levels were significantly higher in homes with elevated levels of air pollution from heating and cooking sources.

2.2.2 NEW JERSEY USA: SCHOENBERG ET AL., 1990

This study was also restricted to women, and included a subset of the subjects selected between 1982 and 1984 to study smoking and other risk factors for lung cancer. Population

controls were selected from three sources depending on the vital status and age of the case: driver licence (<65 years) or health care files (65 years +) for controls of directly interviewed cases, and death certificates for controls of proxy-interviewed cases. Live cases were frequency-matched to live controls on age and race; deceased cases were frequency matched to deceased controls on age, race, and nearest date of death. The analysis was limited to those for whom both interview and home monitoring data was available, and included 31% (411) of the eligible cases and 27% (385) of the eligible controls; the possibility of selection bias was one of several author-cited limitations. Due to budget limitations, year-long radon measurements were limited to the most recent New Jersey residence of minimum 10 years duration in the period 10 to 30 years prior to ascertainment. Covariates included several active-smoking related variables, age (in three categories), education (three levels), 'race' (dichotomous), county of residence, occupation (as a dichotomous variable), vegetable consumption (three categories), and source of information (direct versus proxy respondent).

2.2.3 SOUTH FINLAND: RUOSTEENOJA 1991

Restricted to male subjects, this study was conducted in an area with one of the highest levels of indoor radon world-wide. Lung cancer cases (238) were identified from two sources according to year of diagnosis: 1980 - 1982 from the Finnish Cancer Registry, and 1983 - 1985 from hospital records. Controls (434) were selected from a population register and included all smokers and a random sub-sample of non and ex-smokers from an initial random sample of 1500. Two month winter measurements of radon were attempted in all dwellings occupied by subjects for a minimum of one year between 1950 and 1975 (25 years). The historical time period prior to recruitment for which radon exposure was estimated thus varied across subjects. As a proportion of those 'eligible', the analysis included 75% of cases, and 88% of controls. Covariates included smoking, age, birth cohort, education, and occupational exposures.

2.2.4 STOCKHOLM SWEDEN: PERSHAGEN ET AL., 1992

An extension of an earlier investigation that assessed radon exposure in only a small sample (10%) of subjects' residences, this study, restricted to females, sought year-long radon measures in the livingrooms and bedrooms of all dwellings within the county in which subjects had lived for a minimum of two years during the defined exposure period. This period varied somewhat across subjects, extending between 1945 and the five years prior to subject recruitment. Suspected cases of lung cancer were interviewed in hospitals between September 1983 and December 1985. Those subsequently diagnosed with lung cancer comprised the 210 cases; the remaining 191 individuals comprised the hospital controls. Population controls were selected randomly from a sub-sample of women sharing their date of birth with the cases, as identified from Stockholm County population registers. As a proportion of those 'recruited', the analysis included 96% of cases and 95% of controls. Covariates included age (3 levels), active smoking (4 levels), and municipality of residence, although information on passive smoking, diet, and occupation was also collected.

2.2.5 WINNIPEG CANADA: LÉTOURNEAU ET AL., 1994

Conducted in a city with one of the highest levels of indoor radon among 18 Canadian centres monitored (according to McGregor et al., 1980), this investigation involved both males and females. Histologically confirmed lung cancer cases (aged 35 to 80 years) diagnosed between 1983 and 1990 were identified from the Manitoba Cancer Registry and individually matched by age (± 5 years) and gender with potential controls identified by random selection from the Winnipeg telephone directory. Although the numbers of non-respondents to initial contact was unrecorded during recruitment of controls, 97% of those interviewed were retained in the analysis. Of the 1400 cases eligible, 53% were included; reasons for case exclusion were limited to: illness of the patient, refusal of the proxy, lack of permission by the physician, or inability to trace the patient. One year radon measurements

were sought in all Winnipeg metropolitan area residences in which subjects had reported living for at least one year during the exposure period 5 - 30 years prior to recruitment. One third of reported residences were actually assessed. Covariates included age (6 categories), active cigarette smoking (ever smoking, age at start, duration, daily and pack-years), occupational exposures (3 indicator variables each for job and substances), education, (2 categories) , and country of birth (5 categories).

2.2.6 MISSOURI USA: ALAVANJA ET AL., 1994

The most recent of the four studies in women, this investigation was also restricted to 'white' non-smokers. The 538 primary lung cancer cases were identified from those reported to the Missouri cancer registry between June 1 1986 and June 1 1991, having been diagnosed between 1986 and 1992 (sic); 1183 controls, frequency matched to controls (to a tolerance of 5 years), were identified during an unspecified time interval from one of two sources depending on case age: the Ministry of Revenue driver's licence files (for the 30 to 64 year age group) and Health Care Financing Administration medicare listings (for the 65 - 84 year age group). As a proportion of those 'eligible', 83% of cases and 78% of controls were included in the analysis. Year long radon measurements were attempted in at least one home that the subjects had occupied during the exposure period 5 - 30 years preceding recruitment. Covariates included age, marital status (6 categories), active smoking (previous smoking, years since cessation, and pack-years), passive smoking (5 levels), pre-existing nonmalignant lung disease (dichotomous), education (3 categories plus "unknown"), saturated fat intake (quintiles of kilocalorie intake), and source of information (direct versus proxy).

2.2.7 SWEDEN: PERSHAGEN ET AL., 1994

These investigators defined a study base which included all males and females 35 to 74 years of age who had resided in 1 of 109 municipalities in Sweden at some time between January

1, 1980 and December 31, 1984, and who had also lived in Sweden on January 1947. This established a target population with a high probability of providing long-term measurements of historical radon exposure. The second largest case-control study of residential radon exposure conducted to date, it included 1360 registry-identified cases which were selected to include all 650 women identified, and 850 men (a 40% random sample of eligible males). Two control groups (each of 1500 subjects) were defined and identified through the population registers of Statistics Sweden: in the first group, consisting of 775 females and 725 males, controls were frequency matched to cases on age (to within a 5 year tolerance) and calendar year of residence; the second group was selected based on the same criteria as the first except that it was additionally frequency-matched on vital status (to reduce the potential for information bias); this second control group also excluded subjects dying of 'smoking-related causes' (to avoid over-representation of smoking-related causes of death). As a proportion of those 'eligible', 82% of cases and 90% of controls were included in the analysis. Attempts were made to measure radon levels over a three month time period during the heating season in all dwellings in which the subjects had lived for a minimum of two years during the period between 1947 and three years prior to recruitment. Covariates included age (in 5 year intervals), several smoking-related variables (current status, time-weighted mean consumption, spousal, parental), occupational risk (4 categories), urbanization (dichotomous), and whether the subject habitually slept with an open window (dichotomous).

2.2.8 FINLAND: AUVINEN ET AL., 1996

Nested in a cohort of subjects identified from the Finnish Population Registry and who had resided in the same family home for a minimum of 19 years (from January 1, 1967 or earlier, until the end of 1985), Auvinen et al.'s study was the largest and most recently published case-control study involving direct measurement of radon in subjects' homes. A total of 1973 lung cancer cases (excluding pleural cancer) diagnosed between January 1, 1986 and

March 31, 1992 were identified from the Finnish Cancer Registry, and matched by year of birth and gender to 2885 controls. Both home radon measurements and questionnaire data on covariates were available on 1055 cases and 1544 controls, representing 53% and 54% respectively, of the eligible subjects. Participants were not different from non-participants with respect to gender, age, vital status, or urban/rural residential location. Radon measures were available on average for a greater period of time for cases than for controls; cases reported a median of 38 years lived in the index residence, compared to 35 for controls. Twenty percent of both case and control subjects were excluded due to 'invalid' radon measures. Covariates included active smoking habits (status, intensity, duration, age at start, and dichotomous quit date), intensity and duration of exposure to environmental tobacco smoke, and occupational hazard exposure (dichotomous).

2.3 POTENTIAL FOR SELECTION BIAS

In a case-control study selection bias can arise from : 1) the use of non-comparable criteria to enroll cases and controls (Hennekens & Buring 1987), such criteria typically relating to opportunity for detection of disease⁴ and residence in the study catchment area (Wacholder et al., 1992a); 2) the use of prevalent rather than incident cases, creating the potential for bias due to *selective survival*, if such survival is related to degree of exposure (Kramer 1988); and 3) differential participation which results in final case and control groups that differ from their eligible populations in uncontrolled (and important) risk factors.

Since radon exposure status was not known at the time of subject recruitment, there was relatively minor potential for selection bias among the studies, in spite of some minor inconsistencies with respect to case and control inclusion criteria⁵. All studies recruited

⁴ Mis-classification of disease status is discussed under information bias in section 2.4.2

⁵ See Appendix A Table 1 for comparative details regarding subject selection.

incident cases and all but two studies (Blot et al., 1990, Pershagen et al., 1992) ensured representation of fatal cases through the use of proxies for cases which had deceased by the time of contact⁶. In the latter two studies, the use of rapid reporting of systems to capture incident cases prevented potential bias from selective survival.

All but two (Auvinen et al., 1996; Schoenberg et al., 1990) of seven studies for which there was complete response data included at least 75% of eligible cases and controls. In the New Jersey study, 25% of eligible cases and 35% of eligible controls were excluded because of refusal or loss to follow-up. In Finland (Auvinen et al., 1996), proportions of cases and controls differing by less than 1% were excluded at interview and home measurement, to yield an overall response of 53% and 54% of those cases and controls identified, respectively. Among the seven studies for which numbers of eligible subjects were obtained, the largest difference in overall response proportion between case and control series was 8% in South Finland (Ruosteenoja 1991). Numbers of non-responders were unrecorded during random dialling to recruit Winnipeg controls (Létourneau et al., 1995a). Of those cases eligible, 53% were retained in the analysis. From interview, however, 97% of the case and control series were retained. In spite of the aforementioned discrepancies in response rates between cases and controls, it is difficult to imagine plausible differential selection forces which could have created an uncontrollable bias in the reviewed studies.

2.4 POTENTIAL FOR INFORMATION BIAS

2.4.1 MIS-CLASSIFICATION OF DISEASE STATUS

The consequence of disease mis-classification in a case-control study is the attenuation of estimated odds ratio (OR) if there is a true effect of exposure. Such mis-classification can

⁶ See Appendix A Table 2 for comparative details regarding the use of proxy data.

arise from over-ascertainment of cases and/or under-ascertainment among controls. The latter is unlikely since lung cancer is usually rapidly fatal. With respect to the former, if cases do not in fact have primary (but metastatic) lung cancer, their inclusion dilutes the case group with false positives. According to a Swedish study (Larsson 1971), using National Cancer Registry facilities as the primary data source for lung cancer incidence data leads to over-registration in approximately 2% of the cases. Over-ascertainment was minimized through review of the pathology/histology in the majority of cases among all but one of the eight investigations⁷: Ruosteenoja et al., (1991), relied on hospital records alone. A small bias to the null may have occurred in this study, and to a lesser extent, in Missouri (Alavanja et al., 1994), where slide review was completed in only 75% of cases.

In the second instance, if controls have undiagnosed (primary) lung cancer, exposure would be underestimated in the control group. Although not an important issue among live controls, the use of deceased controls in four studies⁸ ultimately entails reliance on death certificates for ascertainment of cause of death. According to Percy et al. (1981), US death certificates detected 95% of lung cancers listed in hospital records during the 1970's. Such data suggest that disease misclassification among deceased controls would be an important source of attenuating bias only if the relative effect of radon was quite large.

2.4.2 EXPOSURE MEASUREMENT ERROR

2.4.2.1 Consequences Of Errors In Exposure Ascertainment

The consequences of error in exposure ascertainment depend on whether the classification on one axis (e.g. exposure), is independent of (non-differential) or dependent on (differential)

⁷ See Appendix A Table 1.

⁸ See Appendix A Table 2.

the classification on the other (e.g. cases vs. controls) (Dosemici et al., 1990; Kramer 1988; Rothman 1986). *Non-differential* errors can be random, with an average error of zero, or systematic, with exposures being consistently over or under-estimated on average in both case and control series. Purely random error will attenuate estimates of the exposure-disease association (Brenner & Loomis 1994) and will spuriously narrow the estimated confidence interval (Armstrong 1990). Systematic non-differential overestimation of exposure will result in an attenuation of the slope of an exposure-response relationship, whereas underestimation of exposure will lead to overestimation⁹ of the slope. The net effect of *differential* mis-classification of exposure depends on the pattern of bias within exposure-outcome combinations, following no general rule (Kramer 1988). *Both* differential and non-differential mis-classification of *either* the principal exposure under study, or a covariate, will tend to distort the degree of heterogeneity in the OR (Greenland 1980). Exposure measurement errors present in the residential radon data could conceivably be random, systematic, or differential.

2.4.2.2 Overview

Figure 2.1 provides a simplified diagram of error-free radon exposure data; figure 2.2, illustrates deviations from this ideal. Ideally, exposure to radon progeny would be assessed both indoors and outdoors, and continuously over the course of the etiologic time period, through the use of a mobile personal monitor in a prospective study. Personal monitors have only recently been pilot-tested (Litt et al., 1991; Harley et al., 1991) and provide integrative (rendering an average concentration over time) rather than continuous exposures and measure radon gas rather than the progeny of direct interest. Integrative alpha track devices (ATD's) provide area measurements are subject to imprecision and bias, as well as to spatial uncertainty. Exposures outside the home are ignored. Moreover, retrospective estimation

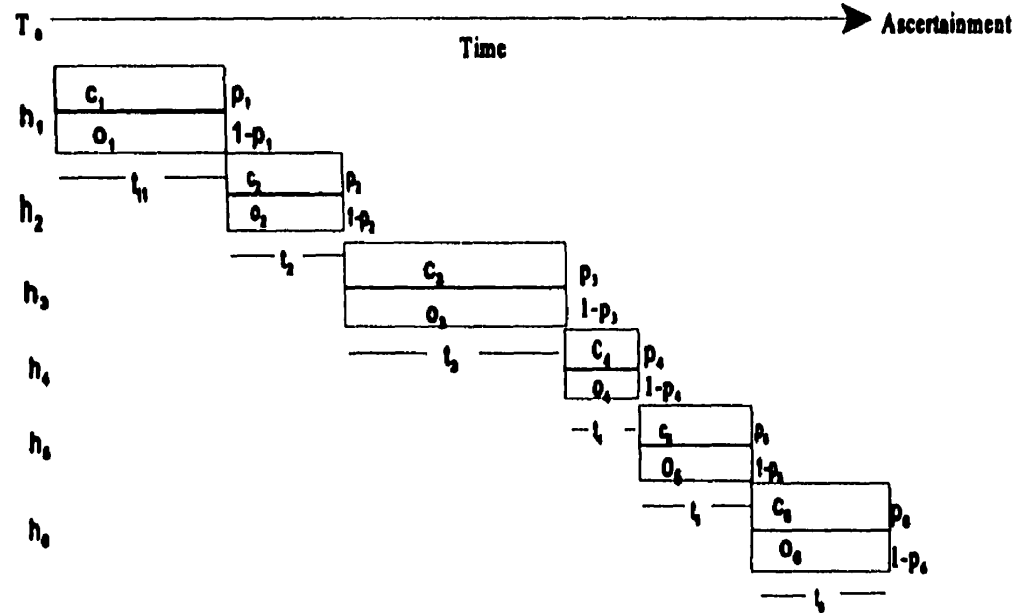
⁹ It should be noted that categorization of exposure can lead to more complex forms of error (Dosemici et al., 1990; Brenner & Loomis 1994).

Figure 2.1

Simplified Diagram of Error-Free Radon Exposure Data

Complete Home History

for subject i where $j = 6$ homes within the appropriate exposure time-window



Legend

- h_j denotes home of subject where $j = 1, \dots, q$
- t_j denotes the duration of residence by subject in home j , in units of Working Months where 1 WM = 170 hours
- c_j denotes the average activity concentration of radon decay products in units of Working Levels, to which subject is exposed in home j
- o_j denotes the average activity concentration of radon decay products in units of Working Levels, to which subject is exposed outside of home j
- p_j denotes the proportion of time subject spends inside home j

Assuming continuous and contemporaneous personal monitoring of Radon Decay Product (RDP) concentration, the cumulative exposure for subject i in WLM could be calculated as:

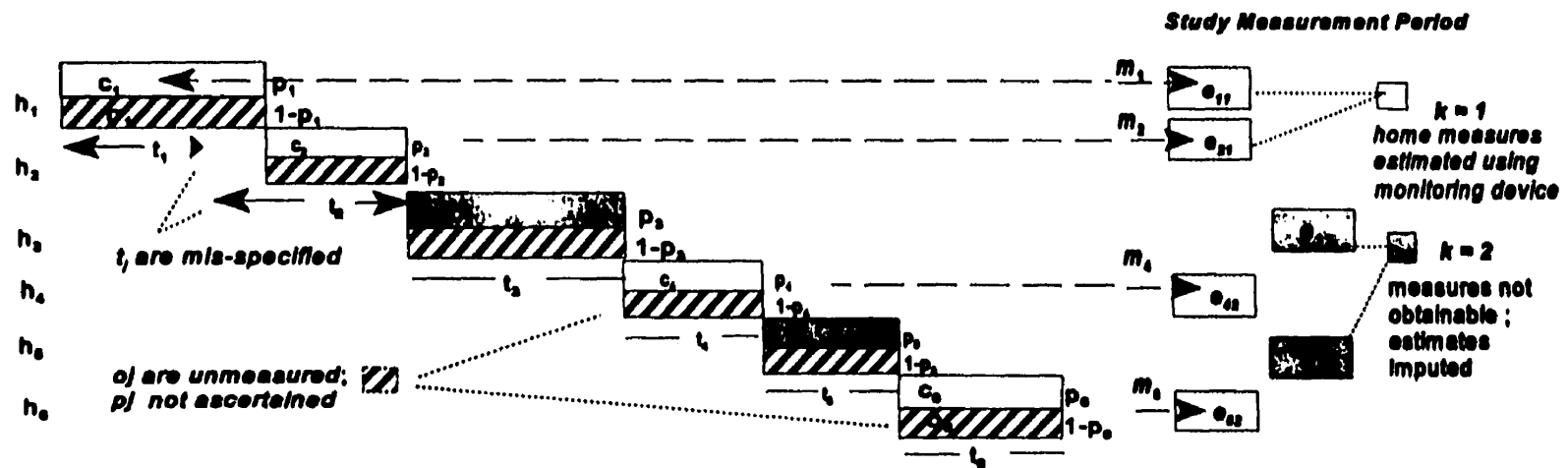
$$WLM_i = \sum_{j=1}^q t_j [c_j p_j + o_j (1-p_j)]$$

Figure 2.2

Errors in Radon Exposure Estimation

Available Home History

for subject i where $j = 6$ homes within an *uncertain* exposure time-window



Legend

- h_j denotes home j of subject where $j = 1, \dots, 6$
- t_j denotes the duration of residence of subject i in home j , in units of Working Months where 1 WM = 170 hours
- c_j denotes the average activity concentration of radon decay products (RDP) in units of Working Levels, to which subject i is exposed in home j
- o_j denotes the average activity concentration of RDP in units of Working Levels, to which subject i is exposed outside of home j
- p_j denotes the proportion of time subject i spends outside of home j
- c_{p_j} denotes the average activity concentration of radon gas (in Bq/m³), as estimated by method k , to which subject i is presumed to have been exposed in home j
- m_j denotes the interval of time between the mid-point of residence by subject i in home j , and mid-point of monitoring of radon concentration

of past indoor exposure entails the estimation of a uniformly valid and etiologically relevant exposure time window (ETW), the identification of a complete home history during the ETW, and access to and monitoring of each home. Since home radon levels are subject to temporal variation due to changes in weather and home ventilation characteristics, even complete current area radon gas estimates may differ from actual past exposures. Inaccurate recall of housing history, and barriers to the measurement of radon in past homes (e.g., refusal of current residents, location outside of study bounds, or destruction), result in missing data. The analysis of exposure histories with gaps entails assumptions about unavailable measures and either implicit or explicit imputation of the missing data. Imputation can itself introduce bias to the estimated cumulative exposure and distort standard errors (Weinberg et al., 1996).

Table 2.1 provides a summary of the sources of error in radon progeny exposure assessment and identifies the section in which each type of error will be discussed¹⁰.

¹⁰

Please refer to Appendix A Table 3, for comparative details discussed throughout the remainder of this section, regarding radon exposure information ascertainment in the reviewed studies.

Table 2.1
Sources of Error in the Estimation of
Cumulative Indoor Radon-Progeny Exposure

Source of Exposure Error	Section¹
Model Uncertainty:	
Mis-specification of the etiologically relevant exposure time window	2.4.2.1
Errors related to measurement of radon indoors:	
Measurement error in the α -particle track device	2.4.2.2
Spatial uncertainties:	
Measurement at one fixed location in a room	2.4.2.5
Measurements limited to one or two rooms only	2.4.2.5
Temporal uncertainties:	
Mis-specification in the duration of monitoring	2.4.2.6
Mis-specification in the duration of residence	2.4.2.6
Use of contemporary measurements to characterize past levels, including:	
Structural changes to homes	2.4.2.6
Meteorological variability	2.4.2.6
Differences in the ventilation habits between the subject and current resident	2.4.2.6
Sources of error in total exposure:	
Model uncertainties:	
Method of dealing with measurement gaps for homes within the exposure period	2.4.2.9
Exclusion of exposures outside the subject's home	2.4.2.4
Inter and intra-individual variation in percentage occupancy over time	2.4.2.4
Uncertainty in percentage occupancy	2.4.2.4
Relationship between exposure and dose:	
Physical factors:	
Activity-weighted particle size distribution	2.4.2.10.1
Equilibrium and conversion of radon concentration to WLM	2.4.2.10.2
Electro-magnetic frequency radiation	2.4.2.10.3
Biologic Factors	2.4.2.10.4

¹ See numbered section for discussion in text

2.4.2.3 Etiologically Relevant Exposure Period

Inappropriate assumptions, implicit or explicit, about the etiologically relevant exposure period results in non-differential mis-classification and attenuation of ORs. Uncertainties involved in the determination of the appropriate ETW include: 1) the suggestion that the

latent interval is not deterministic, but follows some probability distribution (Thomas 1983); 2) the suggestion that the length of the latent period may increase with decreasing exposure rate, therefore being longer in typical residential settings than in mines (Steinhäusler 1988); 3) the suggestion that smoking may reduce the latent period (Steinhäusler 1988, citing Axelson & Sundell 1978; Axelson 1980); 4) the limitation of miner studies to the exposure period corresponding to adult employment; 5) incomplete follow-up of subjects from exposure to death in cohort studies of underground miners (Samet 1989; Lubin et al., 1994a); and, 6) in the most recent analyses of data from 11 miner cohorts (Lubin et al., 1994; NRC 1998) the time interval 25 years or more prior to index date was not further sub-divided for analysis.

In terms of minimum latency, Lubin et al.'s (1994a) estimates among 11 miner cohorts ranged from 0-10 years. Moolgavkar et al., (1993), in a re-analysis of the Colorado miner cohort data, assumed a minimum latency of 3.5 based on the smoking study of Doll & Peto (1978).

Although it is unknown to what extent the availability of data drove assumptions regarding the ETW, these assumptions varied across the residential studies. The maximum induction period varied from 25 to 30 years, and minimum latency included zero, by Blot et al., (1990), three years, by Pershagen et al., (1994) and five years by the remainder of the investigators. The total effective duration of the measurement period, on average per subject, varied from 15 to 26 years. In three studies, moreover, the ascertained exposure period varied across subjects within the study (Pershagen et al., 1992 & 1994; Ruosteenoja 1991).

Pershagen et al., (1992) attempted to weight exposures according to the results of the BEIR IV (NRC 1988) model, resulting in slightly less precision in the effect estimate. Létourneau et al., (1994) provided separate results for two ETW's, 5 - 15 and 5 - 30 years prior to index date; the latter wider interval generally demonstrated higher ORs with more precise

confidence intervals. An analysis of the combined data will provide an opportunity to test varying assumptions about the ETW.

2.4.2.4 Accuracy of Measurement Devices

Aside from the use of charcoal canister measures¹¹ for a small portion of measures in two studies (Schoenberg et al, 1990; Pershagen et al., 1992), all eight studies employed integrative ATD's supplied and processed by one of two main sources: Terradex Corporation, or respective government laboratories. The principal source of ATD inaccuracy arises from counting error. Because of the Poisson distribution of radiation processes, precision of measurement is proportional to counts, the product of monitoring durations (time) and concentrations. Other sources of ATD error result from calibration error [estimated at an additional 8% (coefficient of variation) by Bierma et al., 1989], film sensitivity, and background (Yeager et al., 1991).

Based on the field measurements of Terradex detectors over a measurement period of approximately 3 months at a range of exposure concentrations similar to the equivalent annual average concentration in US homes, reliability, as estimated by the coefficient of variation (CV), ranges from an average of 8% (at 50kBq-day/m³) to 25% at (5 kBq-day/m³) (Scott & Robertson 1991). Concentration-specific precision would have been least in the two studies for which monitors were in place for less than a full year; the South Finnish (Ruosteenoja 1991) and Swedish (Pershagen et al., 1994) studies monitored indoor radon for 2 and 3 months, respectively. In both of these studies, the shorter duration would have been at least partially offset by relatively high average radon concentrations¹². As a result of their relatively higher radon levels, ATD precision would have been greatest in Winnipeg and

¹¹ For equal measurement times, no significant differences have been observed in measurement error between charcoal canisters and ATD's (US EPA 1992a).

¹² See Appendix A Table 4 for a comparison of radon levels in the reviewed study locations.

Stockholm. The impact of unreliability in ATD measures is attenuation of the estimates of ORs and narrowing of their associated confidence bands.

2.4.2.5 Residential Mobility

Although not an error *per se*, studies conducted in areas of high residential mobility will have more homogeneous average and cumulative exposures, reduced inter-subject variability in exposures, and hence reduced power to detect an effect of high cumulative radon levels (Upfal 1995, Lubin et al., 1995a). A study of subjects who lived in an average of 3 homes in their adult lives would require up to three times the number of subjects as one in which subjects lived in a single home, depending on the degree of exposure error (Lubin et al., 1995a). In the only study which provided such data (Létourneau et al., 1994), subjects lived in 9 different homes during the ETW. This high mobility in Winnipeg could offset the gains from having higher radon levels.

Mobility can indirectly result in exposure error by increasing the probability of incomplete missing data by a number of means: 1) housing history may not be accurately remembered; 2) previous homes may be outside the study boundaries. As a case in point, on average only 5 out of 9 ETW residences were within the study bounds of Létourneau et al., (1994); 3) the previous home may no longer exist (having been demolished since subject relocation) or the current resident may be untraceable or refuse to allow placement of monitors; and 5) measurement of previous residences may not have been included in the study protocol, (as in Blot et al., 1990 & Schoenberg et al., 1991). Some investigators attempted to minimize problems arising from mobility by conducting their study in an area with high stability, or restricting their study samples to individuals who had resided for some minimal time within the study boundaries (Pershagen et al., 1994), or in the measured home (Auvinen et al., 1997, minimum 19 years; Blot, et al., 1990, minimum 5 years; Schoenberg et al., 1991, minimum 10 years).

2.4.2.6 Home Occupancy and Outside Exposure

Potential exposure received outside the home was excluded from consideration in all of the residential studies. This involves an implicit assumption that such exposures were either negligible or proportional to in-home exposures. The former would have been more valid in the female-restricted than the mixed-gender studies, while the latter seems doubtful. In their calculation of cumulative exposure in working level months (WLM)¹³, Schoenberg et al., (1991) used a fixed occupancy assumption of 80%. Pershagen et al., (1992) and Alavanja et al., (1994) based their occupancy factors on information gathered at interview; the latter employed a fixed occupancy factor (84%) between and within subjects, while the former did not specify how 'adjustment' for occupancy was accomplished. Based on a Canadian survey, Leech et al., (1996) estimated an overall average of 64% of time spent in the home. The occupancy fraction remains subject to inter- and intra-individual variation. Potential sources of such variation include climate, gender, age, employment (Harley 1991), and lifestyle changes. Not surprisingly, then, the p-values associated with Pershagen's (1992) OR estimates became less significant after 'adjustment' for residential occupancy. If cases have spent more time in the home, the assumption of a constant value would create differential exposure error that would result in a bias toward the null. Auvinen et al., (1996) observed 'no clear differences' in home occupancy, however (median values for cases and controls respectively were 11.5 and 11.7 hours per day in the home).

13

Working Level Month (WLM) is the unit of exposure to radon progeny employed in the mining studies. One WLM is equivalent to exposure to 1 working level (WL) for 170 hours (a typical miner's working month). One WL is any combination of radon progeny in 1L of air that ultimately releases 1.3×10^5 MeV of alpha energy during decay. The SI unit for the rate of decay is Bq/m³; assuming 50% equilibrium between radon gas and its progeny, 1 WL is approximately equal to 37 Bq/m³ (IARC 1988). Equilibrium fraction is discussed in section 2.4.2.4.1

2.4.2.7 Spatial Uncertainty

Spatial uncertainty in radon exposure is roughly related to the number of floors in the home and the distribution of subject time over these floors, each of which is subject to inter- and intra- individual variability. None of the reviewed studies assessed such factors. All but two (Schoenberg et al., 1990; Ruosteenoja 1991) of the eight studies measured radon concentrations in more than one area of the home, thus improving spatial representativity of area measures¹⁴. In three studies (Alavanja et al., 1994; Pershagen et al., 1994, and Auvinen et al., 1996) it is not stated how the two area estimates were used, although one of these indicates that differences were slight. Pershagen et al., (1992) averaged bedroom and living room concentrations, even though a much greater proportion of the 60 to 80% of time an individual spends in the home would be expected to have been spent in the bedroom. Blot et al., (1990) employed the maximum of bedroom and living room estimates for each subject, after analyses using the average of the two or the maximum of the two yielded "similar results". However, 22% of values for the two disagreed by a factor greater than two; unusually high levels of room aerosols in China might explain the differences. Létourneau et al., (1994) were the only investigators who monitored basement levels for all subjects; OR estimates were reported separately. Given the fact that living area concentrations are on average only two thirds that of basement levels, it is important to know how to weight the proportions of time spent at different levels in a home. Although Leech et al., (1996) described the distribution of time spent by Canadians in their basements at varying times of the day, the historical representativity of such data is questionable. If cases were more likely to spend time in their basements, differential error could result in a bias toward the null. The impact of random error related to spatial uncertainty would be a further under-estimation of magnitude of both central estimate of the ORs, and the width of their corresponding confidence bands.

¹⁴ See Table 3 of Appendix A.

2.4.2.8 Temporal Uncertainty

Temporal uncertainty exists in the duration of radon monitoring, the duration of residence in an identified home, and in the historical representativity of current radon measurements. Error in monitoring duration arising from errors in the time of placement and retrieval of dosimeters and in home residency durations are likely to have been random and minimal. A more substantial form of temporal uncertainty relates to the use of current measures of building radon levels to represent past exposure levels.

Radon levels in buildings depend on rates of air entry and exit, circulation, and transformation (attachment, deposition, decay) (Nero 1988a). These may vary diurnally (WHO 1985), seasonally (Martz et al., 1991, Nero 1988a, WHO 1985), and from year to year (Létourneau et al., 1992; Martz et al., 1991). Measurement of annual average concentrations in one or a few homes today which a subject occupied 5 - 35 years ago may not, therefore, accurately reflect those previous exposure conditions. Three main problems affect validity even with complete coverage and precise measures of current radon levels: 1) structural changes to the building (e.g. from aging, remodelling, new furnaces, storm windows), on ventilation rates (and the equilibrium factor); 2) meteorological variation, both seasonally and annually; and 3) changes in the number, age, and behaviour of occupants, exemplified by differences between the window opening habits of the subject (former resident) and current resident if it is a previous home (Archer 1995).

Structural changes to homes is an uncertainty faced to an unquantified degree by all investigators. Since the trend has been to improve energy conservation by tightening homes, current estimates may systematically overestimate historical dose, underestimating ORs. The potential for error would be directly related to the time period between measurement and residency.

The average variability in annual average radon concentrations was estimated by Martz et al., (1991) to be 25.2%, (CV ranging from 7.7 to 51.5%) based on measurements made in 25 Colorado sampling stations monitored over six years. Factoring out the contributions from the measured variability in the response of the detectors reduced this estimate to 22%. Thus, estimating previous years' levels, even without home changes, is subject to considerable random error.

Since radon levels are higher in winter due to increased pressure differences which increase rates of entry of radon (Nero 1988a), the exclusively winter measurements by Ruosteenoja (1991) and Pershagen et al., (1994), (for three, and two months, respectively) would have resulted in a seasonal bias, overestimating annual concentrations, and thereby underestimating the ORs. According to literature cited by Pershagen et al., (1994), the mean radon level is approximately only 10% higher during the heating season. In contrast, Alavanja et al., (1994) observed winter readings (estimated over a three month period) which were more than twice as high as year long readings.

The issue of whether or not subjects or current residents habitually slept with an open window involves not only exposure mis-classification but also potential effect modification. The latter issues will be dealt with in section 2.5.2.7.

Sleeping with an open window (SWOW) increases ventilation (up to 3 times), concomitantly reducing both radon levels *per se* (by 50 - 70% according to the preliminary results of Archer 1995) and the value of the equilibrium factor between radon gas and progeny. SWOW therefore affects both measured radon exposure and the (assumed constant) relationship between exposure and dose. In homes no longer occupied by study subjects and in which the window-opening behaviour of the current occupant differed from that of the study subject, the exposure measured under the ventilation habits of the current occupant would differ systematically from those experienced by the study subject.

Archer (1995) suggested that, "since SWOW is usually a lifetime habit, the radon levels imputed for unmeasured residences, and measurement made when someone else (usually not a subject reporting SWOW) is living there, will need to be adjusted downward" (when not SWOW by an estimated 28%). Archer further suggested that if study subjects who slept with an open window are residents of the previous home of an other subject (as in fact occurred in Winnipeg) who did not sleep with an open window, adjustments would have to be made upward (by an estimated 38%). This would require estimation of the proportion of subject measures requiring adjustment based on the probability of SWOW, and this would be both climate and season dependent. In interpreting the study by Pershagen et al., (1994), however, seasonality is not a problem as all measures were taken during the heating season.

Pershagen et al.'s 1994 study was the only one of eight to ascertain SWOW. Exclusion of subjects who reported SWOW increased the estimate of ERR from 3.4% per 1000 Bq/m³ radon exposure per year to 6.7% per 1000 Bq/m³ per year. Since these investigators did not, however, determine whether the current occupant slept with windows open, their study, although providing information key to understanding the implications of SWOW, was not free from exposure error in this regard.

In Pershagen et al.'s (1994) study, 31% of cases and 37% of controls slept with an open window; this observation not only raises the possibility that inattention to SWOW could result in differential exposure error, but in and of itself lends support to a radon effect. If cases were less likely than controls to have slept with an open window, and if their past homes, as is most likely, was occupied by an occupant who did not, a greater proportion of cases than controls would have been assigned home radon measures that were underestimates of the radon gas to which they were actually exposed. Moreover, the equilibrium between progeny and gas levels would have been higher and hence the measured gas levels would

underestimate the alpha dose actually received¹⁵. Hence, a potentially substantial differential bias toward the null could exist. Even if non-differential, however, inattention to SWOW reduces the signal to noise ratio in estimating an effect of radon on lung cancer.

2.4.2.9 Measurement Gaps & Imputation of Missing Values

Complete coverage of the ETW with home radon estimates is important not only to accurately represent subject exposures but to adequately reflect variation in relevant exposures and hence in the estimated OR. The extent of missing data in the reviewed studies was considerable¹⁶. Six of the eight studies defined an explicit ETW. Coverage of this ETW by dosimeter estimates of radon ranged from a low of 50% by Ruosteenoja (1991) to a high of almost 80% by Alavanja et al., (1994). Five of the six studies defining an ETW attempted radon monitoring in all ETW homes, but only one of the latter restricted subjects to long-term home-residents.

As a result of residency criteria, the three studies which employed radon concentration data from only a single measured home (CONC) compared favourably with the others in terms of number of years of radon monitoring per subject. Median residence time was 20 and 21 years among cases and controls, respectively, in the original New Jersey study (Schoenberg et al., 1990), 24 years for all subjects in the Shenyang study (Blot et al., 1990), and 38 and 36 years among cases and controls, respectively, in the most recent (Finnish) study (Auvinen et al., 1996).

¹⁵ The equilibrium fraction and its relevance to the relationship between exposure and dose is discussed in section 2.4.2.10; potential 'confounding' and effect modification by SWOW are discussed in section 2.5.2.7.

¹⁶ See Table 3 of Appendix A for details.

Among the studies which attempted measurement in all ETW homes, two types of analyses were presented: one which related lung cancer risk to time-weighted average concentration (TWAC) of radon, with each *measured* home concentration being weighted by the time spent in the respective home; and a second which related lung cancer risk to cumulative exposure (CUM), summing dosimeter-estimated home and *imputed* measures over the entire ETW. In three of these studies (Pershagen et al., 1992 and 1994, and Alavanja et al., 1994) both TWAC and CUM analyses were presented; the TWAC analyses assumed that within subjects, the measured homes were representative, on average, of unmeasured homes. Simulation analyses by Weinberg et al., (1996) have demonstrated that this assumption results in a marked bias toward the null in point estimates, and 'badly sub-nominal confidence interval coverage'. In a review of six methods for handling gaps in exposure history data, (assuming a linear ERR model, random missingness, independence among true exposure levels between homes, and no other sources of measurement error), only one method, in which the observed control mean is imputed (OCMI), performed without bias and with coverage of nominal 95% confidence intervals that was close to 95%.

Ruosteenoja (1991) presented only a TWAC analysis, but imputed missing values based on (poorly correlated¹⁷) housing characteristics; the use in this latter study of estimated measures whose standard deviations were much smaller (63 Bq/m³) than those of the measured concentrations (531 Bq/m³) also artificially narrowed the range of exposures and reduced power. Subjects for whom no housing characteristics were available had missing values imputed as the national average home radon concentration for unmeasured homes at ground level, and a fixed level for above-ground residences. To the extent that exposures in South Finland approximated national levels, the latter should have approximated those among controls. Ruosteenoja et al., (1990) and Létourneau et al., (1994) also provided results for

¹⁷ Based on analysis of the entire sample, the pearson's correlation coefficient was only 0.23 in the log scale.

sub-samples of subjects defined on the basis of completeness of radon estimation. Such an approach results in a substantial loss of power, however (Weinberg et al., 1996).

Five studies reported a CUM analysis, each using a slightly different method to impute missing home radon measures¹⁸. Schoenberg et al., supplemented their CONC analysis with one in which cumulative exposure was estimated by imputing the median radon concentration from controls for periods within the targeted ETW not covered by the single measured residence. This method, although not equivalent to, should have closely approximated OCMI in validity. Two studies imputed an estimate of the central radon value (Létourneau et al., using the mean, and Pershagen et al., 1994 the median) across all subjects regardless of outcome. Observed (overall) mean imputation performs almost as well as OCMI, since the mean over all residences is very close to the mean over control residences (Weinberg et al., 1996). Pershagen et al., 1992 imputed missing values based on house characteristics, or assumed the concentration of the previous home, introducing random error. Alavanja et al., 1994 assigned to missing exposure periods the respective mean values for cases or controls. The latter method artificially inflates the difference between the cases and controls if there is a true difference.

The consequences of the various assumptions implicitly or explicitly employed in the reviewed studies to calculate radon exposure thus ranged from reduced power at the least (Létourneau et al., 1994; Pershagen et al., 1992 and 1994; Ruosteenoja 1991), to underestimation (Auvinen et al., 1996; Blot et al., 1990; Schoenberg et al., 1990) or overestimation (Alavanja et al., 1994) at the worst, and complicates inter-comparison of study results. The analysis of the combined data will provide an opportunity to apply a consistent and unbiased imputation procedure.

¹⁸ See Table 3 of Appendix A for details.

2.4.2.10 Relationship Between Exposure and Dose

ATD radon monitors contain a filter which keeps out dust and radon decay products (RDPs) (Yeager et al., 1991). The detector thus measures the alpha particles released from the first decay of the gas (Radon ²²²) to Polonium²¹⁸, and the further decay of radon progeny retained in the dosimeter. The relationship between exposure to radon gas (as measured) and accumulated dose of alpha radiation-induced damage in the cellular DNA of target tissues in the respiratory tract is extremely complex (Samet 1989). Only a simplistic analysis of this relationship will be attempted here. Both biologic (physiologic) and non-biologic (physical) factors mediate the exposure-dose relationship (James 1988; NCRP 1984) as summarized in Table 2.2 (based largely on Samet 1989 and NRC 1991).

Table 2.2

Physical and Biologic Factors Influencing Exposure-Dose Relationships

Physical Factors	Biologic Factors
Activity-weighted particle size distribution	Ventilation pattern
Equilibrium of radon with its progeny	Bronchial morphometry
Electro-magnetic frequency radiation	Mucociliary clearance rate
	Thickness of mucus
	Location of target cells

2.4.2.10.1 Activity-Weighted Particle Size Distribution

The process of radioactive decay (from radon gas) results in a high proportion of RDPs carrying an ionic charge; this charge (Busigin et al., 1981) facilitates RDP attachment to aerosols¹⁹. Even as neutral particles, however, RDPs can adsorb to aerosols. A variable

¹⁹ An aerosol is a system of solid or liquid particles which are dispersed in a gaseous medium, relatively stable, and have a high surface area to volume ratio (Yeh et al., 1976).

proportion of RDPs remain unattached, and appear to form aggregates with water molecules and other chemical species (Duport et al., 1977; Robkin 1987; Samet 1989). The unattached RDP fraction is twice as likely as the attached to adhere to lung tissue (NRC 1991), resulting in a greater deposition efficiency and absorption; the dose of radiation to lung tissue therefore decreases with the proportion of RDP's attached. The probability that RDP's will attach to aerosol strongly depends on the concentration of dust in air (NCRP 1984). In very clean air, RDPs exist mainly in the unattached form; the attached fraction is usually less than 0.1 in indoor environments, due to normal human activity. The smoke from one cigarette will profoundly increase aerosol concentration.

The size of particles determines their penetration through the upper respiratory system and the pattern of deposition within the tracheo-bronchial region. The pattern of deposition in turn affects dose (James 1988; NRC 1991). The geometrical diameters of aerosol particles normally range between 0.001 and 100 microns (Yeh et al., 1976). The specific mixture of RDPs also influences the radiation dose to target cells, although to a lesser degree.

According to NRC (1991 page 13), "recent evidence suggests that the 'unattached' progeny are in ultrafine clusters in the 0.5 to 3 nm size range, and the attached and unattached fractions can not be distinctly separated'.

2.4.2.10.2 *Equilibrium Of Radon Gas With Its Progeny*

To convert the concentration of radon gas as measured by the ATD, to the concentration of radon progeny, one must assume an equilibrium factor (F), defined as the ratio of the potential alpha energy concentration of the decay product mixture to the corresponding concentration if they were in radioactive equilibrium with radon (IARC 1988). In confined spaces (such as mines or homes), the activity concentration of the decay products rarely reaches secular equilibrium with radon gas, as a result of various removal mechanisms, such

as ventilation (which reduces F) (Nero 1988), and deposition (or plateout) of the RDPs onto surfaces (IARC 1988, WHO 1985). Home air exchange rates vary (Murray & Burmaster 1995) among homes, among subjects (by virtue of their ventilation habits), and among studies (due to differences in average temperature). Deposition is affected by attachment²⁰; attachment, in turn, is affected by aerosol concentrations. There is an inverse relationship between the unattached fraction and the equilibrium fraction (James 1988). Lowering the particle concentration increases the deposition rate and lowers the equilibrium factor (F) (Nero 1988a). Estimated values for F range from 0.3 to 0.6 (IARC 1988) based on measurements of indoor air.

The equilibrium factor, together with exposure rate (concentration), duration, and an occupancy assumption enables one to convert radon concentration to the unit of cumulative exposure employed in the mining studies, that of the working level month (WLM). Three sets of study authors analysed cumulative radon exposure in terms of WLM (Schoenberg et al., 1990; Pershagen et al., 1992; Alavanja et al., 1994). Following the EPA (Puskin 1992), all used fixed equilibrium factors of 0.5. Assuming a constant equilibrium factor (even by neglecting an adjustment) for all persons regardless of home dust or smoke levels adds an element of uncertainty to the exposure analysis.

If the actual F is less than the fixed value (0.5) typically assumed, the proportion of RDPs to measured gas is lower than assumed, resulting in an overestimate of dose and underestimate of effect per unit dose. Conversely, an increase in aerosol concentrations, as by the introduction of dust or cigarette smoke (Nero 1988a; Porstendorfer 1987; Jonassen & Jenson 1989) increases the equilibrium factor, shifting the balance of radon (measured) to RDPs (toxic), and RDPs from their unattached (less efficiently absorbed) to their attached (more readily absorbed) forms. If the F value is higher than typically assumed, as for

²⁰ Particle deposition is relatively low, on the order of 0.1/hour compared to deposition of unattached RDPs, which is on the order of 15/hr (James 1988)

example among those exposed actively or passively to tobacco smoke RDP exposure would be underestimated and the corresponding effect measure overestimated.

2.4.2.10.3 *Electromagnetic Frequency Radiation*

A physical factor only recently linked to radon exposures, EMF (electromagnetic frequency radiation) may modify the effect of radon exposure by increasing the tendency of aerosol particles to vibrate; this has been hypothesized to increase the “stickiness” of room aerosols to which radon attaches, increasing their tendency to deposit on other room surfaces, and reducing inhaled radon (O’Brien 1996). The effect of such vibration on absorbed dose is unknown. Since the attached radon fraction is less efficiently absorbed than the unattached however, the potential impact of EMF on radon toxicity is likely to be small.

2.4.2.10.4 *Biological Factors*

Tidal volume and breathing frequency determine flow rates, which together with minute ventilation influence the dose of RDPs to target cells (Samet 1989). Nasal versus oral breathing increases the deposition of unattached RDP in the nose versus the target bronchial epithelium (Samet 1989). Bronchial morphometry, (ie. the sizes and branching patterns of the airways), determines deposition patterns. Both higher mucociliary clearance and thicker mucous layer lining tend to reduce dose per unit exposure. The above biologic factors vary (to differing degrees) with age, gender, physical activity level, health status, exposure to tobacco smoke, airway inflammation (NRC 1991), and ethnicity (Sexton et al., 1993). Such factors should therefore be considered as potential effect modifiers during analytic model development. Only the most important of these, tobacco smoke, will be considered here in detail.

The net effect of either active or passive exposure to tobacco smoke on lung cancer risk represents the summation of factors increasing and decreasing dose to target cells (Samet 1989). Factors associated with smoking which increase risk include: a) exposure to the carcinogenic components of the cigarette smoke itself; b) greater central deposition of RDPs in the lungs; c) increased permeability of the airways; and d) reduced rate of mucociliary clearance (a very small effect according to James 1988). Risk reducing factors include: a) increased thickness of the mucous and mucosal edema due to increased mucus production; b) increased room aerosols (decreasing the unattached RDP fraction); and c) possibly increased room ventilation.

Increased aerosols arising from cigarette smoke in a room would result in a relative underestimate of RDP exposure; and since odds ratios ORs are estimated on the basis of the measured gas concentration, these ORs will be overestimates. On the other hand, if smokers are more likely than others to keep windows open, this higher ventilation rate in homes of smokers, would tend to decrease F. The net balance is uncertain, but both confounding and effect modification are possible, because smoking may be associated with absolute exposure (radon gas), is the most important risk factor for lung cancer, and affects exposure-dose relationships.

It should be stated that if the effect estimates from the residential studies are employed as the basis of a risk assessment, and the same sources of error described above are present in the assessment of population exposure, the exposure errors will essentially cancel out.

2.5 ANALYTIC ISSUES

2.5.1 MODEL FORM

Simple analyses of data from 11 miner cohorts demonstrated that the relative risk (RR) for each cohort was consistent with a linear increase in the risk from exposure measured in working level months (WLM) (Lubin et al., 1995c, 1994a). An exposure-response relationship consistent with linearity has also been demonstrated by Gilbert et al., (1996) in rats exposed to radon, RDPs and uranium ore dust (the latter to replicate the equilibrium conditions of a mine), at levels from 20 - 640 WLMs. At doses where cell-killing is not significant, linearity in dose-response curves is also empirically derivable from cellular radiobiological data (see NRC 1988 page 416). Linearity also has a rational biophysical basis. Inspection of radon levels typical in the residential studies²¹ suggests a range of radon exposures consistent with the range of observed linearity.

Incorrect model specification can lead to biased effect estimates and standard errors, leading to model-based confidence intervals which do not cover the true effect measure at the nominal coverage rate, particularly for extreme exposure levels. All study investigators assumed a logistic regression model in their principal analyses. This assumes an exponential increase in the rate (or rate ratio) of disease with a simple linear increase in the exposure or any of the covariates, and multiplicative joint effects between the exposure and all covariates (Greenland 1989). Only Pershagen et al., (1994) supplemented their main analyses using the exponential logistic model with a linear excess relative risk model, however. This followed Pershagen et al.'s (1992) observation of improved estimation after logarithmic transformation of radon exposure data.

²¹ Refer to Appendix A Table 4 for details.

The estimated rate ratio from the logistic model,

$$e^{\beta x} = 1 + \beta x + \beta^2 x^2/2 + \beta^3 x^3/6 + \dots$$

At very small risks, or when the product βx is sufficiently small (ie. < 0.1), the terms in the squares and higher are negligible, and $e^{\beta x}$ closely approximates $1 + \beta x$ (the RR from an ERR model). Another way of saying this is that the logistic model is close to linear at low exposures. Thus, at the low ends of the curves, little difference in rate ratio estimates is predicted from the logistic versus the linear ERR models. The implications of model misspecification would perhaps be most important when a strictly modelled result is reported for ever-smokers and never-smokers combined, assuming a multiplicative relationship for the relationship between radon and smoking on odds of lung cancer without attempting to fit an interaction term to detect deviations therefrom. This was done in two of the residential studies (Schoenberg et al., 1990; L  tourneau et al., 1994) which did not report results specific to smoking strata (see Appendix A, Table 6).

Although subject to uncertainty arising from limited information on smoking habits (especially relating to cessation) in miners, a sub-multiplicative relationship between for radon and smoking in determining lung cancer risk is supported by numerous analyses (Hornung & Meinhardt 1987; Hornung & Samuels 1981; Lubin et al., 1994a; Lubin & Steindorf 1995; NRC 1988; Whittemore & McMillan 1983; Moolgavkar et al., 1993; Thomas et al., 1985; Yao et al., 1994). Results of experimental animal studies of the relationship between radon and smoking, have been somewhat inconsistent among themselves and with data from miners. Piao and Hei (1993) observed an additive relationship, Monchaux et al., (In Press), a multiplicative one, and Cross et al., (1992) observed mixed results.

What is clear is that more data than has been provided in isolation or inter-comparison from the eight published case-control studies of residential exposure will be necessary to elucidate the possible joint effects of radon and smoking.

The implications of model uncertainty, although expected to be small for radon alone, will be unclear until model sensitivity is examined during an analysis of the combined data.

2.5.2 CONFOUNDING AND EFFECT MODIFICATION

Covariates, including strong risk factors for lung cancer and factors that potentially confound or modify the relationship between radon and lung cancer, will be discussed in this section. A general discussion will precede a brief analysis of specific covariates, to include: active and passive smoking, socio-economic status, age, gender, sleeping with an open window, other environmental exposures, and histologic type.

Radon levels are highly variable between geographic areas, depending on the concentration of naturally present uranium in the local rock and soil (IARC 1988). Within the same geographic area, home radon levels can vary by several orders of magnitude (Marcinowski et al., 1994; White et al., 1992), depending in part on construction methods, type of basement, and ventilation. Although marked differences may exist between neighbouring houses, caused by differences in the air permeability of the subjacent soil and in the substructure and ventilation of the house, there is a substantial clustering of high radon concentration homes by area (Nero 1988b). Thus, since residential area may be associated with smoking (US DHHS 1989, US EPA 1992), occupation (Siemiatycki 1991), socio-economic status (Bourbonnais & Siemiatycki 1996), and other independent risk factors for lung cancer, confounding is of concern. Strong risk factors for the disease in question should be included in logistic models, even if they are not associated with exposure, as their absence can bias estimates of exposure effect (Gail et al., 1984).

Confounding can be reduced or prevented through restriction or matching (with adjustment), and controlled in the analysis through the inclusion of covariates in the risk model. The accuracy with which covariates are measured is critical; both differential and non-differential mis-classification of a confounder lead to residual confounding (Greenland) and to bias in either direction, depending on the direction of confounding (Rothman 1986). Simulations by Becher (1992) demonstrated that categorization of a continuous risk factor into a dichotomous variable can yield high levels of residual confounding. If a confounder is strong and the exposure-outcome relationship is weak or zero, errors in confounder specification could lead to extremely misleading results (Rothman 1986), sometimes increasing net bias (Greenland and Robins 1985). Although Grayson (1987) has suggested 'change-in-estimate' as the useful criterion for confounding, Greenland and Robins (1985) have demonstrated that this criterion can be misleading when mis-classification is present. These results are a matter of substantial concern in light of the potential radon exposure mis-classification discussed above, and the limited quality of many of the categorical covariates employed in the reviewed analyses²².

Four of the eight studies (Auvinen et al., 1996; L  tourneau et al., 1994; Schoenberg et al., 1990, Pershagen et al., 1994) matched subjects on variables not controlled for in their final models, thus inducing some degree of confounding in their final estimates of effect if the matching variables were, indeed, risk factors (Rothman 1986). If model fit was the criterion by which authors retained variables, and since any number of combinations of different covariates might result in a similar model fit (Robins and Greenland 1986), little can be concluded from the variety in the sets of covariates included in final models. Likewise, stepwise selection algorithms (as used by L  tourneau et al., 1994) can easily lead to invalid estimates of effect (Greenland 1989).

²² Please refer to Appendix A Table 5 for study details regarding control of confounding, and Table 6 for details regarding investigation of possible effect modifiers.

Lubin et al., (1994a) in their combined re-analysis of 11 miner cohorts observed that the predicted ERR/WLM decreased with attained age, with time since exposure, with time after cessation of exposure, and was unrelated to age at first exposure. Most importantly, ERR/WLM was 3-fold higher in never smokers than ever smokers. The above factors may be modifiers in a multiplicative model. Only three of the residential studies investigated the potentially modifying effect of age (Alavanja et al., 1994; Auvinen et al., 1996; Pershagen et al., 1992). The other factors, radon progeny dose rate and years since last exposure, are not directly amenable to study in the present setting, the former because of the fixed time window, and the latter because in a residential setting exposure is always non-zero.

2.5.2.1 Active Smoking

With respect to the potentially confounding effect of smoking, Archer (1997) has made the following argument:

"Lower socio-economic groups are known to smoke more and have higher lung cancer rates than the higher socio-economic groups (Syme et al., 1976, Schuman 1977). They also tend to live in smaller and older houses than do higher SES groups (because of their lower income). Small, older houses tend to be less 'tight' than larger newer houses that have higher radon levels (Papastefanou et al., 1994; Cohen et al., 1989, 1991). In addition, smoking odours may prompt occupants to ventilate their houses more than nonsmokers do. Because of the indirect relationships of radon with smoking it will confound the lung cancer-radon relationship in a way that may not be dealt with by the usual smoking adjustments."

Archer suggests that net effect of these factors be a bias toward the null, a surprisingly simple conclusion. It is clear, however, that smoking presents enormous potential problems for adequate control since it is: 1) the strongest lung cancer risk factor; 2) a potential confounder; 3) possibly subject to temporal effect modulation, which will add noise to measurement; and, 4) a potential effect modifier in a multiplicative or additive model, requiring estimation of an interaction term to estimate population-specific overall effects.

Moreover, as demonstrated by Upfal et al., (1995), if the true underlying joint effect between radon and smoking is less than multiplicative, it is very difficult to detect a main effect due to radon in mixed smoking and nonsmoking populations, as the main effect estimate is diluted by the much smaller effect of radon among smokers. This masking of a weak effect by a stronger one is considered by Upfal et al., (1995) as an example of the phenomenon underlying the advice of Rothman and Poole (1988) to restrict studies to lower-risk populations (ie. non-smokers) in order to “strengthen weaker associations”. Even a relatively small error in measuring a large risk factor such as smoking (which explains most cases) may obliterate or spuriously exaggerate an effect of radon. Effectively, the noise from the smoking signal could be mistaken for, or obliterate, any signal from the much smaller radon effect.

Without considering exposure measurement error, mobility, or the employment of a questionable model form, but only the impact of a sub-multiplicative joint effect of smoking and radon as estimated by Lubin et al., (1994a), Upfal et al., (1995) estimated that the number of cases (and equal number of controls) would be 577 in a never-smoking population versus 5651 in a population with a mix of smokers (of 28.5, 40.4 and 31.3 % never, current and former smokers, respectively) in order to detect an attributable risk of 5000 lung cancer deaths per year with power of 0.80 and a two-tailed alpha level of 0.05. This population mix of smokers is better than that achieved in any of the studies. Unfortunately, only one group of investigators (Alavanja et al., 1994) restricted entry to current non-smokers, but allowed ex-smokers.

Unreliability in a surrogate confounder (proxy data on smoking) is much more apt to produce seriously misleading inferences than is unreliability of surrogate measures for disease and exposure (Kupper 1984). The use of proxies for live controls as by (Schoenberg 1990 & Pershagen et al., 1994) may have unnecessarily compromised smoking data. This is supported in the former case by the necessity of an interaction term between smoking and

source of information. In the latter case, the second control group, which had been matched to cases on vital status was later merged with the first control group without reporting whether vital status had an impact on information quality. Moreover, a stronger exposure-response was observed by Alavanja et al., (1994) when he restricted his analysis to live cases (all controls were live).

All but three studies (Auvinen et al., 1996; Pershagen et al., 1992 & 1994) used an imprecise measure of smoking exposure (such as ever-never), or combined intensity and duration of active smoking into a cumulative index (e.g. pack-years) which is known to be inaccurate for this exposure (Thomas 1987, page 73S).

Odds ratios within smoking strata²³ were non-comparable across studies; half of the studies created combined radon-smoking strata by crossing radon and smoking categories (using never- or 'non'-smokers at the lowest study-specific radon level as the reference group for all cross-categories of radon and smoking), while the other half stratified purely by smoking (using the lowest radon level as the reference group within each smoking strata). Moreover, within each stratification method, cut-offs used to define smoking and radon categories varied. No consistent pattern was discernable across studies. A combined analysis will provide an opportunity to form comparable strata across studies for examination of possible trends in radon effect within and between smoking subgroups.

2.5.2.2 Passive Smoking

Only four studies addressed environmental tobacco smoke (ETS); none of these considered its potential to modify the effect of radon exposure. Blot et al., and Schoenberg et al., (1990) both considered "spouse smoking" only; these studies, and that of Auvinen et al., (1996) may have reduced precision by dichotomizing ETS exposure; Alavanja et al., (1994)

²³ See Table 7 of Appendix A.

considered “passive smoking” in a model adjusted also for age. Neither of the other two authors retained this variable in their reported analyses. Since the relative risk associated with exposure to environmental tobacco smoke (estimated as 1.9 by US EPA 1992b) exceeds that expected of residential radon on the basis of the miner studies, lack of accurate control of this factor creates uncertainty in the interpretation of results.

2.5.2.3 Socio-Economic Status

Socio-economic status (SES) must be a surrogate for other, possibly unknown, risk factors, since there is no credible reason for SES *per se* to have a direct effect on lung cancer. SES remains a predictor of lung cancer after adjustment for smoking, occupation and other covariates (Bourbonnais & Siemiatycki 1996). Since SES may also be associated with radon exposure, it is a potential confounder. While none of the reviewed studies considered SES *per se*, all but three studies (Auvinen et al., 1996; Pershagen et al., 1992, 1994) assessed a correlate of SES, namely education. The requirement of listed telephones for controls but not cases in a city in which only 7% of Winnipeg residents without phones would have eliminated those with the lowest income level from the control group. Similarly, the requirement by Schoenberg et al., (1990) and Alavanja et al., (1994) for controls under 65 to be licenced drivers, with no reciprocal requirement for the cases, could have resulted in controls of a higher socio-economic status than cases. Since education is not a perfect proxy for SES, at least residual confounding is a possibility in all studies. Whether income level is important over and above education level should be evaluated in further analyses of combined data.

2.5.2.4 Age

Lubin et al.'s (1994a) re-analyses of cohort studies in miners demonstrated that the ERR of lung cancer associated with radon declines with age and drops sharply after 65 years (Lubin et al., 1994a). Muller et al., (1985) also reported that the rate of increase in RR with age decreased after age 55. Inspection of age-specific ORs is therefore important²⁴. Three studies (Alavanja et al., 1994; Auvinen et al., 1996; Pershagen et al., 1992) provided such data, and all were consistent with the miner results, in that the strongest associations between radon exposure and lung cancer were in their youngest age groups. The internal consistency of the age-specific results lends support for a causal association between residential exposure to radon and lung cancer.

2.5.2.5 Gender

Because females have traditionally spent a greater portion of their time in their home than males, measurements of home radon exposure would tend to be more representative of total radon exposure among females than males. Notably, the three studies which provided evidence of a significant trend in odds ratios among radon categories were among the four (Alavanja et al., 1994; Blot et al., 1990; Schoenberg et al., 1990; Pershagen et al., 1992) which restricted their subjects to females. The exception was the Chinese study by Blot et al., (1990). One study (Ruosteenoja 1990) restricted its subjects to males. The remainder (Auvinen et al., 1996; Létourneau et al., 1994; Pershagen et al., 1994) included both genders. Although the former two studies matched subjects by gender, they did not include it in their final models. Pershagen et al., (1994) included gender in their model, implying its importance. Auvinen et al., (1996) provided gender-specific estimates of radon effect, but did not include gender in the final model (the p-value for the interaction of gender with radon was 0.1).

²⁴ Please refer to Appendix A Table 8 for details regarding treatment of age in the reviewed studies.

In addition to having radon measurements which more accurately represent past exposures, females may have lung cancer susceptibilities different from than of males; female breathing rates and other determinants of exposure-dose relationships differ from those in males (NRC 1991; Sexton et al., 1993). The possible modification by gender of putative radon effects remains an important question to be addressed in analyses of combined data.

2.5.2.6 Other Environmental Exposures

Concurrent exposure to air pollution was explicitly considered by two of the eight studies (Ruosteenoja 1991; Blot et al., 1990), both of which considered it as only a dichotomous variable. If associated with radon exposure, such crude measurement of air pollution would result in residual confounding. Blot et al., (1990) observed significantly higher levels of radon in homes with elevated levels of air pollution from heating and cooking sources. Radon and air pollution were highly correlated so that regression coefficients would have been unstable, resulting in imprecision. Moreover, high levels of room aerosols, to which radon progeny attach, would reduce the absorbed lung radiation dose for a given measured level of radon gas, reducing the magnitude of putative effect. Given all of the above, the interpretability of the Chinese study is rather questionable.

Among those classified as exposed to air pollution in South Finland, OR estimates were reduced, indicative of possible effect modification, but the large confidence intervals enclosed unity (Ruosteenoja 1991); no multi-variable analyses were conducted. The potential importance of air pollution was also suggested by the finding of significant heterogeneity in radon risk with degree of urbanization, a significant relative increase in risk being restricted to the more urbanized area, in Lagarde's (1992) re-analysis of Pershagen et al.'s (1992) Stockholm study. Also indicative is the much greater gradient of response per unit exposure observed by Pershagen et al., (1992) in their Stockholm study (an ERR of 2.7% per WLM) compared to their subsequent nation-wide Swedish study (an ERR of 1.3% per

WLM calculable from their data). An analysis of combined data should attempt to provide results stratified by level of air pollution.

All but three of the residential studies (Alavanja et al., 1994; Blot et al., 1990; Pershagen et al., 1992) considered occupational exposures. Two of the five studies which collected data on such exposures included a significant term for occupational exposure in their final models: 1) in Sweden, Pershagen et al. (1994) used a three point scale for exposure to known and suspect human carcinogens (according to the International Agency for Research on Cancer's coding system); 2) in New Jersey (Schoenberg et al., 1990), used a dichotomous indicator for high versus low risk occupational history. The latter study also collected data on education but excluded this variable from its final model, suggesting possible collinearity between education and occupation.

The most sophisticated occupational exposure assessment methods were employed in the Winnipeg study, which benefited from collaboration with Siemiatycki et al., authors of large-scale case-control study designed specifically to investigate occupational exposures in Montreal. L  tourneau et al.'s coding scheme involved six indicator variables: three of these represented degree of confidence in elevated risk associated with past exposure to a job category; the other three represented the degree of confidence in risk associated with past exposure to one or more of 44 specific agents. Consistent with Siemiatycki et al.'s (1994) observation that adjustment for occupational exposures did not materially affect the odds ratio between smoking and lung cancer, the published Winnipeg model, which included education, excluded the occupational variables. The two other radon studies which investigated occupational exposures, Pershagen et al.'s 1992 study in Stockholm Sweden, and Auvinen et al.'s 1996 study in Finland, also excluded occupational exposures from their final predictive models; these studies did not report having collected information on socio-economic correlates.

The above evidence suggests that unless occupational exposures modify radon risk, their exclusion from radon risk models may be un-important.

2.5.2.7 Sleeping With an Open Window

As mentioned previously, only one set of authors (Pershagen et al., 1994) investigated the impact on the radon-lung cancer relationship of sleeping with an open window (SWOW). Subjects not reporting SWOW had double the ERR of lung cancer per unit exposure concentration than subjects who did report SWOW. Lubin (1994) described the SWOW habits of current inhabitants as confounding measured values. For confounding to occur, SWOW would have to be an independent risk factor for lung cancer (in the non-radon exposed) as well as being associated with exposure (in the controls). SWOW meets the latter condition through its previously discussed effect on radon concentrations. Both Lubin (1994) and Archer (1995) appear to interpret the heterogeneity in stratum-specific results (for subjects reporting SWOW versus those who did not) as evidence that SWOW is an independent risk factor. More precisely, however, the issue of SWOW can be seen to satisfy only the definition of an effect modifier, the effect of SWOW on lung cancer risk being mediated by its effect on exposure-dose relationships, specifically, on the F factor assumed constant when translating measured concentrations to assumed RDP exposure. Although confounding could be possible if smokers may be more likely than non-smokers to sleep with open windows (accounting for some of the variability in the radon smoking relationship), in the presence of strong effect modification, assessment of confounding is complex and subjective, and of secondary importance to the elucidation of the former.

2.5.2.8 Histologic Specificity

No histologic specificity was observed by Létourneau et al., 1994 or Blot et al., 1990, consistent with their lack of finding of a main effect of radon on lung cancer overall. No

overall pattern can be discerned amongst the rest of the studies, each of which demonstrated some “significant” pattern. It has been implied that the lack of consistency between the miner studies, allegedly showing a specificity for small cell carcinoma, and residential studies, which do not show this specificity, brings into question the existence of an effect of residential radon exposure on lung cancer risk. Such an argument appears to have little support from the data. The evidence for specificity even in miner studies is weak (Samet 1989), and both heterogeneity of type in whole tumours (Roggli et al., 1985; Saba et al., 1983) and the potential for mis-classification of samples therefrom, are large (EPA 1992; Steinhäusler 1988; Thomas et al., 1993).

A recent miner study by Yao et al., (1994) suggested that both adenocarcinoma and small cell are associated more strongly with radon exposure than squamous at younger ages, for both smokers and non-smokers. With respect to smoking, however, a change in the distribution of lung cancer types over time associated with the introduction of filtered cigarettes, from a predominance of squamous with unfiltered cigarettes to a predominance of adenocarcinomas with filters, has been suggested by research in both rats (Monchaux et al., In Press) and humans (Levi et al., 1997). The majority of the tobacco-related studies recently reviewed by the EPA (1992) demonstrated an increase in risk for lung cancer for all four major histologic groups of lung cancer; a more recent study also failed to show any predisposition toward a specific histologic type of cancer (Hegmann et al. 1993). According to the most recent NRC report (1998), histologic type has not proven to be an indication of the causes of lung cancer either in the general population or among miners occupationally exposed to radon. Historical observations of a weak association between specific histologic types of lung cancer and radon were likely thus a reflection of mis-classification, confounding and/or type I error.

2.6 COHERENCE OF THE EVIDENCE

2.6.1 BIOLOGIC PLAUSIBILITY AND EVIDENCE OF CARCINOGENICITY

The biologic plausibility of radon-induced carcinogenesis is based on the fact that alpha particles released during the radioactive decay of radon are high linear energy transfer (LET) electromagnetic radiations, which can produce localized and severe damage in biological systems (Hobbs and McLellan, 1986). Alpha particles which 'hit' or transfer energy to cellular DNA could therefore result in genetic mutations, which could lead to cancer. Empirical evidence of the carcinogenicity of radon includes *in vitro* studies demonstrating radon's potential for cellular transformation in rodent cells (NRC 1994); *in vivo* experimental studies in three animal species (IARC 1988), demonstrating its lung carcinogenicity and (in rats) synergy with tobacco smoke exposure (NRC 1988; Cross et al 1990, 1992; Cross 1994); as well as human cohorts in five types of mine environments (Lubin et al., 1994a, 1995c). The evidence available even prior to 1988 led the International Agency for Research on Cancer to classify radon²²² as a known human carcinogen (IARC 1988).

2.6.2 MAIN LIMITATIONS OF THE REVIEWED STUDIES

Ideally, for a study to quantify an effect of radon on lung cancer risk, it should be free of selection bias, utilize reliable and valid radon measurement techniques, obtain complete measurements on all subjects for a minimum of 40 years prior to the index date, be conducted in a low residential mobility area with a wide range of radon exposures, contain a large proportion of non-smokers, and include a sufficiently large number of subjects to detect a small magnitude of effect. As can be seen from the summary of limitations outlined in Table 2.3, such ideals were unable to be met in any individual study. The most important limitations are highlighted below:

Table 2.3

Summary of Main Study Limitations

- [↓] Denotes an attenuation of the estimate which affects (ie. is not reflected in) the confidence bounds; for radon measurement error, number of arrows indicates assumed relative magnitude of bias, taking in to account duration and seasonality of monitoring, and average concentration level;
- × Denotes an excluded factor;
- EM Effect modification; ETW: Exposure time window; SES: Socio-economic status; SWOW Sleeping with an open window not considered

Location / Author	Selection & Disease Misclassification	Radon Information	Smoking Information	Other Covariates	Power (# Cases)
Shenyang China: Blot et al., 1990	No discernable source of bias	Dosimeter imprecision [↓] Occupancy & other exposure to radon [↓] Only 1 home measured [↓] Variable ETW & no latent period assumed [↓] Spatial & temporal uncertainty [↓] Equilibrium conversion ?	Active smoking - residual confounding ? Passive smoking - spouse only (dichotomous)	♀ Only - good ! Pollution - residual confounding; SES - confounding ? Age - EM not considered SWOW [↓] + ?	308 Mobility ↓ Exposure moderate Smokers and non-smokers mixed ↓
New Jersey USA: Schoenberg et al., 1990	No discernable source of bias in spite of author-cited concern	Dosimeter imprecision [↓↓] Only 1 home measured - [↓↓] Occupancy & other exposure to radon [↓] Spatial & temporal uncertainty [↓] Equilibrium conversion ?	Active smoking in 86% of cases; Passive smoking - spouse only (dichotomous)	♀ Only SES-residual confounding ? SWOW [↓] + ? Age - EM not considered	433 Mobility ↓ Exposure low - ↓ Smokers and non-smokers mixed ↓
South Finland: Ruosteenoja 1991	No histopathologic confirmation of cases [↓]	Winter measures [↓] Dosimeter imprecision [↓↓] Coverage of ETW [↓↓] Imputation for TWA [↓] Variable ETW [↓] Occupancy & other exposure to radon [↓] Spatial & temporal uncertainty [↓] Equilibrium conversion ?	Active smoking - 98% cases Passive smoking - ×	♂ Only - good ! SWOW [↓] + ? SES-residual confounding ? No multivariable adjustment	238 Mobility ↓↓ Exposure high - good ! Smokers and non-smokers mixed ↓

Location / Author	Selection & Disease Misclassification	Radon Information	Smoking Information	Other Covariates	Power (# Cases)
Stockholm Sweden: Pershagen et al., 1992	No discernable source of bias	Dosimeter imprecision [11] Occupancy & other exposure to radon - [1] Coverage of ETW [11] ? could have varied differentially Spatial & temporal uncertainty [1] Equilibrium conversion ?	Active smoking in 81% of cases; passive smoking - X	♀ Only - good ! Age - EM observed but no interaction term in model; SES - confounding ? SWOW [1] + ?	201 Mobility I Exposure moderate Smokers and non-smokers mixed I
Winnipeg Canada: Létourneau et al., 1994	Potential controls not recorded.	Dosimeter imprecision [1] Coverage of ETW - only 5/9 homes in study area [11] Occupancy & other exposure to radon [1] Spatial & temporal uncertainty [1] Equilibrium conversion ?	Active smoking in 97% of cases; Passive smoking - X	Gender - EM ? Age - EM ? SES-residual confounding ? Ethnicity - observed to vary by outcome but not adjusted; SWOW [1] + ?	738 Mobility II Exposure moderate Smokers and non-smokers mixed I
Missouri USA: Alavanja et al., 1994	24% of cases unconfirmed [1]	Dosimeter imprecision [111] Coverage of ETW [1] Occupancy & other exposure to radon [1] Spatial & temporal uncertainty [1] Equilibrium conversion ?	Current non-smokers - good ! Passive smoking - in one model	♀ Only - good ! Age - EM observed but no interaction term added SES-residual confounding ? SWOW [1] + ?	538 Mobility I Exposure low
Sweden: Pershagen et al., 1994	—	Winter measures [1] Dosimeter imprecision [11] Coverage of ETW [1] Crude CUM measurement [1] Occupancy & other exposure to radon [1] Spatial & temporal uncertainty [1] Equilibrium conversion ?	Active or unknown smoking in 81% of cases; Passive smoking - X	Gender - EM ? Age - EM ? No adjustment for 2 matching variables (cal year of res & resp type); SES confounding ? SWOW [1] + ?	1281 Mobility I Exposure moderate Smokers and non-smokers mixed I linear model I
Finland: Auvinen et al., 1996	—	Dosimeter imprecision [11] Only 1 home measured [1] Variable ETW [11] Occupancy & other exposure to radon [1] Spatial & temporal uncertainty [1] Equilibrium conversion ?	Active smoking in 91% of cases	Smoking - res. confounding SES confounding ? SWOW [1] + ?	1055 Smokers and non-smokers mixed I

2.6.2.1 Measurement Error

Random error of measurement attenuates effect measures. Simulations conducted by Lubin (1995a) assuming a true radon effect as estimated by the BEIR IV report (NRC 1988), demonstrated that under “realistic conditions” of mobility (three moves - moderate mobility compared to some of the reviewed studies), and exposure error (whose extent was only guessed), for a study to have 90% power to reject the null hypothesis at an alpha level of 0.05 would require 7000 - 18,000 cases (for equal case: control ratio). Sample sizes in the reviewed studies ranged from 201 to 1281 cases. Moreover, Lubin’s analysis did not consider the potentially confounding or modifying effect of covariates.

2.6.2.2 Effect Modifiers & Required Sample Size

Sample size requirements generally tend to increase in the presence of effect modifiers; potential modifiers included smoking, gender, age, air pollution, and SWOW. Power may be further compromised by mis-classification of the estimates of interacting exposures (Samet & Lambert 1991). Without considering exposure measurement error, mobility, or the employment of an incorrect model, but only the impact of an assumed sub-multiplicative joint effect of smoking & radon, Upfal et al., (1995) estimated that studying a mixed population of smokers and nonsmokers would require an order of magnitude (10 times) more subjects than studying nonsmokers alone; moreover, the mix of smokers and non-smokers which these authors assumed was more favourable than that present in the reviewed studies (with the singular exception of Alavanja et al., 1994 who studied current non-smokers).

2.6.2.3 Incremental Risk per Unit of Exposure

Since the miner studies were based on a linear model of RR, the odds ratio at typical residential exposure levels is uncertain, but expected to be small. Based on a 25 year

residential exposure to the US EPA & US Department of Health and Human Services (1994) action level (150 Bq/m³)²⁵ a relative risk of less than 1.2 was predicted.

2.6.2.4 Level of Exposure

Few studies were conducted in areas where radon exposures were high and heterogeneous. Moreover, dosimeter measurement error is larger (in relative terms) at lower exposure levels. The larger relative error may further attenuate an already small anticipated effect.

2.6.2.5 Comparability of Study Populations

The studies were conducted among target populations which differed geographically and climatically, resulting in differing exposure conditions and levels. The prevalence of potentially important modifying factors, and degree of effective control of possible confounding variables also varied across study, limiting the comparability of overall results. Although small stratum-specific sample sizes limited precision, several studies demonstrated exposure response relationships within smoking²⁶- and age²⁷- specific strata, which were either statistically significant or of sufficient magnitude to warrant further investigation.

2.6.3 CONSISTENCY OF THE RESULTS

Figure 2.3 (abstracted from Lubin 1994) overlays the relative risk estimates from the first seven published residential studies to that extrapolated from the BEIR IV risk model assuming constant exposure between 35 and 65 years of age. The first and perhaps most

²⁵ Only 1 study had average exposures above this level

²⁶ Please refer to Appendix A Table 7 for details regarding treatment of active and passive smoking in the reviewed studies.

²⁷ Please refer to Appendix A Table 8 for details regarding treatment of age in the reviewed studies.

obvious observation to be made from Figure 2.3 is that overall precision (as evidenced by the widths of the category-specific confidence intervals) is so extremely (but understandably) low that the results are at once consistent with both the BEIR IV miner-based model and a no-effect model. Recalling the numerous factors affecting accuracy and precision in estimation, added to analytic issues which obscure study inter-comparison and interpretation, important effects seem plausible. Even a small elevation in RR would be important because it would result in large numbers of excess cases from widespread population exposure.

The second point to be noted with respect to Figure 2.3 is that the category- mean odds ratios from different studies are not directly comparable because of the dependence of categorical ORs on baseline reference group, which differed by study. Table 2.4 details these reference groups and the ORs associated with comparison groups from the final logistic regression models among all subjects, and highlights in bold those category-specific estimates which were statistically significant; only those of Pershagen et al., (1992 and 1994) stand out. As emphasized by Samet (1994), however,

"...findings of studies whose upper confidence intervals for risk estimates extend to values of potential public health concern and overlap with the range of risks projected from miner studies can not be discarded on the basis of lack of "statistical significance". It would be erroneous to label such results as "negative" on the basis of an imprecise estimate of risk."

Moreover, four of the eight studies had category-specific ORs which suggested exposure-response trends; in three of these (Alavanja et al., 1994; Pershagen et al., 1992; and Schoenberg et al., 1990) such trends achieved statistical significance.

Figure 2.3²⁷

Category-specific Odds-ratios For Radon Concentration
Estimated From Seven Case Control Studies

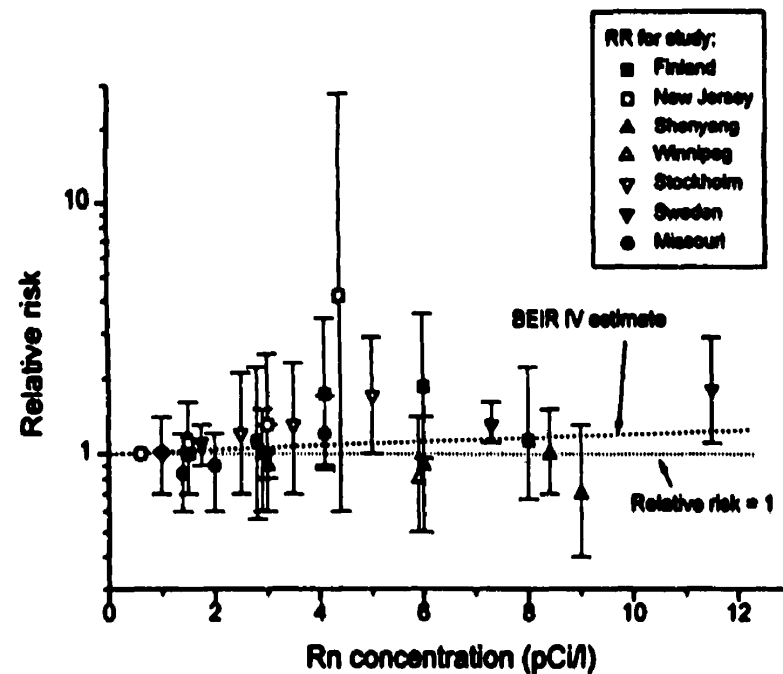


FIGURE 1. Relative risks (RR) for lung cancer by picocuries per liter (pCi/l) for the seven case-control studies of residential radon (Rn) exposure listed in table 1. Also shown are a plot of predicted relative risk for residential exposure from the age of 35-65 years using the model of the Committee on the Biological Effects of Ionizing Radiations (BEIR IV) (2) adjusted to 1 pCi/liter as the referent exposure and a plot of a relative risk of 1.0.

Table 2.4

Reference Exposure Categories and Associated Results

Location / Author	Radon Concentration (TWA) in Bq/m ³				Cumulative Radon Exposure			
	Reference Category for OR	Exposure Categories	Reported OR (90% CI) [95% CI]	Trend or Slope (continuous exposure)	Reference Category (Bq/m ³ -year)	Category Cutpoints	Results	Trend or Slope (continuous exposure)
Shenyang China: Blot et al., 1990	3.7 - 73	74 - 147 147 - 292 ≥ 296	0.9 0.9 0.7	No	--	--	--	--
New Jersey USA: Schoenberg et al., 1990	0 - 36	37 - 73 74 - 147 ≥ 148	1.1 [0.8-1.7] 1.3 [0.6-2.9] 4.2 [1.0-7.5]	Trend p=0.04 one-sided	< 925	925 - 1849 1850 - 3699 3700 - 5735	1.2 (0.8-1.9) 0.9 (0.4- 2.2) 7.2 (1.0-50.3)	RR 3.4% per WLM (trend p = 0.09 one-sided)
South Finland: Ruosteenoja 1991	0 - 108	109 - 150 151 - 196 197 - 264 ≥ 265	0.8 [0.4-1.5] ^a 1.7 [1.0-3.0] 1.8 [1.0-3.0] 1.2 [0.7-2.1]	N.S. (pg 72)	--	--	--	--
Stockholm Sweden: Pershagen et al., 1992	0 - 75	76 -110 111 -150 ≥ 151	1.2 [0.7-2.1] 1.3 [0.7-2.3] 1.7 [1.0-2.9]	Trend p=0.05 continuous; p < 0.02 categorical	< 1250	1251 - 2500 2501 - 5000 ≥ 5001	1.4 [0.8-2.6] 1.7 [1.0-3.1] 2.3 [1.1-4.5]	ERR 2.7% per WLM or 15% per kBq/m ³ yr (p = 0.01) ^b
Winnipeg Canada: Létoirneau et al., 1994	--	--	--	--	0 - 1800	1801 - 3600 3601 - 7200 ≥ 7201	0.6 [0.3-1.2] ^c 0.8 [0.4-1.6] 1.6 [0.9-2.7]	No
Missouri USA: Alavanja et al., 1994	3.7 - 29.5	30 - 44 45 - 63 63 - 91 91 - 566	1.0 [0.7-1.4] ^d 0.8 [0.6-1.2] 0.9 [0.6-1.3] 1.2 [0.9-1.7]	Trend p = 0.01 ^e	First Quintile	2nd Quintile 3rd Quintile 4th Quintile 5th Quintile	1.5 1.3 1.6 1.9	WLM estimated Trend p = 0.06 (direct interview)

Location / Author	Radon Concentration (TWA) in Bq/m ³				Cumulative Radon Exposure			
	Reference Category for OR	Exposure Categories	Reported OR (90% CI) [95% CI]	Trend or Slope (continuous exposure)	Reference Category (Bq/m ³ -year)	Category Cutpoints	Results	Trend or Slope (continuous exposure)
Sweden: Pershagen et al., 1994	≤ 50	50 - 80 80 - 140 140 - 400 >401	1.1 [0.9-1.3] 1.0 [0.8-1.3] 1.3 [1.1-1.6] 1.8 [1.1-2.9]	ERR ^f = 0.18 [0.06-0.37]	--	--	--	ERR ^g 0.23 [0.07-0.51]
Finland: Auvinen et al., 1996	≤ 48	49 - 96 97 - 196 197 - 396 > 396	1.0 [0.8-1.2] 1.0 [0.8-1.3] 0.9 [0.6-1.4] 1.2 [0.7-1.9]					

Table 2.4 Notes:

TWA Time-weighted average for the studies in which measurement was attempted in more than one home per subject: these included all except Blot et al., (1990) and Schoenberg et al., (1990); Trend test not indicated due to sharp drop in OR for highest concentration.

- a** Adjusted for smoking and including both measured and estimated radon concentrations (222 cases and 379 controls). Based on measured homes only (smaller sample size of 164 cases and 334 controls) results for the upper 4 exposure levels compared to the reference group were: 1.12 [0.56], 1.73 [0.87-3.44], 1.85 [0.96 - 3.58], 1.13 [0.57 - 722]. Continuous estimate: an increase of 100 Bq/m³ associated with a risk of 0.3% [-0.3,1.0].
- b** After logarithmic transformation of cumulative radon exposure treated as a continuous variable and adjusting for occupancy (p= 0.01 without replacement for missing radon levels or adjusting for occupancy) the excess RR increased almost linearly with estimated cumulative radon exposure- having a slope of 15% per 1000 Bq/m³.
- c** Bedroom measurements for the 454 participants for whom measurement coverage was a minimum of 75% of the 5 - 30 year period prior to the index year; results also provided for basement measurements, for a shorter index period (5 - 15 years), and for all 1476 subjects, including those with less than 75% of measured values.
- d** Adjusted for age only.
- e** The most significant (p = 0.01) trends were observed for radon treated as a categorical variable (continuous variable results p = 0.04) and when adjustment was made for age and saturated fat, or for age, pack-years of previous smoking and years since smoking cessation, but CI's were not given for these analyses and no data was provided for which simultaneous adjustment was made for all of the above variables. All subjects non-smokers for the 15 years prior to recruitment.
- f** ERR = Excess Relative Risk per unit of radon; Unit of exposure defined as 2.8 pCi of radon per liter; the above estimate excludes subjects reported to sleep near an open window; importantly- no dose-response relationship was observed when the analysis included those subjects who reported to sleep beside an open window. No other investigators considered this variable in their analyses.
- g** Unit of exposure defined as a combination of the average period for which (time-weighted average) exposures were estimated (32.5 years) and a radon level of 2.7 pCi per liter- with missing measurements replaced by median levels- and after exclusion of subjects who slept with an open window.

To avoid the problem implicit in Figure 2.3 and Table 2.4 of reference-category dependent odds ratios, Table 2.5 provides a quantitative comparison of the few analyses in which effects have been estimated by modelling cumulative exposure as a continuous variable.

Table 2.5
Estimates of Radon Effect Based on A Continuous Measure of Exposure

Type of Exposure	ERR/WLM (%)	RR/WLM (%)	Investigators	Study Population
Underground Mining		0.5 - 4.0	BEIR IV, NRC 1988	Parallel analysis of 4 miner cohorts
Residential Exposure		3.4	Schoenberg et al., 1990	New Jersey
		0.8	Derived from results of Létourneau et al., 1994 ²⁹	Winnipeg; minimum 75% coverage of the exposure time-window
	2.7 ³⁰		Pershagen et al., 1992	Stockholm, Sweden
	1.3 ³¹		Pershagen et al., 1994	Sweden; excluded subjects who slept with an open window
Underground Mining	0.5 (combined) Range for individual studies: 0.2 - 5.1		Lubin et al., 1994a	Combined analysis of 11 miner cohorts

Estimates of ERR per WLM from analyses of the most recent data from 11 miner cohorts by Lubin et al., (1994a) ranged from 0.2% to 5.1%, with a combined estimate of 0.5%, without

²⁹ Based on a reported OR of 1.18 [0.95, 1.46] per 3750 Bq/m³-years, which corresponds to 21 WLM (assuming an occupancy factor of 0.8 and an equilibrium factor of 0.5), or an OR of 1.008 [0.98, 1.04] per WLM.

³⁰ Assuming an occupancy factor of 0.8 and an equilibrium factor of 0.5; no confidence limits reported.

³¹ Pershagen et al.'s (1994) ERR estimate of 0.23 was based on exposure to an average of 2.7 pCi/l for 32.5 years, or 87.75 pCi/l-years. Conversion of these units to WLM, (assuming an occupancy factor of 0.8 and an equilibrium factor of 0.5) results in an estimated ERR/WLM of 0.013 or 1.3%. Bracketed numbers are 95% confidence limits.

considering heterogeneity. As is evident from Table 2.5, ERR/WLM estimates available or calculable from published residential radon studies were consistent with those estimated in the individual miner studies. Estimates of ERR/WLM were 1.3% in Sweden as a whole (derived from Pershagen et al., 1994) and 2.7% in Stockholm Sweden (Persshagen et al., 1992). An RR/WLM of 0.8% calculated from data provided by Letourneau et al., (1994) and of 3.4% estimated by Schoenberg et al., (1990) compare well to the BEIR IV estimate of 0.5 - 4.0% (NRC 1988).

2.6.4 CONCLUSIONS

Lubin (1995a) concluded that "the degree of consistency between studies reported to date is compatible with the consequences of population mobility, mis-specification of cumulative radon-progeny exposure, and the inability to measure radon levels accurately in past residences". Adding to this list the many other uncertainties explicated herein pertaining to uncertainties about the empirical induction period, model form, covariate control, differences among study populations, and most importantly, perhaps, the predominance of and inability to accurately separate out the effects of active and passive smoking, one can only conclude that the demonstrated consistency of the reviewed studies is in fact surprisingly good.

An analysis of the combined data from case-control studies of residential radon is not only warranted, but will be necessary to provide an analysis which is standardized in such a way as to provide adequate comparability of lung cancer risk estimates.

3 METHODS

3.1 WORKING DEFINITIONS AND APPROACH

This thesis is an endeavour in quantitative research synthesis, defined here as the synthesis of epidemiologic data gathered to address a particular question. Table 3.1 distinguishes between two approaches to quantitative research synthesis and provides common definitions and terminology.

Table 3.1

Approaches to Quantitative Research Synthesis

	Meta-Analysis	Analysis of the Combined Primary Data
Other Terms in Common Use:	Meta-Analysis of the Literature (MAL)	Meta-Analysis of the Patient Data (MAP) Combined Analysis Combined Re-Analysis Pooled Analysis Collaborative Re-Analysis
Working Definition:	A systematic review of evidence involving the use of statistical methods to synthesize the <i>summary level results</i> (or secondary data) from two or more similar investigations ¹ .	A systematic re-analysis of <i>individual-level</i> (or primary) <i>data</i> from two or more similar investigations

Because of the large number of terms that are commonly used to describe the two main approaches to quantitative research synthesis, a brief justification of the terms adopted in this thesis follow. The Oxford English Dictionary defines the suffix ‘meta’ as ‘of a higher or second-order kind’. The use of primary data is clearly a first order task that should be differentiated from the analysis of secondary or summary level data in a ‘meta-analysis’. Stuart and Parmar’s (1993) terms ‘meta-analysis of the literature’ (MAL) and ‘meta-analysis

¹ For other definitions see L’Abbe et al. 1987, Leviton and Cook 1981, Louis et al. (1985), Victor 1995

of the patient data' (MAP) fail to make this fundamental distinction, nor do they recognize the differences in time, expense and effort involved in the undertaking of the two approaches. Moreover, the subjects and design context of a meta-analysis are more generalized than 'patients' in a randomized clinical trial, which the MAP terminology subsumes. The next three alternative terms listed in Table 3.1 for the analysis of primary data are imprecise, since it is the data and not the analysis which is 'combined' or 'pooled'. The term 'collaborative re-analysis' is also inaccurate, as I am the sole analyst of data contributed by a group of collaborating primary investigators. I have thus referred to my approach to the quantitative synthesis of research concerning the relative risk of lung cancer associated with residential radon exposure as an 'analysis of the combined primary data' from several similar investigations. Meta-analysis is a term which I reserve for syntheses which utilize the summary-level effect estimates from several similar investigations.

There is a growing body of literature with respect to the practice of meta-analysis². The focus has been on the experimental epidemiological research design - the randomized clinical trial (RCT), to the relative exclusion of non-experimental designs (case-control, cohort, and cross-sectional studies). Indeed, some researchers believe that meta-analysis should be restricted to RCT's (Mann 1990; Stewart and Parmar 1993, who also cite Yusuf 1987). There are, however, examples of meta-analyses (Licciardone et al. 1990, Lubin & Boice 1997,) and 'combined analyses' (Cardis et al. 1995, Lubin et al. 1994a, Lubin et al. 1994b) of data from non-experimental studies, several of these in the field of radiation exposure. The same general principles as apply to combining data from randomized clinical trials apply to that of non-experimental studies, (Spitzer 1995, Beral 1995), with the added caution that the

² See, for example, an entire issue of the Journal of Clinical Epidemiology, (Volume 48, Number 1, January 1995), devoted to the Potsdam International Consultation on Meta-Analysis; For guidelines to assess the quality of a meta-analysis see L'abbe et al. 1987; For suggested procedures in the conduct of meta-analyses see Cook et al. 1995, L'Abbe et al., 1987, Mann 1990, Thacker 1988; For statistical methods see Fleiss 1993, Laird and Mosteller 1990, Greenland 1987, Greenland and Longnecker 1992, Whitehead and Whitehead 1991. Eddy et al. (1992) describe both procedures and statistical methods.

absence of random allocation of subjects to exposed and unexposed groups places a much greater burden on the investigator to control for potential biases and ensure comparability of the data³. This makes the use of primary rather than secondary data not only preferable (Farewell 1989), but perhaps essential (Beral 1995). The collection and re-analysis of individual level data provide the least biased and most reliable means of addressing questions not satisfactorily resolved by individual investigations (Stuart and Parmar 1993).

Farewell (1989) described the process of combining epidemiologic information as involving two fundamental tasks: 1) to determine if a set of studies examine the same exposure-disease relationship, and 2) to provide an analysis of the combined statistical evidence regarding the exposure-disease relationship. As for the first task, all of the studies included in the present pilot analysis evaluate the relationship between lung cancer and residential exposure to radon. In carrying out the second task, one must address the comparability of studies (NRC 1992, Farewell 1989). Essentially, study comparability refers to whether the different populations are reasonable samples of a broader population of interest (Eddy et al., 1992, Farewell 1989), and whether the types of exposure being considered are representative of those of interest (Farewell 1989). Since all of the studies included here employed the same basic case-control design and methods for radon exposure assessment (as reviewed in Chapter 2), at a basic level, comparability is assured *a priori*.

In their treatise on the topic of combining statistical information, NRC (1992) considered it appropriate to combine when 1) information sources are 'similar enough', and 2) when the 'similarity judgment' implied by the first criterion can be expressed as a 'model' or 'framework' for combining. The framework adopted in this thesis is that of a common statistical model, comprising radon exposure, estimated in similar fashion from each study,

³ Eddy et al. (1992) describe the propensity of various designs to different types of biases (page 67), and offer corrective procedures for many (chapter 15) assuming the availability of appropriate data.

and analysed as a continuous variable, a common constellation of covariates, including risk factors, the common model itself being derived through parallel analyses across site. Statistical information (that is summary estimates of effect) should *not* be combined when interaction effects (modifiers of the effect of the main exposure) are meaningfully large (NRC 1992). This is because even inclusion of an interaction term will result in an overall effect averaged over the distribution of effect modifiers represented in the studied population. In such a case, modifier-specific results (e.g. radon risk among specific age, gender, or smoking sub-groups) should be reported separately, and combined across study only within such strata.

Neuberger (1992) suggested the following procedures for the conduct of a 'combined re-analysis' of residential radon studies: 1) finding all relevant studies (both published and unpublished); 2) establishing study inclusion and exclusion criteria with respect to outcome and exposure; 3) establishing criteria for the comparability of studies with respect to sources of subjects, adequacy of data on active and passive smoking, and occupational and other information; and, 4) gaining access to the data and compiling it into a common data format.

These preparatory procedures are described for the most part in sections 3.2 to 3.5; smoking and occupational data are further discussed in the context of statistical model development; section 3.6 describes how I analysed the data in parallel across the three sites to develop a common multivariate logistic model for the prediction of the rate of lung cancer associated with radon exposure; section 3.7 describes the methods employed in combining the results of the parallel logistic analyses; methods undertaken to examine sensitivity of the results to exposure time window definition and model form (additive versus multiplicative in the rate ratio) are described in section 3.8.

3.2 SELECTION OF STUDIES FOR THE PILOT ANALYSIS

This analysis includes all North-American case-control studies of residential radon and lung

cancer which had been published prior to 1996 that involved directly measured radon gas concentrations in at least one residence for each subject. Three studies met these criteria. In order of publication, these were conducted in New Jersey USA, Winnipeg Canada, and Missouri USA, with original analyses by Schoenberg et al. (1990), Letourneau et al. (1994), and Alavanja et al. (1994), respectively. This order of presentation will be retained throughout the thesis.

By restricting ourselves to published studies we exclude only one completed but unpublished study known to the author and an international steering committee on residential radon and lung cancer. This study, undertaken in Florida, involved 200 cases of lung cancer. The principal investigator declined to analyse this data-set because the population base was migratory and measured radon levels were very low (Krewski, personal communication); the information value of this study was therefore considered low relative to other studies (either completed or presently being undertaken). It is unlikely that the international committee formed in 1995 to pool related data (US DOE/CEC 1995) is unaware of other studies of residential radon and lung cancer. Moreover, since roughly half of the studies published to date have interpreted their results as negative, there does not appear to be a publication bias in this field. Further, because of the large potential public health impact of such a widespread exposure and the large amount of controversy in the literature, it is doubtful that there has been any reluctance on the part of authors to submit for publication adequately-designed studies which yield negative findings.

3.3 STUDY POPULATIONS

3.3.1 OVERVIEW OF THE INCLUDED STUDIES

Table 3.2 summarizes the key distinctive features of the three included studies. The primary data from 1,756 cases and 2,363 population-based controls were available for analysis. The

Table 3.2

Overview of the Included Studies

Study Location & Authors of Original Analysis	Subjects Retained in Original Analysis		Proportion of each Gender	Proportion Never- Smokers		Radon Concentrations (Bq/m ³) From ATD Measures Among Controls		
	Cases	Controls		Cases	Controls	Mean	Min	Max
New Jersey USA Schoenberg et al. 1992	480	442	100% ♀	14 %	52 %	25	3	189
Winnipeg Canada Létourneau et al. 1994	738	738	34% ♀ 66% ♂	3 %	30 %	141	2	999
Missouri USA Alavanja et al. 1994	538	1183	100% ♀	70 %	83 %	63	4	459
Totals	1756	2363						

♀ Female

♂ Male

Bq/m³ Becquerels per cubic metre of air; the Becquerel is the Standard International Unit for activity of radon per unit air mass (NRC 1991).

two USA studies were restricted to females, who have historically spent a greater proportion of their time in the home than men; one third of the Winnipeg cases were female; controls were frequency matched by gender to cases.

By further restricting its target population to current non-smokers (70% of cases reported never having smoked cigarettes), the Missouri study allowed for estimates of radon risk relatively free of the strong impact of smoking. The majority of cases in the other two sites were smokers (86% in New Jersey; 97% in Winnipeg).

Among the base populations (reflected by the controls), radon exposure, expressed as mean ambient concentration in Becquerels per cubic metre (the standard international unit for radon activity concentration), was highest in Winnipeg (141 Bq/m³), where it was more than double that in Missouri (63 Bq/m³), whose radon levels were in turn more than double those in New Jersey (25 Bq/m³). Taking into account radon levels and numbers of cases, the most precise estimates of risk will be provided by Winnipeg, followed by Missouri and New Jersey.

3.3.2 SELECTION OF SUBJECTS FOR EACH STUDY

3.3.2.1 New Jersey USA

The subjects in the New Jersey investigation by Schoenberg et al. (1990, 1992) were a subset of those selected originally for a study of smoking and other determinants of lung cancer, which took place between 1982 and 1984 (Schoenberg et al., 1989). The generic eligibility criteria for this 1989 study were female gender and current residence in New Jersey.

Cases included all female residents of New Jersey newly diagnosed with lung cancer from August 1982 through September 1983. Case ascertainment was extensive and involved a

rapid reporting system established with hospital pathology departments, and periodic review of hospital pathology records and of state cancer registry and death certificate files. Population-based controls were selected at random from the New Jersey population through driver's licence (for live cases under 65 years of age), Medicare (for live cases over 65 years of age), and death certificate (for deceased controls matched to deceased cases) files. Controls were matched to cases by vital status at the time of ascertainment (live versus deceased), age (within 5 year age strata), and 'race'. Deceased cases were further matched by closest date of death, to deceased controls.

Subjects were considered eligible for the radon sub-study if they had lived in one or more 'index residences' for a total of 9 or more years during the period 5 - 30 years prior to diagnosis (cases) or selection (controls) (Schoenberg et al. 1992)⁴. 'Index residences' had to satisfy two criteria: 1) Each had to have been a subject's residence for a minimum of 7 years; and, 2) Of the period 5 - 30 years prior to ascertainment, 14 or more years had to have been spent in one to three houses which could be measured.

Details of the inclusion and exclusion of potential subjects from those identified as eligible are provided in Figure 3.1. The radon sub-study subjects were 'fairly representative of the women in the original study except with respect to age, race, and education' (Schoenberg et al., 1990). Compared with the original lung cancer study, the radon sub-study subjects were older, more likely to be of 'white race', and more educated than those original subjects for whom radon measurements were not ultimately available. Since all of these factors were

⁴ The eligibility criterion noted here is a more relaxed criterion than that in the original publication (Schoenberg et al. 1990) reviewed in Chapter 2, which required that subjects lived in a single 'index residence' for at least 10 years during the period 10 to 30 years prior to diagnosis or selection. The former criterion assumed a minimum 10 year latency between relevant radon exposure and lung cancer diagnosis, based on Whittemore and McMillan (1983), an assumption called into question by data (Howe et al. 1986; NRC 1988) indicating a shorter (5 year) latency period. The later criterion not only added more subjects, but added more measured houses for some subjects, as the initial budget limited measurement to one house per subject. The expanded radon study involved measures in up to 3 homes per subject.

included as control variables in multivariate regression models, selection bias arising from these measured factors was not an important concern. To the extent that such factors differed from those in the other radon studies, however, comparability of populations (generalizability) may be important *if* these factors modify radon-attributable lung cancer risk.

According to Schoenberg et al., a subject was ultimately included in their 1992 analysis only if there were radon measurements or estimates covering at least 9 years of the 25 year target exposure period. While constructing year by year exposure estimates, however, I found that the 922 subjects reportedly meeting the above criteria in fact included 263 subjects (129 cases and 134 controls) for whom there were apparently no measured radon data. These subjects were not included in the main analysis, but the availability of personal covariates on these subjects enabled a comparison among those interviewed of those included to those excluded due to lack of home monitoring data.

Figure 3.1
Subject Inclusion/Exclusion in New Jersey

Cases		Controls	
1306	Incident cases were identified from August 1982 through September 1983.	1449	Eligible non-smoking controls identified
↓	312 Exclusions for the following reasons: - no contact attempted because of physician refusal or inadequate waiting period (81); - subject untraceable (14); - no respondent available for interview (57); - refusal (160).	↓	454 Exclusions for the following reasons: - subject untraceable (45); - no respondent available for interview (46); - refusal (363).
994	Completed the personal interview and constituted the cases included in the 1989 study of lung cancer determinants	995	Completed the personal interview and constituted the controls included in the 1989 study of lung cancer determinants
↓	514 excluded due to: - Refusal or loss to follow-up of respondent or inadequate address-specific information (168); - Subjects did not meet 'index residence' inclusion criteria (165); - Inaccessibility of previous residences, ie., destroyed or refusal by or inability to contact current resident(s) (181).	↓	553 excluded due to: - Refusal or loss to follow-up of respondent or inadequate address-specific information (152); - Subjects did not meet 'index residence' inclusion criteria (176); - Inaccessibility of previous residences, ie., destroyed or refusal by or inability to contact current resident(s) (255).
480	Included in 1992 analysis and retained in data files.	442	Included in 1992 analysis and retained in data files.
↓	- 129 exclusions by VS Catalan due to lack of monitoring data in any home.	↓	- 134 exclusions by VS Catalan due to lack of monitoring data in any home.
351	Cases included in the present analysis.	308	Controls included in the present analysis.

3.3.2.2 Winnipeg Canada

Winnipeg was selected by L  tourneau et al., (1994) as the site of this investigation because it had the highest levels of indoor radon among the 18 Canadian cities monitored by McGregor et al. (1980). An inclusion criterion for all subjects was having a current Winnipeg residence. Cases were subject to the further criterion that prior written consent had to have been provided by the attending physician; population controls were only eligible if their home was equipped with a telephone (99.5% of Winnipeg residences) which was listed in the Winnipeg telephone directory (7% of phone numbers were unlisted). The restriction of the comparison (base) population to those with listed telephone numbers eliminated the most poor (and possibly also the most wealthy) members of society. Concern about potential selection bias on the basis of this non-comparable inclusion criterion is, however, minimized by the ascertainment and control of socio-economic status.

Incident cases of primary lung cancer occurring among residents of Winnipeg between 1983 and 1990 were identified every three weeks from the Manitoba Cancer Registry. Potential controls were selected at random from the Winnipeg telephone directory (estimated to include 92.5% of the Winnipeg population) and subsequently matched to the cases by age (+/- five years) and gender.

Details of the inclusion and exclusion of potential subjects from those identified as eligible are provided in Figure 3.2. Thirteen subjects (9 cases and 4 controls) included in the original study by L  tourneau et al. (1994) were excluded by the present author because they lacked information on smoking status. L  tourneau et al. (1994) assumed in their analyses that the latter individuals were smokers, and confined their adjustment for smoking to ever/never smoking status. In addition, I excluded 38 subjects (28 cases and 10 controls) found to have no home radon monitoring data. Covariates for these latter excluded subjects were compared to those for the remaining subjects.

Figure 3.2

Subject Inclusion/Exclusion in Winnipeg

Cases		Controls	
14091	Potential incident cases were identified between 1983 and 1990 [1400 of these ultimately met eligibility criteria - see below]	?	Eligible controls identified by random digit dialling
↓	- 650 Exclusions due to illness of patient and refusal of proxy; lack of permission by attending physician, or inability to trace patient.	↓	- Unknown number of non-respondents.
759	Completed the telephone interview	759	Completed the full telephone interview.
- 6 case-control pairs excluded because the cases did not in fact have primary lung cancer - 3 case-control pairs excluded because the cases had mesothelioma considered unrelated to radon - 12 case-control pairs excluded because controls were not properly matched to cases by age or gender			
738	Eligible for year-long radon measurement in Winnipeg residences occupied during the previous 5 - 30 years and included in the original (1994) analysis	738	Eligible for year-long radon measurement in Winnipeg residences occupied during the previous 5 - 30 years and included in the original (1994) analysis
↓	37 Excluded by VS Catalan due to: - lack of information on smoking status (9); - lack of home radon monitoring data (28).	↓	24 Excluded by VS Catalan due to: - lack of information on smoking status (4); - lack of home radon monitoring data (10).
701	Cases included in the present analysis	724	Controls included in the present analysis

3.3.2.3 Missouri USA

Alavanja et al. (1994) conducted this investigation among non-smoking women. Potential subjects were considered eligible if they were female, aged between 30 and 84 years at the time of diagnosis (cases) or interview (controls), of 'white' race, and, non-smokers (either lifetime non-smokers or former smokers who had ceased using all tobacco products 15 or more years prior to the interview).

Incident cases of primary lung cancer were identified from the Missouri Cancer Registry (estimated to capture 95% of cancer cases in Missouri) between June 1, 1986 and June 1991. A population-based sample of controls was selected from two sources, depending on the age of the cases, to whom the controls were ultimately frequency matched in 5 year age strata. For women between 30 and 64 years of age, names and addresses were selected randomly from the driver's licence files of the Missouri Department of Revenue. Although the proportion of the general population possessing a driver's license is not known, socio-economic status, the only known correlate of both driver licencing and lung cancer, was controlled in the analysis. For women between the ages of 65 and 84 years, names and addresses were randomly selected from lists of Missouri women provided by the Health Care Financing Administration, which included an estimated 95% of women of this age.

Details of the inclusion and exclusion of potential subjects from those identified as eligible are provided in Figure 3.3 on the following page.

Although this study was similar to the former two in seeking proxy respondents for cases who were deceased or too ill to be interviewed, it differed from the two former studies in selecting only live (directly responding) controls. The proportions of each respondent type (direct versus proxy) by case/control status and study site are provided in Table 2 of Appendix A.

Figure 3.3

Subject Inclusion/Exclusion in Missouri

Cases		Controls	
?	Incident cases were identified between June 1, 1986 and June 1991	1587	Eligible non-smoking controls identified
↓		↓	- 50 non-respondents to initial screening interview
650	Eligible cases verified through a screening interview	1527	Responded to the initial screening interview
↓	- 32 non-respondents	↓	- 125 non-respondents
618	Completed the telephone interview	1402	Completed the full telephone interview
↓	80 excluded due to: - Refusal of the respondent (10) or the next-of-kin or land-lord (11) to facilitate radon measurements; - Inaccessibility of previous residences, ie., location out of state or destroyed or inability to locate (59).	↓	219 excluded due to: - Refusal of the respondent (21) or the next-of-kin or land-lord (21) to facilitate radon measurements; - Inaccessibility of previous residences, ie., location out of state or destroyed or inability to locate (177).
538	Cases had year-long radon measurement in at least one Missouri residence occupied during the 5 - 30 years prior to ascertainment.	1183	Controls had year-long radon measurement in at least one Missouri residence occupied during the 5 - 30 years prior to ascertainment.

3.3.2.4 Summary of Subject Inclusion

Table 3.3 summarizes the above information regarding study subject inclusion. Subject recruitment took place between 1982 and 1991 inclusive. Ascertainment (defined as the date of diagnosis for cases, and interview for controls) began 1 - 2 years later for controls than cases, but finished in the same year in all three sites. All studies matched subjects by age (5 year strata or +/- 5 year tolerance); New Jersey additionally matched on 'race', while Missouri restricted on this factor⁵. The two USA studies restricted to females, while Winnipeg matched on gender. All three studies made use of cancer registries to identify incident cases; New Jersey supplemented this information with a rapid reporting system established with hospital pathology departments, and periodic review of hospital pathology records and death certificate files. All three studies made an attempt to confirm the diagnosis of primary lung cancer; only 76% of the cases were histologically confirmed in the non-smoker-restricted Missouri study. Inclusion of non-primary lung cancer cases in the latter study would be expected to result in a bias toward the null.

All three studies were population-based. Both of the USA studies exploited health care files to recruit controls 65 and over, and driver licencing files to recruit controls under 65. Winnipeg made use of the telephone directory which includes an estimated 92.5% of the population. Socio-economic status may therefore be an important control variable; education was ascertained in each study, and income from two of the three.

Missouri subjects had the highest response rates (83% among cases, 78% among controls). The Winnipeg study had a response rate of 50 % among eligible cases; 93% (650 of 699) of the exclusions (among those actually eligible) were related to refusal by the physician to consent to patient inclusion, inability to obtain a proxy for a patient too ill to respond, or

⁵ Ethnicity was addressed in Winnipeg through the collection of information on country of origin.

Table 3.3

Population Selection

Location / Author	Dates & Source(s) of Subject Ascertainment		Matching	Method of Lung Cancer Case Confirmation	Number And Proportion Of Eligible Subjects Retained In Original And Present Analyses	
	Cases	Controls			Cases	Controls
New Jersey USA, Phase I: Schoenberg et al. 1990, 1992	1982 - 1983 New Jersey Cancer Registry & supplemental information	1982 - 1984 1. Live Cases: Driver licence (<65 yrs); Medicare files (65 +) 2. Deceased cases: Death Certificates	Vital status (P); 1. Live Cases: age & race (F) 2. Deceased Cases: age, race, closest date of death (P)	Pathology reports were reviewed by Health Dept. physicians (no slide review)	<u>1990</u> : 433 ♀ = 31% of the 1306 who met original (1989) criteria [△] <u>1992</u> : 480 ♀ = 37% of 1306 <u>Present analysis</u> : 351 ♀ = 27% of 1306	<u>1990</u> : 402 ♀ = 27% of the 1449 who met original (1989) criteria [△] <u>1992</u> : 442 ♀ = 31% of 1449 <u>Present analysis</u> : 308 ♀ = 21% of 1306
Winnipeg Canada; Létourneau et al. 1994	1983 - 1990 Manitoba Registry	1985 - 1990 Winnipeg telephone directory	Age (P) Gender (P)	All cases histologically Confirmed	<u>1994</u> : 738 (488♂ 250♀) <u>Present analysis</u> : 701 (237♂ 464♀) = 50% of the 1400 eligible	<u>1994</u> : 738 (488♂ 250♀) <u>Present analysis</u> : 724 (246♂ 478♀) (total eligible unknown)
Missouri USA Phase I: Alavanja et al. 1994	1986 - 1991 Missouri Cancer Registry	1988 - 1991 Driver Licence (30 - 64 yrs) Medicare Files (65 - 84 yrs)	Age (F)	76% of cases histologically confirmed	<u>1994 and Present Analyses</u> : 538 ♀ = 83% of 615 identified	<u>1994 and Present Analyses</u> : 1183 ♀ = 78% of the 1587 eligible

DX = Diagnosis;

P = Pairwise matching;

F = Frequency Matching;

△ = Residence inclusion criteria restricted eligibility to radon study.

inability to trace the patient. The total number of eligible controls in Winnipeg was unrecorded during 'random digit dialling' (Létourneau et al. 1995a). In this, the only study to employ pair-wise matching, reciprocal exclusion was employed - ie., controls of excluded cases were also excluded by the original authors; the original analysis presented two subsets of subjects according to completeness of radon data. In the present analysis, a very few subjects (28 cases and 10 controls) for whom no home monitoring data was actually available for radon estimation, were excluded.

Although 76% of eligible cases, and 69% of eligible controls completed the personal interview, radon data was available for only 27% of eligible cases and 21% of eligible controls in the New Jersey study. Unavailability of home radon monitoring data was responsible for the exclusion of 67% of cases and 60% of controls post-interview. The main reasons for this included inaccessibility for measurement of previously inhabited residences and failure of traceable homes to meet residential inclusion criteria. Inadequacy of subject-provided residential history information required to locate previous homes for measurement, or subject refusal or loss to follow-up were responsible for the remainder.

Clearly, logistic rather than motivational factors were prominent in determining final participation rates. Although such low response rates in an observational study would typically give rise to concern for potential selection bias, under- or over-representation of cases or controls with high or low radon exposures is unlikely if radon varies in an unpredictable way from home to home⁶, as is widely assumed (e.g. Létourneau et al. 1995a). The recently raised suspicion that socio-economic status may be related to radon levels

⁶ The concentration of radon gas in indoor air is a function of its rate of entry into the building and the effective ventilation rate removing radon from the building (WHO 1985). The rate of entry can vary by at least three orders of magnitude, depending on the radon content of the soil gas, the permeability of the soil and building foundation, and the pressure difference between the building atmosphere and the soil gas. Radon concentrations of even adjacent buildings may differ considerably, and reliable estimates are not possible without measurement.

(Archer 1997)⁷ was not an important concern since, as mentioned previously, correlates of this factor were measured. Moreover, response rates were only slightly lower amongst controls than cases in all sites, minimizing concern for selection by un-measured factors. Available interview data on excluded subjects was, however, compared to that among those retained to investigate possible differences in attrition which could affect the generalizability (external validity) of the results.

3.4 RADON EXPOSURE INFORMATION

Table 3.4 below summarizes the study characteristics relating to radon exposure assessment. All three studies targeted the exposure time window (ETW) 5 to 30 years prior to ascertainment (date of diagnosis for cases, and interview for controls), with the exception that Missouri initially used date of interview as the time zero for both cases and controls. The present analysis uses ascertainment date to define a standard time axis for all three studies.

Radon ascertainment began with the completion of a home history during the subject interviews. Homes occupied during the subject's lifetime were identified by a sequence number, address, and start and end date of residence. Homes fulfilling study-specific residence criteria (column 2 of Table 3.4) were then targeted for in-home radon monitoring. If no residences met the criteria, the subject was dropped from the study. The relatively stringent residence inclusion criteria in New Jersey accounted for its low overall response rate (calculated as proportion of 'eligible' subjects included in the final analysis). Whereas a maximum of 3 residences were monitored in New Jersey, all in-state (Missouri) or within-city (Winnipeg) residences that had been occupied by the subject for a minimum of 1 year

⁷ Lower socio-economic status (SES) groups tend to live in smaller and older houses than do higher SES groups. Small, older houses tend to be less 'tight' (ie., have higher ventilation rates) than larger newer houses that have higher radon levels (Papastefanou et al. 1994; Cohen et al. 1989, 1991).

Table 3.4

Radon Exposure Information

Location / Author	Originally Targeted Exposure Time Window (ETW)	Residential Inclusion Criteria	Placement of Dosimeter	Apartment Measurement	Method and Duration of Radon Measurement[△]
New Jersey USA Phase I: Schoenberg et al. 1990	5 - 30 years prior to ascertainment	Minimum 9 years coverage by up to 3 houses during ETW [□]	Living Area (76%); Basement (5%); 4 Day Charcoal Canister (8%)	For 22 cases and 17 controls who lived in apartments above 2 nd floor assigned 14.8 Bq/m ³	1 year ATD τ
Winnipeg Canada: Létourneau et al. 1994	5 - 30 and 5 - 15 years prior to ascertainment	All Winnipeg (city) residences of ≥ 1 year during ETW	Bedroom and Basement (reported separately)	Measured (bedroom only)	2 consecutive 6 month CR-39 ATD
Missouri USA Phase I: Alavanja et al. 1994	5 - 30 years prior to interview	All Missouri (state) residences of ≥ 1 year during ETW	Bedroom and kitchen	Measured	1 year ATD τ

τ Terradex Corp. supplied and read Alpha Track Detectors (ATD).

$\triangle$ Refers to method of measurement in each subject residence.

$\square$ See section 3.3.3.1 for further details.

during the ETW were targeted in the other two studies. All of the studies differentiated between basements and living area (including bedrooms and kitchens) when placing dosimeters for radon measurement. By using dosimeters in both bedrooms and basements, the authors of the Winnipeg study were able to derive a regression equation for predicting one from the other; explained variance was 95%⁸. With the exception of 22 cases and 17 controls in New Jersey who lived in apartments above 2nd floor (who were assigned an assumed value of 14.8 Bq/m³), measurements appear to have been made in both apartments and single family dwellings alike. Alpha track measures were available for a full year on each monitored home with the exception of 8% of New Jersey measures obtained using 4 day charcoal canisters⁹. Terradex Corp. supplied and read the dosimeters for the two American studies; Winnipeg used CR-39¹⁰ detectors supplied by the Manitoba Cancer Treatment and Research Foundation. Integrated measurements yielding an annual average radon concentration were available for at least one home for all subjects included in each study site (as defined in the bottom lines of Figures 3.1 to 3.3).

3.5 DATA PREPARATION

This section briefly describes the procedures used to verify or 'clean' the primary data,

⁸ Living area radon levels were on average two thirds those in the basement within a home.

⁹ Although no placement information was recorded for the canister measures, it seems reasonable to assume these were placed in living areas as opposed to basements. Charcoal canisters and alpha track detectors are methods of integrated sampling used to measure the total number of radiation events resulting from the decay of radon and its progeny over a period of time (Krewski et al. 1989). Because of the shorter duration of canister measures, they are less reliable.

¹⁰ CR-39 is an abbreviation for Columbia Resin 39, and refers to the chemical diethylene glycol bis (allyl carbonate) (Bartlett 1985). CR-39 is a type of plastic film on which alpha particles etch tracks which are later counted. Although it is unknown what type of plastic is used in Terradex dosimeters, it is presumed to be of no net importance. According to Yeager (1991), a background correction (representing surface defects from manufacturing) is made and subtracted from observed track density in computing net track density which is then multiplied by the responsiveness of the plastic material to determine exposure.

derive a common electronic format for each type of information supplied, and impute values for missing data on radon and covariates; it concludes with a summary of the constellation of common variables available for analysis.

3.5.1 DATA VERIFICATION AND FORMATTING

The importance of this phase of the analysis can not be over-stated. It entailed a year of full time effort, ensuring the most reliable, internally consistent, and comparable (between studies) data-sets derivable from the originally collected data. Throughout the process, all procedures were meticulously documented. The use of a single (and highly commented) data management program file for each site ensured that coding changes were automatically maintained.

Only records for the subjects comprising the samples defined on the bottom lines of Figures 3.1 to 3.3 were retained. A frequency table of values was created for each variable; suspicious or extreme values were identified and investigated. Discrepancies in definitions were resolved. For each variable, common coding was developed which retained the highest order of information (ie, continuous, interval, ordinal, dichotomous). This coding scheme is described in Appendix A, and summarized in Figure 3.4.

Fifty-year arrays were created for the year-by-year covariables. For each of 50 years prior to ascertainment, 11 separately coded and indexed (1 - 50) variables (identified on the right hand column of Figure 3.4) were created.

Basement radon concentration was ascertained only in Winnipeg for the full sample of subjects. In addition, since the Winnipeg questionnaire did not ascertain the intensity of smoking by co-residents in each subject home, the aggregate intensity of smoking by co-residents for each home was estimated. This was done by calculating the average intensity

Figure 3.4

Common Data

Outcome: Lung Cancer Case/Control Status and Histologic Type									
Static Variables:				Year by Year Variables:					
	NJ	Wpg	Miss				NJ	Wpg	Miss
Age & Year at Ascertainment	✓	✓	✓	<i>For each of 50 years prior to ascertainment:</i>					
Gender	✓	✓	✓						
Source of Information (Self, Proxy)	✓	✓	✓	Home Sequence Identifier			✓	✓	✓
Ethnicity	✓	✓	✓						
Subject Smoking:	✓	✓	✓	Subject Smoking Intensity			✓	✓	✓
Status ¹	✓	✓	✓	Number of Other Smokers in Home			✓	✓	✓
Intensity	✓	✓	✓	Intensity of Smoking by Others in Home			✓		✓
Start Age	✓	✓	✓						
Stop Age	✓	✓	✓	Living Area Radon Concentration			✓	✓	✓
Duration ²	✓	✓	✓	Living Area Radon Estimation Method			✓	✓	✓
Years since Smoking Cessation	✓	✓	✓	Basement Radon Concentration				✓	
ETS Exposure	✓	✓	✓	Basement Radon Estimation Method				✓	
Education	✓	✓	✓						
Family Income		✓	✓	Proportion of Time Spent in Home					✓

1 Categories for smoking status were: Never-smoker, Cigarette-smoker, Pipe or Cigar Smoker, or Mixed (cigarette and pipe or cigar) smoker

2 Duration was adjusted for periods of cessation which were followed by relapse prior to stop age (final quit date) or ascertainment.

NJ New Jersey;

Wpg Winnipeg;

Miss Missouri

ETS Environmental Tobacco Smoke

of cigarette-smoking by co-residents for New Jersey and Missouri combined, and multiplying this value by the number of co-resident smokers specific to each ETW year.

3.5.1.1 Internal Consistency

The internal consistency of the site-specific data-sets was exhaustively examined. For each variable, one or more internal comparisons with other variables in the data-set were sought and performed. Examples of such attempts at cross validation in the personal covariates and radon exposure files (respectively) include: 1) verification that a subject's attained age was greater than or equal to the age at which the subject was reported to have quit smoking; 2) verification that within subjects, the home sequence identifier changed consistently with the home radon concentration recorded for a given calendar year. Where inconsistencies were identified, their origin was discerned and the problem corrected.

3.5.1.2 Replication of Original Descriptive Analyses

This step involved the restriction (in the case of New Jersey) or expansion (in the case of Winnipeg) of the data-set to the subject samples defined in the original publications (as defined in Table 3.4 above). In the case of New Jersey, the 1991 publication by Schoenberg et al. was selected for verification since it contained more descriptive statistics. The descriptive statistics provided in the original paper were then verified against those in the reconstituted data-set.

3.5.1.3 Examination of Descriptive Statistics and Creation of Common Categorical Variables

Analyses were performed on the final subject samples defined in section 3.3 (Figures 3.1 to 3.3). Following inspection of uni-variate statistics across site for each covariate, categories

were created with common cut-off points across site which would provide sufficient stratum-specific proportions (by case/control status) for both descriptive and regression analysis. The radon distributions among cases were analysed across site to define an appropriate common reference group and exposure categories. Bi-variate descriptive analyses involved the calculation of the 25th, 50th (median), and 75th percentile radon concentration within strata defined by the covariate; this was done separately by outcome (case/control status).

3.5.2 IMPUTATION OF MISSING DATA

As mentioned in the literature review, Weinberg et al., 1996 used simulation to assess six methods¹¹ for handling gaps in exposure history data, in the context of radon risk assessment:

Weinberg et al. made the following assumptions:

- 1) The relation between cumulative radon exposure and risk of lung cancer follows the Excess Relative Risk (ERR) model, as developed by the NRC (1988);
- 2) No other sources of measurement error; and,
- 3) Random (non-differential) missingness of exposure data with respect to: a) disease status b) length of occupancy, or, c) the true exposure level.

Under a credible non-null model (based on the BEIR IV report), only 'observed control mean imputation' (OCMI)¹² and 'observed mean imputation' (OMI) performed without bias, and

¹¹ The 6 methods reviewed by Weinberg et al. 1996: [1) Overall Mean Imputation; 2) Status-Specific Mean Imputation; 3) Time-Weighted Averaging; 4) Complete Subjects Only Analysis; 5) Zero Impute; and, 6) Observed Control Mean Imputation] are defined in Chapter 2.

¹² OCMI refers to the imputation (or replacement) of missing periods of exposure by the mean observed amongst controls.

with coverage of nominally 95% confidence intervals that was close to 95%. Weinberg et al. explained the OCMI¹³ results as arising from the Berkson error structure which the method imposes. If R is the true cumulative radon exposure, and R^* is the exposure assumed when missing concentrations are assigned the observed control mean, the mean of R conditional on any particular value of R^* is the assigned exposure, R^* . As a consequence (Armstrong 1990), if the rate ratio depends linearly on R , then the dose-response can be validly estimated using R^* instead of R .

In a linear model (or in an exponential model where the product of the regression coefficient and the exposures are relatively small, ie., at the lower, linear end of the curve), observations close to the mean exert little influence on the slope of the regression line, by contrast to those further away, which exert more leverage (Weinberg et al. 1996). Since OCMI brings cumulative exposures closer to the mean it reduces their influence on the slope. OCMI does not, therefore, introduce bias to the estimation process. In the words of Weinberg et al. , 'the estimate is not biased compared to the parameter that would be estimated with no missing data'. However, power (and precision) are 'inevitably reduced when exposure data are missing'.

OCMI was employed to impute missing radon concentrations; within each site, the mean value among control homes was imputed for historical periods for which respective home monitoring data was unavailable. The defensibility of the assumptions on which the validity of this method depends is discussed below:

¹³ The results of OMI were almost as good as OCMI because the mean over all houses, in this instance, closely approximates the mean over control houses.

- 1) In the exponential model¹⁴ [where $RR = \exp(\beta x)$ with an exposure variable designated x and β its regression coefficient relating exposure to rate ratio of disease], with normal Berkson error, there is no attenuation provided that the error variance does not depend on the approximated exposure (Armstrong 1990). This, condition, that RIR^* is normally distributed is a strong assumption and 'in general, (in non-normally distributed Berkson error settings) you will get some bias, though hopefully it will be small' (Weinberg, personal communication). Normality issues aside, however, since the exponential curve is linear at its low end, i.e., where the product βx is small, which is the case in with residential radon and lung cancer, bias arising from imputation is not expected. An exponential model was assumed by the original investigators of the eight case-control studies published to date; two of these investigations (Schoenberg et al. 1990, Pershagen et al. 1994) also included results based on an ERR model. The relative fit to the data to the two model forms was not explicitly compared, however. In the present study, sensitivity analyses¹⁵ were undertaken to evaluate the relative fit of the possibly attenuated logistic model to a linear ERR model free of attenuation arising from the method of imputation.
- 2) Dosimeter measurement error is known to exist: for a three month alpha track detector measurement the coefficient of variation is 8 - 25% (Scott and Robertson 1991), depending on (inversely related to) the radon concentration. This error is of the classical type, and so will attenuate estimates of effect.
- 3) a) The assumption of missingness that is non-differential between cases and controls will be examined during data analysis;

¹⁴ The exponential model combined with case-control sampling induces the logistic model; under the rare disease assumption or with incidence-density sampling, the odds ratio (OR) predicted by the logistic model estimates the rate ratio (RR) predicted under the exponential model.

¹⁵ Model sensitivity analyses are described in section 3.8.2.

- b) There is a relationship between missingness and length of occupancy, since measurement was only attempted if the subject had resided in a home for a minimum time period (7 years in New Jersey, 1 year in Winnipeg and Missouri). The impact of this violation is unknown.
- c) Whether missingness is un-related to the true exposure level can not be tested, but seems a reasonable assumption.

If missingness can be related in an informative way to true radon levels, Weinberg et al. (1996) suggested stratifying on predictors, (such as house-type, floor level, underlying geology, or the window-opening behaviour of occupants), then calculating the control mean within these strata and imputing missing values in a stratum-specific manner (provided that the assumptions under part 2 hold within strata). Under a hypothetical stratification system that explained about half the variability in the radon concentrations (on a logarithmic scale), a precision of prediction close to that of complete data was observed in simulations by Weinberg et al. As one of the predictors (house type) suggested by Weinberg et al. (1996) was available from the New Jersey data-set, its potential value for stratified OCMI of radon exposure was investigated.

OCMI (with or without stratification) was also used to impute missing values for covariates as follows:

In New Jersey:

- The missing value of education for 1 case was imputed with the observed control mean education level.
- For one control subject, missing values for age at start of smoking, age when stopped smoking, duration of smoking, years since cessation, and average number of cigarettes smoked per day, were replaced by the respective control mean values.

In Winnipeg:

- Eight missing values (7 cases, 1 control) for education category were imputed with the subject's income-stratum-specific control mean; 306 missing values (165 cases, and 141 controls) for income were imputed with the control mean for the subject's education-stratum-specific control mean; where both income and education were missing for a subject, the respective overall control mean was used.
- Missing or invalid values for the following variables were replaced with the observed control mean specific to the smoking-history category of the subject: missing values for 2 subjects for age started smoking; questionable values (<7) of age started smoking (for 17 subjects); an inconsistent value (5) of age when smoking stopped for one subject who reported starting smoking at 20; a missing value for one subject for average number of cigarettes smoked per day.

In Missouri:

- Fifty-two missing values (for 26 cases and 26 controls) of education category were imputed with the control mean education level specific to the subject's income stratum. Two-hundred four missing values of annual income at interview (for 83 cases and 121 controls) were imputed with the control mean income level specific to the subject's education stratum.
- Missing values were replaced by the respective control mean value for 21 subjects with missing age at start of smoking, 2 subjects (1 of whom also had missing age at start) with missing age when smoking stopped, and 16 subjects with missing average number of cigarettes smoked per day.

There was one exception to the use of observed control mean as the method of imputation. This occurred in the New Jersey and Winnipeg data-sets during the calculation of environmental tobacco smoke exposure. Missing values for the start date of a subject's cohabitation with an other smoking household member were replaced with the subject's birth

date for first degree relatives, and as the date at age 18 in the case of others. Missing values for end date of cohabitation were replaced with the date at age 18 for first degree relatives, and date of ascertainment for others. Since there was much greater variability between subjects than within with respect to the date of start and end of co-residence with a smoker, the former strategy seemed like a more valid approximation than would have been imposed by the use of data from all control subjects. Missing values for the intensity of smoking by smoking household members was, however, imputed by stratum-specific OCMI, ie., using the intensity of smoking specific to the type¹⁶ of smoking co-residents, among controls.

3.6 DEVELOPMENT OF A COMMON LOGISTIC MODEL

In this section I discuss the methods employed in developing a common logistic model to predict the rate ratio for lung cancer associated with radon exposure. I will address in order: 1) how I accounted for matching at the design stage; 3) my general analysis strategy; 4) the regression variables; and 5) potential effect modifiers.

3.6.1 MATCHING

One of the three studies (Winnipeg¹⁷) employed pair-wise matching of controls to cases; the results obtained by the original investigators using conditional and unconditional logistic regression were similar (Krewski, personal communication). The New Jersey and Missouri studies employed frequency matching within 5 year age groups. The New Jersey study also frequency matched by ethnicity, respondent type (direct versus proxy), and, for proxy respondents, closest date of death. Since gender, age, ethnicity and respondent type all

¹⁶ Types of smoking co-resident included: cigarette-only smokers, cigar-only smokers, pipe-only smokers, cigarette and cigar smokers, cigarette and pipe smokers, and cigar and pipe smokers.

¹⁷ In Winnipeg controls were individually matched to cases by age (+/- 5 years) and gender.

constitute large strata, and since the inclusion of stratification (matching) variables as independent variables in a multivariate unconditional logistic regression model gives the same results as a conditional regression model in this circumstance (Hosmer and Lemeshow 1989; Rothman 1986), a conditional analysis is not warranted.

Breslow and Day (1990) recommend that ‘matching should be accounted for in the analysis whenever it has been incorporated by design’, since the reduction in variance estimates from an *unadjusted* analysis is not expected to be great, and appropriate and flexible methods for adjustment exist. Following such guidance, and further extending this for the purpose of consistency of statistical treatment and related inferences across study sites, each variable by which subjects were matched in any study was included as a covariate (or a set of dichotomous indicator variables) in the regression models for every site for which data was available, or for which such adjustment was relevant.

Default covariates arising from matching at the design stage will thus include the following: age at ascertainment (all sites), gender (Winnipeg only), respondent type (New Jersey only), and ethnicity (New Jersey and Winnipeg).

3.6.2 ANALYSIS STRATEGY

The potential pitfalls in multi-variate data analysis are well documented in the general epidemiologic literature. Robins and Greenland (1986), in arguing against the use of automated covariate selection procedures, have warned that multi-variate analyses of large data sets can result in numerous combinations of different covariates producing the same model fit. One can readily imagine that the numbers of possible models, each with differing sets of covariates and associated coefficients, across three different data-sets, could quickly render post-hoc disentanglement a highly questionable endeavour. By contrast, limitation of models to relatively few variables based on *a priori* knowledge lends credibility and

interpretability to the results (Robins and Greenland 1986; Vandenbrouke 1987). A structured and systematic approach has therefore been taken to the parallel analysis of the three case-control studies.

Hosmer and Lemeshow (1989) have noted a recent move among quantitative epidemiologists to include all 'scientifically relevant' variables in the model, irrespective of their contribution to the model. The rationale is to provide as complete control of confounding as possible within the given data set, based on the fact that it is possible for individual variables not to exhibit strong confounding, but when taken collectively, confounding can be present in the data (Miettinen and Cook 1981). Although I decided to err on the side of inclusion of variables that are known or suspected *a priori* to be 'important' (as risk factors or effect modifiers), my main concern was *not*, however, with confounding, as the literature reviewed in Chapter 2 provided little evidence of any demographic or other measured variable - itself a true cause of lung cancer - being predictive of radon exposure. Rather, my concern was with a risk factor disproportionately larger than radon - namely tobacco smoke exposure¹⁸ - and its correlates, and with the comparability of estimates obtained across the three sites - hence the objective to derive a common constellation of covariates for control across site.

A seemingly little known fact is that the exclusion of strong risk factors from logistic regression models can result in bias *regardless of whether or not they potentially confounders*. This follows the findings of Gail et al. (1984), who demonstrated that the logistic is one of the few members of the exponential family of models for which the asymptotic biases resulting from omission of important covariates are non-zero¹⁹ even if these variables are effectively balanced between study groups by design .

¹⁸ The RR associated with active smoking was expected to exceed that of radon by at least two orders of magnitude.

¹⁹ The bias is to the null for exposure coefficients > 0 and away from the null for exposure coefficients < 0 .

As a general and conservative strategy, radon exposure was considered only after preliminary analyses with respect to matching factors, *a priori* risk factors, and other covariates which were considered potentially important; effect modifiers were considered thereafter. Each model-fitting exercise included the examination of correlations amongst variable coefficients, changes in model deviance, the Schwartz Criterion, the Akaike Information Criterion, and the Hosmer-Lemeshow lack of fit test (SAS Institute Inc. 1990).

3.6.3 TREATMENT OF COVARIATES

Although a thorough review of the literature is beyond the scope of this methodologic discussion, I will provide a brief justification for each of the potential regression covariates in order of their anticipated importance as follows: known risk factors for lung cancer (active and passive tobacco smoke exposure and socio-economic status), matching variables (age, gender, ethnicity, source of information), and optional control variables (other environmental exposures, and year of ascertainment). A discussion of radon exposure estimation follows in section 3.6.5. Section 3.6.6 provides a summary of first order candidates for regression. Potential modifiers (second order terms) of the radon and smoking effects are discussed in section 3.6.7. As a convention, I will retain this order for discussion of the results.

3.6.3.1 Active Smoking

Cigarette smoking was first associated with lung cancer in a case-control study published by Wynder and Graham in 1950; Wynder (1997) notes with historical interest that on reviewing their data in the spring of 1950, the chief statistician at the American Cancer Society (EC Hammond) did not at first believe the strength of the correlation between smoking and lung cancer, assuming instead that some sort of error, either of interview bias

or statistical analysis, had been committed²⁰. Later this same year, however, Doll and Hill (1950) published their first and similar study on the subject. Numerous retrospective and prospective studies have followed.

A review by Wu-Williams and Samet (1994), summarized crude odds ratio point estimates for ever-smoking which ranged from 7.3 to 12.8, and 2.1 to 9.4, among men and women, respectively. The adjusted population attributable risk for cigarette smoking among men was estimated in a recent Montreal-based study (Siemiatycki et al. 1995) to be 90%. Among US women over the age of 35, only an estimated 14% of lung cancer deaths are thought to be unrelated to active or passive exposure to cigarette smoke (EPA 1992).

Moolgavkar et al. (1993) inferred on the basis of their re-analysis of radon and smoking data from a cohort of miners and smoking data from Doll and Peto's cohort study of British doctors, that radon and active cigarette smoke exposure affect the same processes (first mutation rate and intermediate cell division kinetics or 'promotion'), in the multistage process of carcinogenesis. Based on observations that lung cancer risk increases with early age at initiation and decreases with the time since cessation, the EPA (1992) similarly presumed tobacco smoke to act both as a tumour initiator and promoter. The same 5-50 year exposure time window to be applied to radon (in section 3.6.5) will also, therefore, be applied to active and passive smoking.

In the present analysis, four alternative models for the control of smoking are fit to each of the three data-sets. The strategy was to identify a model which first explained the maximum possible variance in the data arising from smoking, the most important *a priori* predictor, (and the default covariates), *before* considering radon exposure. The most typical model for

²⁰

Hammond collaborated with D Horn and went on to make what Wynder (1997) considers 'outstanding' contributions to the elucidation of the smoking-lung cancer association (Hammond and Horn 1958a, 1958b).

the treatment of smoking in previous radon studies was examined first; in this model, intensity and duration are combined into the simple cumulative index of pack-years. Alternative models were based on the work of Doll and Peto (Doll 1971; Doll and Peto 1976; Doll and Peto 1978; Doll et al. 1980; Peto and Doll 1984) in modelling separately the contributions of intensity and duration of smoking to lung cancer risk. Goodness of fit of the four alternative approaches to the modelling of smoking was the main criterion for their evaluation.

3.6.3.1.1 Alternative Models Based on Intensity and Duration of Active Smoking

A. Models Based on The Simple Product of Intensity and Duration Of Smoking

The first type of model included categorical variables for the default covariates²¹, with intensity and duration of smoking combined into “pack-years”, the product of packs of cigarettes smoked per day, and duration smoked in years up to the period five years prior to ascertainment²². This first model was specified as follows:

$$\text{Log (Rate Ratio)} = \sum \alpha_i X_i + \sum \beta_k I_k \quad (3.1)$$

Where:

Log is the natural logarithm;

Rate Ratio is rate of lung cancer in subjects with covariate values of $x_1, x_2, \dots$, and a given smoking level (indicated by indicator variables) relative to the rate in those with zero values for $x_1, x_2, \dots$ and the reference level of smoking in pack-years;

²¹ To be summarized in section 3.6.4.11; radon was not a ‘default covariate’.

²² For example, 20 pack-years if one smoked 4 packs per day for 5 years, or 2 packs per day for 10 years, or 1 pack per day for 20 years.

X_i	is the i^{th} covariate ²³ (other than those in the active smoking subset), with
α_i	designating the associated regression parameter (coefficient);
$\sum$	denotes summation over the covariates or smoking indicator;
I_k	is an indicator of whether a subject was in the k^{th} category for pack-years of smoking (with all $I_k = 0$ if in the reference category), with
β_k	designating its respective regression parameter.

As reiterated by Thomas (1987), Doll and Peto (1978) demonstrated that it is not correct to combine intensity and duration of smoking into a cumulative exposure index (such as pack-years above). The following two sets of models (B & C) incorporate the advice of Thomas that preliminary analysis begin with intensity and duration considered jointly in a multivariate fashion.

B. Models With Separate Terms for Log Intensity and Log Duration Of Smoking

The second type of model was based on the cohort study of British doctors, which features the most extensive follow-up results ever reported for smoking. In the publication that most specifically considered the quantitative association of intensity and duration of smoking with lung cancer risk, Doll and Peto (1978) observed that lung cancer rates were related to the square of the cigarettes per day (plus the constant, six, to allow for background rates in nonsmokers), and to the 4.5th power of duration of smoking. Duration was calculated as attained age minus 22.5 (average age at start of smoking plus three for latency). Doll and Peto's (1978) model was as follows:

$$\text{Age-specific Rate} = k (\text{cigarettes/day} + 6)^2 \times (\text{age} - 22.5)^{4.5}, \quad (3.2)$$

²³ In preliminary models, categorical representation was used for all variables, with levels as those defined in Tables 4.1.1 to 4.1.3. Indicator variables were used for those with greater than two levels. For variables demonstrating linearity in logits between category midpoints, categorical variables were replaced by single continuous variables.

where the proportionality constant, k , was equal to 0.273×10^{-12} . Let the term (cigarettes/day + six) be denoted by the term intensity, and the term (age - 22.5) by the term duration* (in recognition of the difference between it and the more usual interpretation as smoking stop age minus smoking start age). The rate ratio for men whose values of intensity and duration (denoted by i and d^* , respectively) differ by one arbitrarily defined unit relative to other men (whose values are denoted $i-1$, and d^*-1) can be expressed as:

$$\text{Rate Ratio} = i^2 (d^*)^{4.5} / (i-1)^2 (d^*-1)^{4.5}. \quad (3.3)$$

Equation 3.3 implies that the log of the rate ratio comparing men with intensity at level i and duration at level d^* , with those with intensity at level $i-1$ and duration at level d^*-1 , can be expressed as:

$$\text{Log (Rate Ratio)} = 2 [\log(i) - \log(i-1)] + 4.5 [\log(d^*) - \log(d^*-1)] \quad (3.4)$$

Allowing the coefficients for log intensity and log duration to be fit to the data, the portion of the model dealing with smoking has been re-specified as follows:

$$\text{Log (Rate Ratio)} = \beta_1 [\Delta_1 \log(\text{intensity})] + \beta_2 [\Delta_2 \log(\text{duration}^*)] \quad (3.5)$$

Where:

- Rate Ratio is to the relative rate of lung cancer in subjects with log smoking intensity and log smoking duration* differing by one unit value designated by Δ ;
- Intensity is the average number of cigarettes smoked per day plus six;
- Duration* is calculated as when smoking stopped (or age at ascertainment if the subject was still a smoker at the time of ascertainment), minus four (for latency), minus age at start smoking among cigarette smokers; among never-smokers

'duration*' is calculated as age at ascertainment, minus four, minus the average value of age at start smoking among smokers; and, β_1 and β_2 are fitted parameters for the $\log(\text{intensity})$ and $\log(\text{duration}^*)$ terms, respectively; $\exp(\beta_1)$ and $\exp(\beta_2)$ represent the amount by which the rate is increased for every unit change in $\log(\text{intensity})$ and $\log(\text{duration}^*)$, respectively.

Since age at ascertainment was implicitly controlled in the calculation of duration*, it was removed from the set of default covariates in this model.

Doll and Peto's (1978) model (equation 3.2) fit well in the subgroup of their subjects who: 1) were less than 80 years of age; 2) smoked a maximum of 40 cigarettes per day; and, 3) were either lifelong non-smokers or 'regular' smokers²⁴ of cigarettes (pipe or cigar smokers were excluded) since early adulthood (ages 16-25). In the present analysis, model 3.3 has been fit both to the full set of study subjects (within each site) as well as to the subset of subjects (of both genders) *restricted* to the three above criteria, with the exception that the third restriction was not implemented in Winnipeg as this would have excluded 97% of subjects.

C. Models With Separate Terms for Intensity and Duration Of Smoking

Whereas the "B" models considered the log of smoking intensity and duration, the C models considered intensity and duration of smoking in their untransformed units. It was therefore no longer necessary to add a constant to smoking intensity. In addition, in place of Doll and Peto's term for smoking duration*, which involved the subject's attained age, duration was simply considered as the reported number of years the subject had smoked. The type C

²⁴ Regular smokers were defined as those who had not quit smoking more than once who and had not changed their intensity of smoking by more than 5 cigarettes per day.

model is given in equation 3.6.

$$\text{Log (Rate Ratio)} = \beta_1 [\Delta \text{ intensity}] + \beta_2 [\Delta \text{ duration}] \quad (3.6)$$

where the terms are as defined previously (for equation 3.5), except that for equation 3.6,

Duration was calculated as age at cessation (for ex-smokers) or age at ascertainment (for continuing smokers, minus four (for latency), minus age at start of smoking; among never-smokers, duration was set equal to zero²⁵.

D. Models With Terms for Log Intensity Of Smoking and Attained Age

An additional set of models was considered, which controlled for attained age in place of smoking duration. This followed the observation of Doll (1971) that intensity of smoking was linearly related to lung cancer risk within separate age strata.

$$\text{Log (Rate Ratio)} = \sum \alpha [\Delta \text{ age}] + \beta_2 [\Delta \log (\text{intensity})]. \quad (3.7)$$

3.6.3.1.2 *Recency of Active Smoking*

As indicated earlier, all of the alternative formulations of smoking intensity and duration (models of type A through D) above included a four year lag interval prior to ascertainment; that is, only intensity and duration of smoking in the fifth through fiftieth years prior to ascertainment were considered. I used a separate model term to adjust for recency of smoking, based on repeated observations (in both retrospective and prospective studies) of a decline in lung cancer rate ratios with increasing time since smoking cessation (DHHS

²⁵ Breslow and Day (1987) described the choice between the variable I have termed duration* and duration as somewhat arbitrary.

1990).

In preliminary analyses, a dichotomous term discriminated between subjects who had smoked within the 15 years prior to ascertainment, and those subjects who had not. The latter condition effectively defined the smoking-specific inclusion criteria for the Missouri study. In later analyses, an ordinal term²⁶ applicable to all three studies was derived by subdividing the period 14 or more years prior to ascertainment into sub-categories similar to those employed by Alavanja et al. (1994) in their descriptive analyses.

3.6.3.1.3 Age at Start of Smoking

Earlier age at start of smoking increases the relative rate of lung cancer (Hegmann et al., 1993; EPA 1992). Risk model types "A" through "D" were therefore fit with and without additional terms for age at start of smoking.

3.6.3.2 Environmental Tobacco Smoke

Environmental tobacco smoke (ETS) exposure was included as a default covariate, and retained regardless of the precision of its estimated parameter, the weight of evidence having been sufficient for the US Environmental Protection Agency (1992) to classify ETS as a known human lung carcinogen. Most of the studies evaluated in the EPA risk assessment ascertained passive exposure to cigarette (as opposed to pipe or cigar) smoke, as was also the case in the included radon studies.

²⁶ The ordinal term for recency of smoking had 5 levels: 0 - never-smokers; 1 - current smokers (at ascertainment); 2 - ex-smokers who quit 14 - 22 years prior to ascertainment; 3 - ex-smokers who quit 23 - 35 years prior to ascertainment; 3 - ex-smokers who quit more than 35 years prior to ascertainment

3.6.3.3 Socio-Economic Status

After adjustment for age, gender, and calendar period, social class was observed to be an independent risk factor for cancer mortality in a 1995 'combined analysis' of nuclear industry workers (Cardis 1995). Bourbonnais and Siemiatycki (1996) assessed the relationship between thirteen types of cancer and four indicators of social class (family income, years of education, an occupational history index, and a composite index of the three). Covariates for all cancer sites included: smoking of cigarettes, pipes and cigars, drinking of beer, wine, 'hard liquor', tea, coffee, and soft drinks, eating of foods containing beta carotene, and an occupational history 'dirtiness' index. In the lung cancer analyses, many recognized occupational risk factors specific to this site were also included. The adjusted lung cancer odds ratios [and 95% confidence intervals] for the four social class indicators (lowest tertile compared to highest tertile) were 1.4 [0.9, 2.1], 1.9 [1.3, 2.9], 2.8 [1.9, 4.2], and 2.5 [1.6, 3.8], for education, occupation index, income, and a composite index, respectively. This finding of a stronger odds ratio for income and the composite index than for education or occupation index was a general finding across the 13 cancer sites.

Education level, available in all three study sites, was included as a default variable. The importance of income (available only in Winnipeg and Missouri) over and above education was investigated in each type of model.

3.6.3.4 Age

All three studies matched controls to cases on the basis of age at ascertainment. Age was therefore considered as a model covariate, either implicitly, as a component of the smoking duration* variable, or explicitly, as a separate variable.

3.6.3.5 Gender

As implied in section 2.5.2.5, gender is a possible modifier of the effect of radon. Two of the studies - New Jersey and Missouri - were restricted to the female gender. Since the Winnipeg study was matched on gender, it was considered as a default covariate in this site.

3.6.3.6 Ethnicity

The variable 'race' which distinguished 'white' non-hispanics, 'blacks', hispanics, aboriginals, Asian/Pacific islanders, and others, was a matching variable in the New Jersey study, and thus was included as a default covariate. The Missouri subjects were restricted to 'white' race. Another factor related to ethnicity, 'country of origin', was observed to significantly discriminate between cases and controls in the original Winnipeg study (Létourneau et al. 1994), but it was excluded from their final model because its inclusion did not significantly improve model fit (Krewski, personal correspondence). This factor will, however, be included in the analyses by default, based on my earlier-stated criteria.

3.6.3.7 Other Environmental Exposures

The evidence for an impact of occupational exposures on radon risk assessment was reviewed in section 2.5.2.6. Information on occupational exposures was collected in only two of the three included studies. The New Jersey study included a dichotomous indicator for 'high-risk occupation'. Létourneau et al.'s coding scheme involved six indicator variables: three of these represented degree of confidence in elevated risk associated with past exposure to a job category; the other three represented the degree of confidence in risk associated with past exposure to one or more of 44 specific agents. While the former study included high risk occupation but excluded education from its final model, the latter included education and excluded occupation. Collinearity between these terms is a probable explanation.

Considering the lack of comparability of the occupational variables and their probable correlation with socio-economic factors to be included by default, occupational exposures were excluded from consideration in the present analysis.

The distinction between urban versus rural residence was investigated by three published radon studies: Stockholm Sweden (Pershagen et al. 1992), Sweden (Pershagen et al. 1994), and Finland II (Auvinen et al. 1996). The rationale given in the Stockholm study for including a model term for 'county of residence' was to control for ambient air pollution, diet, and referral patterns for cases and hospital controls. A separate report on the Stockholm study by Lagarde (1992) noted that the relative risk was significant for radon exposure only in the most urbanized area, indicative of effect modification. A relationship to air pollution was presumed. Pershagen et al.'s 1994 later analysis on a wider Swedish study included a model term for urban versus rural living, but noted that none of their covariates had strong effects. The third study to collect data on urban/rural residence did not discuss this factor in their results. As information on this 'factor' was unavailable from the pilot sites it could not be further considered.

3.6.3.8 Source of Information

Since the New Jersey study employed frequency matching on this factor, using proxy controls when the case was deceased or too ill to be interviewed, source of information met my criteria for a default covariate. By virtue of restriction of controls in Missouri to those who were live and self-reporting, and could hence provide the most accurate information, adjustment for this variable was not relevant in this site. Although associated with lung cancer in the Winnipeg site, (65% of cases were able to respond directly to the interview, compared with 90% of controls), source of information was certainly not a risk factor (or correlate thereof); whether or not the subject was directly or indirectly interviewed was determined after the diagnosis of this usually fatal disease. As a 'spurious confounder'

(Greenland and Robins, 1985) quality of information arising from proxy responses should only be “adjusted for” if it has been (inadvisably) used as a matching variable. Although inclusion in the Winnipeg analysis of a term for source of information (as a main effect) made essentially no difference in odds ratio estimates for radon exposure, or any other covariate, this term was excluded from the final model in this site.

3.6.3.9 Year of Ascertainment

Breslow and Day (1980) suggested that a learning effect on interview technique may occur such that the quality of smoking information may increase over the duration of the study; if year of ascertainment varied between cases and controls, then so would the quality of smoking information. These authors suggested that if case acquisition lasted more than 1-2 years, year of ascertainment should be included as a stratification variable. For this reason, this factor was considered in preliminary models.

3.6.4 RADON EXPOSURE ESTIMATION

Radon exposure was included in regression models only after much model fitting with respect to the aforementioned factors. This section discusses the etiologically relevant time period, including lag period, earliest effective exposure time, and exposure time windows, and the method used to calculate radon exposure within exposure time windows.

In their combined analyses of 11 miner cohorts Lubin et al. (1994a) operationally defined a lag interval as a period in which exposure does not influence lung cancer incidence. This involves the definition of a minimum latency/detection time, ie., the minimum time between

the carcinogenic transformation of a cell and diagnosis of lung cancer²⁷. Radon exposure from underground mining is characterized by discrete individual exposure periods defined by start and end dates of employment. In the residential context, however, exposure is always non-zero, accumulated continuously for the lifetime of an individual. One must therefore define, in addition to a lag time which ends exposure assessment, an earliest effective exposure time, at which exposure ascertainment begins. In the residential setting, then, an etiologically relevant time period is sought such that its complement (ie., the time outside this interval) does not influence lung cancer incidence.

3.6.4.1 Lag Time

Although it has been noted (Thomas 1983) that the lag interval follows some probability distribution, it has been considered deterministic for practicality. Lubin et al. (1994a) analysed variation in model deviance as a function of the length of the lag interval, and observed that the maximum likelihood estimates for lag interval in the 11 cohorts were within zero to ten years of mortality; a five year interval was selected by these authors. Consistent with Lubin et al. (1994a) and the original analyses of the eight case-control studies of residential exposure published to date, this value of five years has been retained for the present analyses.

3.6.4.2 Earliest Effective Exposure Time

There is no theoretical basis on which to assume that there is indeed a period prior to the lag interval during which exposures are unrelated to lung cancer risk. There may, however, be quantitative differences in carcinogenic risk in children compared to adults beyond that of

²⁷

Note that mortality was the considered as the outcome in the miner cohort study, diagnosis is the outcome in the case-control studies. Since lung cancer is usually rapidly fatal, the difference in lag interval should be negligible.

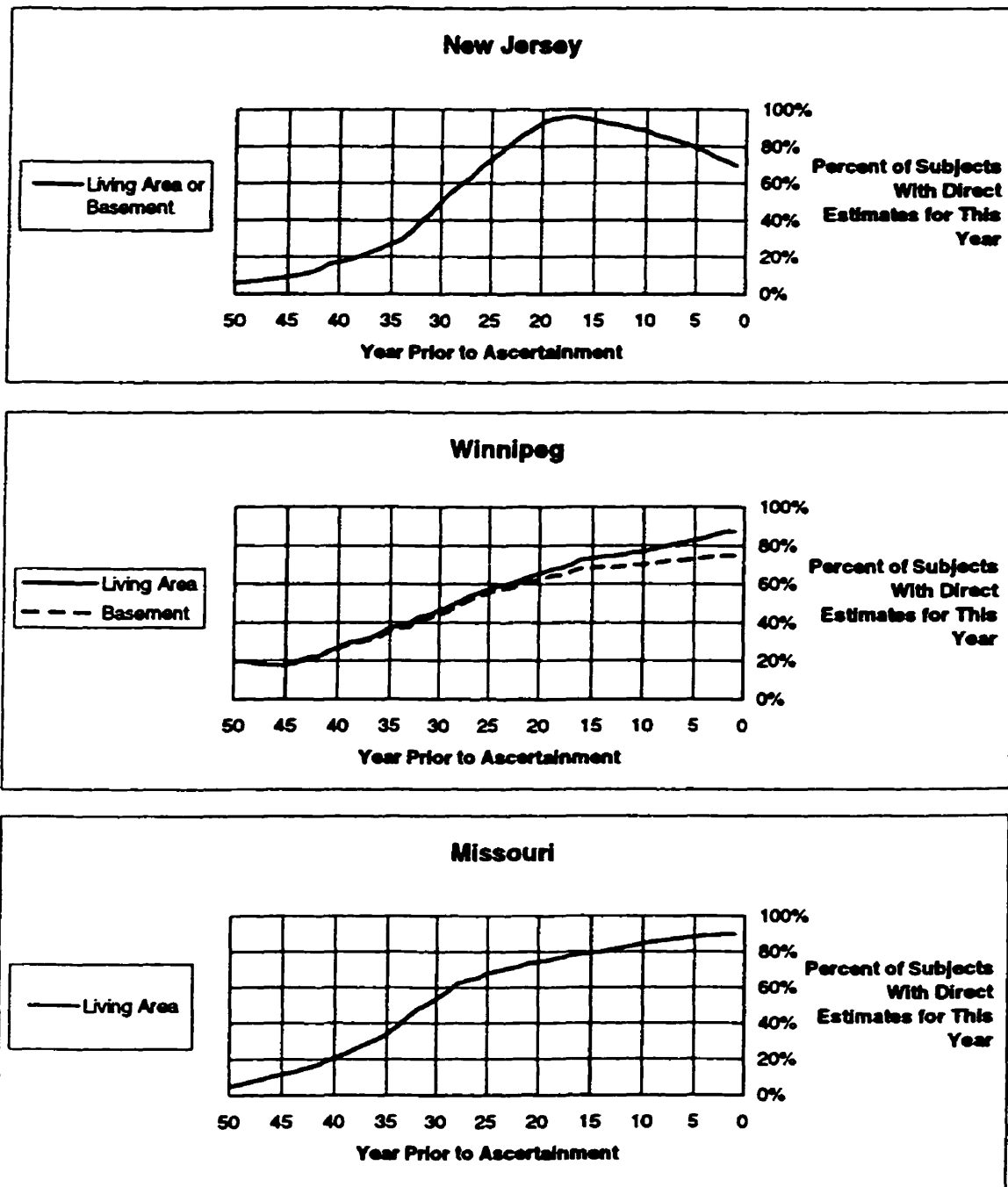
a longer exposure period; exposures early in life may in fact be of greater significance than those occurring later (Farland 1992). It seems prudent, therefore, to utilize all available data, ie., up to and including 50 years prior to ascertainment.

3.6.4.3 Exposure Time Windows

In their combined analyses of four miner cohorts, the BEIR IV committee (NRC 1988) considered three distinct time intervals for exposure: 5 - 10, 10 - 15, and ≥ 15 years prior to ascertainment, and estimated separate coefficients for the relative potency of cumulative exposures therein. Since no clear differences were observed between risks due to exposures in the first two intervals, these were combined into two exposure windows: 5 - 15 years and ≥ 15 years (NRC 1988).

Figure 3.5 illustrates the availability of measured radon exposure in the three pilot sites. The amount of available measures over time varies between sites, since it is largely governed by the residency-related subject inclusion criteria specific to each study (as seen in Table 3.2). The proportion of subjects for whom home radon estimates is available is highest between 16 and 21 years prior to ascertainment for New Jersey, whereas for Missouri and Winnipeg, exposure data is most complete at 1 year prior to ascertainment, declining monotonically thereafter. In general, however, over 60% of the home radon measures in the time period 5 - 24 years prior to ascertainment have direct estimates. This is compared with an average of only 34% in the time period or more years prior to ascertainment. Both for consistency with the treatment of the miner data, and because of the paucity of data in the period 25 -50 years prior to ascertainment, the further sub-division of the latter interval was considered impractical and uninformative. The 5 - 50 year interval for which (at least partial) exposure data was received from the primary investigators was therefore sub-divided into three non-overlapping time windows: 5 - 14 years, 15 - 25 years, and 25 - 50 years, analogous to the time windows defined in the most recent analysis of updated information on miner cohorts

Figure 3.5
Completeness of Radon Exposure Data



by the BEIR²⁸ VI committee (NRC 1998).

The BEIR VI committee estimated weighting factors for each exposure interval under two types of models (using either duration or concentration in addition to age and cumulative exposure), to account for the exposure-rate effect. The exposure rate or dose-protraction-enhancement effect refers to the modification by rate of exposure of the effect of cumulative exposure - the effect of accumulating a fixed cumulative exposure (integration of the concentration by time curve) is enhanced when this exposure is received over a longer duration (or analogously), at a lower dose rate (concentration). The weights derived by the BEIR VI committee from these two models were very close, as expected.

Since a fixed exposure period (of 5 - 50 years prior to ascertainment) is being assessed for each subject in each case-control study (most of whom were over 50 years of age), the average radon concentration (in Becquerels per cubic meter) over the period differs from the cumulative exposure (in working level months) only by a constant scaling factor, if a fixed equilibrium fraction and occupancy factor are assumed²⁹ (see equation 3.10 below). Protraction-enhancement effects can only be explored when cumulative exposure can be fixed, and either duration or concentration allowed to vary. In a case-control study with age-matched subjects and fixed retrospective time windows of exposure anchored not to age but to ascertainment year, only intensity of radon exposure *or* cumulative exposure are free to vary (each vary directly with the other). The two sets of BEIR coefficients, each of which resulted from considering one additional variable than that available for study herein, differed only marginally. As a default assumption, these two sets of coefficients were simply averaged for use in time window weighting. The weights were re-scaled to sum to one for

²⁸ BEIR is the acronym for the Biological Effects of Ionizing Radiation, a subcommittee of the National Research Council.

²⁹ In fact, gender-specific occupancy factors were developed in Chapter 4; within gender, however, there is a direct relationship between cumulative exposure in WLM's and average concentration.

the analysis of average concentration. Sensitivity of the results to differing sets of weights is addressed in section 3.8.1.

3.6.4.4 Calculation of Radon Exposure Within Time Windows

Let the subscript $t_1 - t_2$ denote the exposure time window for years t_1 to t_2 prior to ascertainment, and c_y denote the annual average radon concentration for the indexed year (subscript y). For the time window t_1 to t_2 of duration $Y = t_2 - t_1$, the mean concentration of radon exposure (in Bequerels/m³) is simply $\sum c_y / Y$. Let $WLM_{t_1-t_2}$ denote the cumulative exposure to radon decay products in working level months within this same time window. This cumulative exposure is calculated as follows:

$$WLM_{t_1-t_2} = \sum c_y \div 3700 \times 51.56 \times 0.5 \quad (3.8)$$

Integration of annual exposure concentrations	Conversion of Bq/m ³ to Working Levels	Conversion of years to Working Months (170 days) assuming 100% home occupancy	Assumed Fixed Equilibrium Factor

3.6.4.5 Home Occupancy

Residential radon exposure is representative of total radon exposure only to the extent that exposure to radon outside the home is small relative to home exposure, proportional to home exposures, or small and equivalent across subjects. Data on radon levels in workplaces or other indoor environments was unfortunately unavailable in any of the present investigations. Since outdoor concentrations are only a small (but variable³⁰) proportion of those indoors,

³⁰ Gorsek et al. (1990) measured indoor and outdoor concentrations in the Saleska Valley (Yugoslavia); average indoor concentrations by building material were 32, 53, and 75 Bq/m³ for concrete, ordinary brick, and fly ash bricks, respectively. The outdoor concentration varied

exposures from outside the home will be small when the amount time spent in other indoor locations is low. Table 3.5 displays Canadian survey data which were available at the time of the analyses, regarding the proportions of time spent in major locations.

Table 3.5
Canadian Home Occupancy Data[†]

Age Group	Sample Constitution: N (%)		Percentage of Time Spent in Major Locations			
	Female	Male	Home	Other Indoor [‡]	Outdoors	In Vehicle
Under 12 Years	155 (44.3)	195 (55.7)	71.0	17.0	8.5	3.5
12 - 18 Years [°]	91 (45.3)	110 (54.7)	67.8	20.5	8.6	3.2
Over 18 Years	925 (52.8)	825 (47.1)	64.4	24.0	5.5	6.0
Total	1130 (100.0)	1171 (100.0)				

† Based on data extracted from Leech et al. 1996.

‡ Included work/school, restaurants/bars, and other indoor environments (other than the subject's home).

° Some totals not equal to 100% due to rounding error.

Although varying with age, the large majority of time is spent in the home. In addition to variability with age, further variability can be expected between subjects by gender and calendar period (ie., cohort effects).

Although subject to inter- and intra-subject variability, however, some assumptions can be made about the limits of uncertainty with respect to the proportion of subjects' time which was spent in the residence. At the lower end of the extreme, persons employed outside the

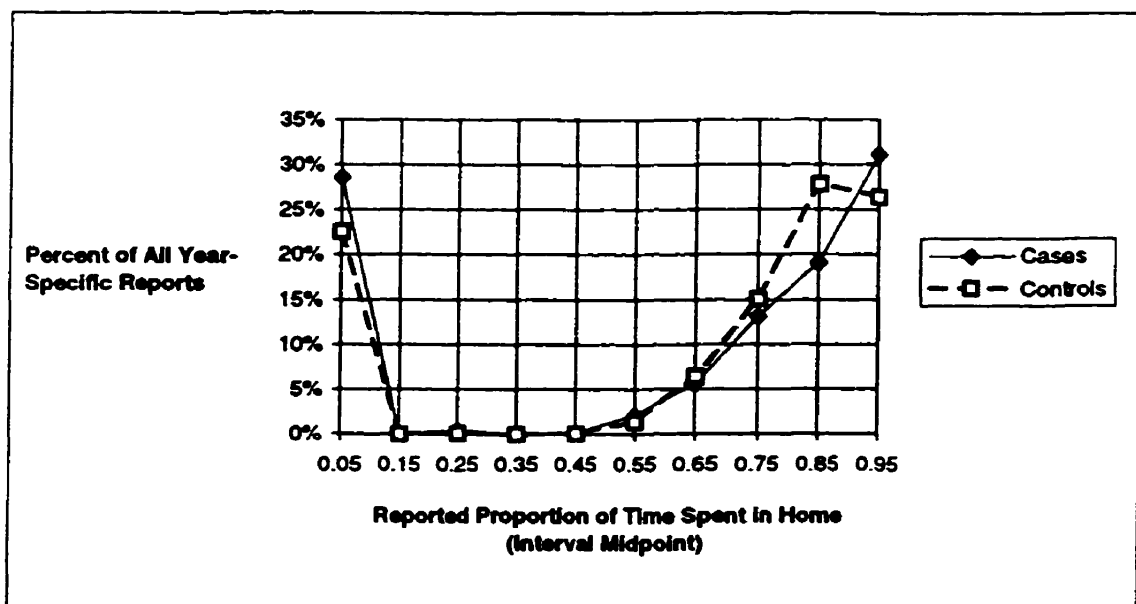
between 5 and 19 Bq/m³ depending on climatic conditions. Zhai and Feng (1993) observed outdoor radon concentrations in China there an average of 8% of the average indoor concentration. Indoor concentrations are related to building ventilation (Zhai and Feng 1993), which is strongly influenced by basement window opening (Letourneau et al. 1992b), and to climatic conditions (Letourneau et al. 1992b). Steck and Field (1998) observed a correlation of 0.6 between indoor radon and county median outdoor concentration in Minnesota.

home spend a reasonable minimum of 8 hours sleeping, one hour before retiring, and one hour on rising, totalling 10 of 24 hours, or 42% of each work day, in the home. Assuming further that this minimum occupant spent only one day out of two at home on the weekend, then a total of 84 ($= 10 \times 6 + 24$) of 168 hours, or 50% of the week would be spent at home. At the other end of the spectrum, women involved with primarily domestic responsibilities could reasonably spend up to 90 or 100% of their time in the home. Both the New Jersey and Missouri studies were restricted to women, assumed to have spent the majority of their targeted exposure time window (in the years 5 - 35 years prior their ascertainment between 1982 and 1991, ie., between 1947 and 1986) in their residences. In addition, the Missouri study subjects were specifically asked the average number of hours per day that they spent outside each home in their housing history. This information was used to calculate the proportion of time spent in the respective home in each of the 5 - 50 years prior to ascertainment.

Figure 3.6 provides a frequency (percent) distribution of these year-specific estimates of home occupancy among cases and controls. This data shows similar bimodal distributions by outcome. Although possibly reflecting a sub-population of frequent travellers, the lesser mode, at 10%, lends a degree of incredulity to this data; in fact the authors themselves chose not to use these estimates, preferring to make a fixed assumption of 84% occupancy by all residents in their calculation of cumulative exposure in working level months. Although reflecting some uncertainty, the data in Figure 3.6 do not suggest that variability in home occupancy is related to outcome status. If generalizable to the other two sites, this data suggests that assuming a fixed occupancy factor will not create bias, but rather simply reduce precision.

Figure 3.6

Estimated Home Occupancy In Missouri USA



Prior to analysing the above data, an assumption of 100% occupancy was made in calculating cumulative exposure in WLMs amongst all subjects. This factor is clearly an over-estimate which would limit the comparability of the WLM exposure units between those in this study and those amongst miners. After this analysis, however, a weighted average, using relative frequencies as weights, of the occupancy estimates between 0.33, (which corresponds to 8 hours per day) and 1.0 of 80% was calculated for cases and controls combined. This value was employed for the female respondents in all three sites (including 34% of Winnipeg subjects). Information available from the Canadian Human Activity Pattern Survey (Leech et al., 1996) (summarized in Table 3.5 above) was employed to inform an educated guess as to home occupancy among the Winnipeg males. The aforementioned survey interviewed 2381 persons between November 8th 1994 and August 30th 1995 in six Canadian cities. Unfortunately, gender-specific analyses were not conducted. These data do, however, point to obvious variability according to age; in addition, the historical representativity of these estimates is subject to uncertainty. The same uncertainty exists, however, with respect to the Missouri data collected between 1986 and 1991; given the other uncertainties associated with the use of a fixed estimate for a factor which varies by age, and calendar time, the two sets of data were considered adequately comparable. Assuming that the female-specific occupancy rate (proportion) among Canadian adults was 80%, the male-specific rate was estimated as $[(64.4 - 0.528(80))/0.471] = 47\%$. This value was less than the lowest reasonable extreme assumption of 50% estimated above; the latter rather conservative value (equal to the former rounded to one significant digit) was nonetheless assumed for the 66% of Winnipeg subjects who were of male gender.

Since average radon concentration does not take occupancy into account, but differs only from the cumulative measure by a constant which now includes a gender-specific occupancy factor, the impact of the above assumptions on the results will be readily deducible from differences in the results of analyses in these two radon measures, and then only in Winnipeg where both genders are represented. The WLM analysis can therefore be seen as a type of

analysis of sensitivity to occupancy assumptions.

3.6.4.6 Time Window Weighting

As a generality, let $w_{t_1-t_2}$ denote radon exposure (in WLM or Bq/m³, for cumulative or mean concentration, respectively) in the time window designated by $t_1 - t_2$. Define w^* as “effective exposure”, with θ_{5-14} , θ_{15-24} , and θ_{25-50} defining the relative contributions to the total exposure from the three time windows.

$$w^* = \theta_1 w_{5-14} + \theta_2 w_{15-24} + \theta_3 w_{25-50} \quad (3.9)$$

For the three separate time windows of 5 - 14, 15 - 24, and 25 - 50 years (inclusive) prior to ascertainment, the respective weighting coefficients are given in Table 3.6.

Table 3.6
Time Window Weighting for the Default 5 - 50 Year Radon Analysis

Window	Window Weight (θ)		Annual Weight Within Window	
	Average Radon Concentration	Cumulative Radon	Average Radon Concentration	Cumulative Radon
5 - 14 years	0.4494	1	0.0449	0.1
15 - 24 years	0.3371	0.75	0.0337	0.075
25 - 50 years	0.2135	0.475	0.0082	0.0183

Note that, concordant with the earlier observations by Lubin et al. (1994a) of the same miner cohorts, the BEIR VI committee's (NRC 1998) empirically derived weights reflect decreasing importance of more distant exposures. As a practical matter, such a weighting scheme accords lesser importance to the least certain radon estimates.

3.6.5 SUMMARY OF FIRST ORDER CANDIDATES FOR REGRESSION

Table 3.7 lists by site the covariates included in regression equations by default, and those included optionally, based on fit, collinearity with other variables, and other considerations, prior to the addition of a term (or terms) for radon exposure.

In all, over seventy distinct models were considered before final model selection. Intermediate models are specified in Appendix C. A first order term was retained if it was a significant predictor in at least one site, and/or regardless of statistical precision, if it was part of the set of *a priori* risk factors, so long as it was not importantly collinear with other model components. Concern for collinearity was raised by a correlation between regression coefficients greater than 0.40.

Table 3.7
Candidates for Regression

STAGE OF ANALYSIS	OCCURRENCE IN STUDY SITE		
	NEW JERSEY	WINNIPEG	MISSOURI
1. Default Covariates	Tobacco Smoke-Related Variables [◊]	Tobacco Smoke-Related Variables [◊]	Tobacco Smoke-Related Variables [◊] ⊠
	Education	Education	Education
	Age at Ascertainment*	Age at Ascertainment*	Age at Ascertainment*
	⊠	Gender*	⊠
	Ethnicity*	Ethnicity	⊠
	Respondent Type*	Respondent Type	⊠
2. Optional Covariates	--	Income	Income
	Year of Ascertainment	Year of Ascertainment	Year of Ascertainment
3. Main Exposure	Weighted Radon	Weighted Radon	Weighted Radon

- Matching variable in this study site.
 - ⊠ Restriction factor in this site. The New Jersey and Missouri studies were restricted to females. The Missouri study was further restricted to 'white' race, non-smokers (either lifetime never-smokers, or persons having quit 15 or more years prior to interview), and among controls, live respondents.
 - Indicates a variable for which information was not gathered or which was considered irrelevant in the primary study.
 - ◊ The complement of variables to represent active smoking was determined on the basis of preliminary analyses taking into account the potential for inter-correlation amongst the smoking-related variables, and the separate contributions of intensity and duration.
- ETS Environmental Tobacco Smoke Exposure.

The overall impact of the control variables was assessed in each site by comparing the coefficients (and standard errors) for radon as estimated in the final standard multivariate logistic model, with those estimated in a univariate logistic model with weighted radon as the single regression variable.

3.6.6 EFFECT MODIFICATION

The potential for effect modification was explored (either by stratification or the addition of interaction terms) only for those variables whose potential to modify the relationship between the exposure and outcome were suggested on the basis of *a priori* evidence. Since the power to detect a radon effect was itself expected to be moderate at best (Lubin et al. 1995a), and interaction tests have even lower power than those for first order (or 'main') effects, a requirement for a significant ($p < 0.05$) change in the estimate effect of radon seemed unreasonable, as it is the *relative* strength of interaction terms that is important (Pearce 1989). As a decision criterion, interaction terms which were significant at $p < 0.20$ in at least one site were retained in the common model. As a practical matter, interaction terms were centred to minimize collinearity between these terms and the first order terms involved in them.

3.6.6.1 Potential Modifiers of the Smoking Effect

For the purpose of trying to fit the observed patterns in risk as carefully (comprehensively) as possible, interaction terms with smoking were investigated for gender, age at ascertainment, and source of information. The gender interaction term was based on Peto and Doll's (1984) observation of excess rates of lung cancer in women which were approximately two thirds of those observed in males. The age interaction term was based on data given by Doll (1971) (presented by Breslow and Day 1980 on page 64) which shows that the relative risk of lung cancer in smokers compared to non-smokers varies, (increasing, but neither linearly nor multiplicatively) with age strata. The postulated interaction between smoking intensity and source of information term followed Schoenberg et al.'s (1990) finding of the significance of such a term (between log of the average smoking rate and respondent type) which could be explained by differences in the quality of smoking information between self and proxy respondents.

3.6.6.2 Potential Modifiers of the Radon Effect

As discussed in the literature review (section 2.5.2), *a priori* candidates for modification of a radon effect included age, gender, smoking, air pollution, and sleeping with an open window. All but the latter two, for which information was unavailable in these studies, were investigated; interaction terms to test for non-homogeneity of radon effect were fit for age, gender, the subject's average smoking intensity, and environmental tobacco smoke exposure. As mentioned earlier, descriptive analyses were also produced by levels of each covariate, providing mean radon concentration for cases and controls within each strata. This analysis, although weighting radon exposures in the 50 years prior to ascertainment equally, provided another, more basic means of evaluating the potential for effect modification. Potential between-site differences in smoking effects were investigated by comparing estimates of relative risk associated with standard categories of pack-years from a standard logistic model which controlled for matching variables.

The form of the relationship between smoking and radon in determining lung cancer risk has important implications for the targeting of lung cancer risk reduction efforts; in the case of a multiplicative inter-relation, efforts may be more effectively directed to reducing exposure to tobacco smoke than to radon. Given the anticipated imprecision in the radon risk estimates to detect important modifiers of a putative effect, separate (stratified) analyses by smoking were considered crucial. Although two different indexes of smoking were planned as separate stratification variables, (smoking intensity, and recency of smoking), since the effect of one modified the other, and subjects in Missouri were restricted with respect to recency, separate analyses were conducted within level of recency with adjustment for smoking intensity.

3.7 COMBINED LOGISTIC ANALYSES

3.7.1 FIXED VERSUS RANDOM EFFECTS

If study-specific statistical estimates are 'similar enough' to be combined (see section 3.1), yet another 'similarity judgment' is required as to how the studies will be combined (NRC 1992). This amounts to asking the following questions: 'are the studies sufficiently homogeneous in design and patient sampling protocols...' (NRC 1992), and I would add, *measured* covariates (risk factors and effect modifiers), and *estimable* determinants of aerosol particle-size distribution, RDP equilibrium, home occupancy, and external sources of exposure³¹, '...that they are measuring the same thing', ie., a single *fixed* effect? Or are they estimating different, ie., *heterogeneous*, underlying effects? In the first case, a *fixed effects* assumption would be appropriate, whereas in the latter, one must assume and accommodate, *random effects*. Berlin et al. (1993) summarized the differing interpretations of effect estimates from the two types of model as follows:

'A fitted value from a random effects model μ_i represents the mean of true slopes (b_i) among studies with covariates z_{ik} ³². In contrast, the fitted value from the fixed effects model, b_i , represents the true slope for any study with covariates z_{ik} .' (Berlin et al., page 226).

3.7.1.1 Fixed Effects Models

Fixed effect models assume no differences in the underlying effect among the populations

³¹ In the present analysis, only one of these factors were estimated, and this (occupancy), only in one site.

³² Here i indexes the individual studies, and k , the covariates within each study.

studied (NRC 1992). One logical choice³³ of statistical procedure in this case is the combination of the different independent estimates of the same parameter using the inverse of their variances as weights (Laird and Mosteller 1990), as follows:

$$\hat{\beta} = \frac{\sum s_k^{-1} \hat{\beta}_k}{\sum s_k^{-1}} \quad (3.10)$$

Where:

$\hat{\beta}$ denotes the pooled estimate of overall effect (for radon) for all sites,

$\hat{\beta}_k$ represents the estimate of the single-valued model parameter β in site k (where in our case $k = 1, 2, 3$), and,

s_k denotes the estimated variance of $\hat{\beta}_k$.

If the estimates are uncorrelated, then the variance of this overall parameter estimate effect is the inverse of the sum of the inverse variances:

$$Var(\hat{\beta}) = \frac{1}{\sum \frac{1}{s_k}} \quad (3.11)$$

3.7.1.2 Random Effects Models

Equation 3.11 will result in a confidence interval which is artificially narrow, since the

³³ Another natural choice, according to Laird and Mosteller (1990) is equal importance, with combining weights of 1 for each estimate.

confidence interval for the overall estimate of effect reflects the random variation within each study, but not potential heterogeneity between studies (Thompson and Pocock, 1991). According to the latter authors, random effects models assume that (even if there is no within study variation due to finite samples) the true treatment effects in the different possible universe of studies are randomly positioned about some central value and that the individual studies observed are representative of some hypothetical population of studies. Such models take into account both random variation within studies as well as heterogeneity between them. When heterogeneity is present, the confidence interval will be wider than it would have been on the fixed-effects assumption, thus introducing a degree of statistical caution (Laird and Mosteller 1990; Whitehead and Whitehead 1991) and model realism.

The two-stage method recently applied to radon data among miners (Wang et al. 1996) was employed.

The first stage involves obtaining study-specific estimates of effect. These estimates and their standard errors $\{\hat{\beta}_k, s_k; k = 1, \dots, K\}$ ($k = 3$) are used as input for the second stage, at which the study-specific estimates are combined.

Stage 1:

Following development of a common logistic model, site-specific estimates were obtained from an analysis in which the data from all three sites were combined (jointly analysed). Except where variables varied importantly between site, this model assumed that the covariates could be estimated by common coefficients across site. Two indicator variables were used to denote site. Interactions between the site indicator and the radon term provided a mechanism for estimating three site-specific radon terms. This model for the log (Rate Ratio) included the following sum:

$$+ \gamma_{NJ} I_{NJ} + \gamma_{MISS} I_{MISS} + (\beta_{radon} + \delta_{NJ} I_{NJ} + \delta_{MISS} I_{MISS}) w^* \quad (3.12)$$

where I_{NJ} and I_{MISS} are indicators of whether a subject was in the New Jersey or Missouri study, γ_{NJ} and γ_{MISS} reflect the different case-control ratios in the New Jersey and Missouri studies relative to Winnipeg, and β_{radon} , $\beta_{radon} + \delta_{NJ}$, and $\beta_{radon} + \delta_{MISS}$, are the radon “slopes” in Winnipeg, New Jersey, and Missouri, respectively. The variances for these stage 1 estimates for New Jersey and Missouri were calculated according to the formula for the variance of a sum.

Stage 2:

The second stage begins by defining:

$$\bar{\beta} = \frac{\sum_k s_k^{-1} \hat{\beta}_k}{\sum_k s_k^{-1}} \quad (3.13)$$

a weighted average of the three site-specific radon slopes, (which is the pooled estimate of overall effect assuming fixed effects);

$$\hat{\tau} = \frac{\sum_k s_k^{-1} (\hat{\beta}_k - \bar{\beta})^2 - (K - 1)}{\sum_k s_k^{-1} - \frac{\sum_k s_k^{-2}}{\sum_k s_k^{-1}}} \quad (3.14)$$

an estimate of the heterogeneity of the slopes between sites, net of within site variation (if the estimate of tau is negative, tau is set to zero); and

$$w_k = \frac{(\hat{\tau} + s_k)^{-1}}{\sum_k (\hat{\tau} + s_k)^{-1}} \quad (3.15)$$

a new weighting term (which is inversely proportional to an amalgam of the within and between study variances) to be applied to the three site-specific slopes. The pooled estimate $\hat{\beta}$ of the mean slope over all possible sites (based on the sample of three) is then given by:

$$\hat{\beta} = \sum_k w_k \hat{\beta}_k \quad (3.16)$$

Provided the $\hat{\beta}_k$'s are statistically uncorrelated, the variance of this estimate is given by:

$$Var(\hat{\beta}) = \frac{1}{\sum \frac{1}{\hat{\tau} + s_k}} \quad (3.17)$$

Note the similarity in form between equations 3.11 (fixed effects variance) and 3.17 (random effects variance). The latter adds an estimate of between site variance to the within site variances.

A statistical test for homogeneity of the $\hat{\beta}_k$ among sites is given by:

$$\chi_{homog}^2 = \sum_k s_k^{-1} (\hat{\beta}_k - \bar{\beta})^2 \quad (3.18)$$

which has a chi-square distribution with $K - 1$ degrees of freedom.

The shrinkage estimator of the site-specific effect β_k is given by:

$$\hat{\beta}_k^* = \frac{s_k \hat{\beta} + \hat{\tau} \hat{\beta}_k}{s_k + \hat{\tau}}, \quad (3.19)$$

which is termed the 'site-specific stage two estimate', and is a weighted mean of the pooled and site-specific estimates. Shrinking the site-specific estimates toward the pooled (mean) estimate reflects the fact that the stage one site-specific estimates differ both because of random (within site) and systematic (between site) differences. By 'shrinking' one hopes to improve the overall mean square error across the three sites, but possibly at the expense of one or more of the individual mean square errors. One could also think of the stage two estimate as a weighted average of two legitimate estimates, one estimated within a given study, and the other an average across studies. The stage two estimate thus incorporates information from the entire (combined) data-set .

3.7.2 ADEQUACY OF USING COMMON REGRESSION COEFFICIENTS FOR COVARIATES

The adequacy of common coefficients for the covariates included in the analysis of the combined data was assessed in three ways:

3.7.2.1 Combination of Site-Specific Model Estimates

The overall point and interval estimates of RR for radon obtained from combining the estimates within each site from a) a single regression analysis applied to the combined data, so that coefficients for covariates were constant over site, were compared to those from b) separate regressions (with the same parameters) fitted for each site in parallel, so that coefficients for covariates could vary from site to site.

3.7.2.2 Deviance of Sum of Site-Specific Versus Combined Model

The sum of the $-2 \log$ likelihoods of the common model run separately in each site was compared to the $-2 \log$ likelihood of the common model run on the combined data (with site

by radon interactions and smoking by site interactions as necessary). The resulting statistic was assumed to follow a Chi squared distribution with degrees of freedom equal to the difference in number of fitted parameters for the parallel analyses and the analysis of the combined data.

3.7.2.3 Significance of Site by Covariate Interaction Terms in An Analysis of the Combined Data

Interaction terms were created for each site and each covariate and an automated (stepwise) regression analysis performed. A p-value of 0.2 was selected for entrance to and retention in the model. A type I error rate of 0.2 was consistent with that adopted formerly adopted (in section 3.6.7) for interaction terms in general.

3.8 SENSITIVITY ANALYSES

3.8.1 TIME WINDOW WEIGHTING

In addition to the default time window weighting as defined in Table 3.6 above, two additional weighting scenarios were applied, effectively eliminating the most distant exposure windows in which annual exposure data are increasingly missing and therefore imputed. Imputation of gaps in subject's reconstructed exposure history results in Berkson type error and reduced power to detect departures from the null model.

The three weighting scenarios are analogous to analysing three distinct combinations of time windows as follows: 1) Three separate time windows of 5 - 14, 15 - 24, and 25 - 50 years prior to ascertainment, respectively. Decreasing weight is given to more historically distant exposures, where uncertainty (due to missing data and representativity of currently measured for historical exposures) is greatest; 2) Two separate time windows, the first summarizing

exposure 5 - 14 years prior to ascertainment, and the second, 15 - 24 years. The most distant and uncertain time window is excluded; and, 3) A single time window of 5 - 14 years, observed to be the most predictive interval in the BEIR VI analysis (NRC 1998). The two most distant time windows are excluded.

The alternative weights are given in Table 3.8 below (for estimation of cumulative radon exposure and average radon concentration, respectively), and compared with the default weighting scenario and with the hypothetical scenario in which each year in the entire exposure period 5 - 50 years prior to ascertainment was weighted equally. The expression of these window weights as annual weights within windows reinforces the very small (or zero) weights accorded historical exposures occurring 25 or more years prior to ascertainment.

Table 3.8
Time Window (and Annual) Weights For Estimation of
Cumulative Radon Exposure Under Three Scenarios

Time Window in Years Prior to Ascertainment	Time Window Weighting Scenario			
	5 to 50 (Equal Weights For Each Year)	5 -50 Default [‡]	5 - 24	5 - 14
5 - 14 years	0.2174 (0.0217)	1.0000 (0.1000)	1.0000 (0.1000)	1.0000 (0.1000)
15 - 24 years	0.2174 (0.0217)	0.7500 (0.0750)	0.7500 (0.0750)	0.0000 (0.0000)
25 - 50 years	0.5652 (0.0217)	0.4750 (0.0183)	0.0000 (0.0000)	0.0000 (0.0000)
Sum of weights	1	2.225	1.75	1

‡ Derived from BEIR VI analysis (NRC 1998).

Table 3.9 provides these weights re-scaled to sum to one for analysis of average radon concentration. Within each scenario, the three time windows maintain the same ratio of weights for the estimation of cumulative radon and radon concentration.

Table 3.9

Time Window (and Annual) Weights For Estimation of Average Radon Concentration Under Three Scenarios

Time Window in Years Prior to Ascertainment	Time Window Weighting Scenario			
	5 to 50 (Equal Weights For Each Year)	5 -50 Default [#]	5 - 24	5 - 14
5 - 14 years	0.2174 (0.0217)	0.4494 (0.0449)	0.5714 (0.0571)	1.0000 (0.1000)
15 - 24 years	0.2174 (0.0217)	0.3371 (0.0337)	0.4286 (0.0429)	0.0000 (0.0000)
25 - 50 years	0.5652 (0.0217)	0.2135 (0.0082)	0.0000 (0.0000)	0.0000 (0.0000)
Sum of weights	1	1	1	1

Derived from BEIR VI analysis (NRC 1998).

3.8.2 MODEL FORM

The common logistic model developed in section 3.6 was consistent with the treatment of radon in the original analyses of residential case-control studies. Previous combined analyses of radiogenic hazards, both of cancer among nuclear industry workers (Cardis et al. 1995) and of lung cancer among underground miners, adopted an additive model of relative risk, specifically the excess relative risk (ERR) model (NRC 1988, Lubin et al. 1994a). The multiplicative and additive models are members (each as special cases) of the family of general relative risk models (Moolgavkar and Venzon 1987). At small relative risks, (i.e.

when $\beta w < 0.1$ in an ERR model where w is units of radon exposure), the logistic model predicts RR estimates close to that of an ERR model. Although the differences between the predictions of the two models are expected to be small, a set of analyses was run with the linear excess relative risk model.

As in relative risk models in general, the lung cancer incidence rate $r(x, z, w^*)$ is modelled as a product of the background incidence rate among those unexposed to any risk factors $r_0(x)$ and an exposure-response function $RR(z, w^*)$:

$$r(\underline{x}, \underline{z}, w^*) = r_0(\underline{x}) RR(\underline{z}, w^*) \quad (3.20)$$

where:

- $r(\underline{x}, \underline{z}, w^*)$ denotes the lung cancer incidence rate;
- $r_0(\underline{x})$ denotes the background incidence rate, in those with no (zero) radon exposure;
- $RR(\underline{z}, w^*)$ denotes the rate ratio for lung cancer of two populations with the same baseline risk as determined by their respective covariates, one having w^* units of radon exposure and the other no radon;
- $\underline{x}$ denotes a vector of covariates describing the background incidence rate;
- $\underline{z}$ denotes a vector of covariates which potentially modify the exposure-response relationship; and,
- w^* denotes effective radon exposure (in working level months).

Breslow and Day (1980) illustrated, with several examples from the literature of cancer epidemiology, that the relative risk provides a stable measure of association in a wide variety of human populations.

Assuming a homogeneous population in respect to covariates, denote by “ w ” radon exposure in continuous units, and the rate ratio associated with radon exposure relative to absence of

exposure ($w = 0$) by $RR(w, \beta)$. Define, for illustrative purposes, two covariates, c_1 and c_2 , which do not modify the effect of w .

In the exponential regression model,

$$\begin{aligned} RR &= \exp(\alpha_1 c_1) \exp(\alpha_2 c_2) \exp(\beta' w) \\ &= \exp(\alpha_1 c_1 + \alpha_2 c_2 + \beta' w) \end{aligned} \quad (3.21)$$

and in an additive (excess) relative risk model,

$$\begin{aligned} RR &= \exp(\alpha_1 c_1) \exp(\alpha_2 c_2) (1 + \beta' w) \\ &= \exp(\alpha_1 c_1 + \alpha_2 c_2) (1 + \beta' w), \end{aligned} \quad (3.22)$$

assuming that the background rate is of the multiplicative form of equation 3.23,

$$r_0(\underline{x}) = \exp(\underline{\alpha} \underline{x}) \quad (3.23)$$

with:

- $\underline{x}$ denoting a vector of covariates describing the background incidence rate and including a constant; and,
- $\underline{\alpha}$ designating the corresponding vector of parameters, with the one associated with the constant being a non-separable function of the sampling fraction of controls and the absolute background rate of lung cancer.

A consequence of the multiplicative assumption for the baseline rate is that the logistic and linear models fit differed only in their treatment of radon exposure.

While the exponential model can be fit with regular logistic regression, the ERR can not, since it requires a link function that is non-linear in the coefficients. The exponential regression model assumes an exponential increase in the rate ratio³⁴ as function of w , and multiplicative joint effects between the exposure and each of the covariates (Greenland 1989). The excess relative risk (ERR) model assumes an linear increase in the rate ratio as a function of w . Since incorrect model specification can lead to biased effect estimates and standard errors, and model-based confidence intervals which do not cover the true effect measure at the nominal coverage rate, and since OCMI was expected to induce some bias in the non-linear model, both types of models were fit. Following the development of a common multivariate logistic model, sensitivity to model form of the estimate of RR as a function of radon was assessed by fitting an additive ERR model which assumed the same set of covariates as in the multiplicative model.

For the analyses of the combined data, terms for study site are added reflecting the varying case to control ratio across site and possible heterogeneity in the radon effect; the resulting theoretical model for the ERR case is:

$$r(x, w^*) = \exp(\alpha x + \gamma_{NJ} I_{NJ} + \gamma_{MISS} I_{MISS}) [1 + (\beta'' + \delta_{NJ} I_{NJ} + \delta_{MISS} I_{MISS}) w^*], \quad (3.24)$$

where β'' denotes a unit increase in excess relative risk per unit effective radon exposure for the reference site (Winnipeg). The other components as defined previously for expression 3.12. As before, β'' , $\beta'' + \delta_{NJ}$, and $\beta'' + \delta_{MISS}$, are the fitted radon slopes for the three sites, Winnipeg, New Jersey, and Missouri, respectively.

³⁴ In a case-control study which employs incidence density sampling, the odds ratio estimates the incidence density ratio or rate ratio (RR) of disease. Incidence density sampling implies that incident cases and controls are selected periodically over the same duration of the study, and that controls are selected from a roster of individuals whose exposure distribution is representative of the base population which gave rise to the cases.

In order to compare the fit of the logistic to the ERR model, the predicted exposure-response curves (with radon expressed as a continuous variable) predicted from the linear and logistic models, together with data-points which represented estimates of RR within radon categories, were plotted. The model predictions and categorical data point estimates were all adjusted for the final common constellation of covariates in exponential form as assumed in equation 3.23.

4 RESULTS

The results are presented in the following order: first, the covariate distributions amongst subjects included and excluded following completion of the personal interviews are analysed in order to evaluate the potential for differential selection; second, information relating to the reconstruction of radon exposure histories is provided to evaluate the potential for information bias; third, the relationships amongst radon exposure, outcome, and covariates are evaluated to ascertain the potential for confounding and effect modification; fourth, the data gathered with respect to cigarette smoking evaluated for consistency across sites and with an external historical comparison; fifth, the development of a common logistic regression model for the prediction of lung cancer is elaborated; sixth, the results of the final analysis of the combined data are summarized; and seventh and finally, the sensitivity of the results to time window definition and statistical model form is analysed.

4.1 COVARIATE DISTRIBUTIONS AMONGST INTERVIEWED SUBJECTS

Tables 4.1, 4.2, and 4.3 compare the covariate distributions of subjects included and excluded following interview completion, by outcome status, in New Jersey, Winnipeg, and Missouri, respectively. This analysis includes all subjects excluded post-interview in Winnipeg and Missouri. In New Jersey, however, data on excluded subjects were limited to the 20% of cases and 20% of controls excluded post-interview by the present author because of the lack of any home radon monitoring data on these subjects. The inclusion of these subjects using imputed radon data for their entire exposure time window, as apparently done in the original analysis by Schoenberg et al. (1992), was non-informative.

Table 4.1

**Covariate Distributions Among Included and Excluded Subjects:
New Jersey USA**

Covariate	Cases		Controls	
	Frequency (%) Among Included	Frequency (%) Among Excluded [†]	Frequency (%) Among Included	Frequency (%) Among Excluded [†]
Number of Subjects	351 (100)	129 (20) [†]	308 (100)	134 (20) [†]
Age at Ascertainment ^{•*}				
<55	52 (15)	18 (14)	52 (17)	17 (13)
55 - 64	117 (33)	46 (36)	107 (35)	49 (37)
65 - 74	135 (36)	44 (34)	96 (31)	47 (35)
75 - 79	35 (10)	16 (12)	30 (10)	11 (8)
≥80	22 (6)	5 (4)	23 (7)	10 (7)
Calendar Year of Ascertainment*				
1982	121 (34)	55 (43)	58 (19)	24 (18)
1983	230 (66)	74 (57)	222 (72)	102 (76)
1984	0 (0)	0 (0)	28 (9)	8 (6)
Respondent [•]				
Subject	200 (57)	67 (52)	169 (55)	71 (53)
Proxy	151 (43)	62 (48)	139 (45)	63 (47)
Gender				
Female	351 (100)	129 (100)	308 (100)	134 (100)
Male	0 (0)	0 (0)	0 (0)	0 (0)
Education (Highest level undertaken) ^{•*}				
Primary	23 (7)	20 (16)	42 (14)	16 (12)
Secondary	227 (65)	80 (62)	171 (56)	80 (60)
Post-Secondary	101 (29)	29 (22)	95 (31)	38 (28)
Annual Household Income	Not Available			
Race [•]				
"White"	344 (98)	119 (92)	297 (96)	126 (94)
"Other"	7 (2)	10 (8)	11 (4)	8 (6)
Smoking History*				
Never-Smoker ^{†‡}	49 (14)	19 (15)	164 (53)	69 (51)
Cigarette Only Smoker	302 (86)	110 (85)	144 (47)	65 (49)

Covariate	Cases		Controls	
	Frequency (%) Among Included	Frequency (%) Among Excluded [†]	Frequency (%) Among Included	Frequency (%) Among Excluded [†]
Timing of Cigarette Smoking*				
Never Smoker	49 (14)	19 (15)	164 (53)	69 (51)
Current Smoker (<14 years since cessation)	286 (81)	105 (81)	116 (38)	49 (37)
Former Smoker (≥ 14 years since cessation)	16 (5)	5 (4)	28 (9)	16 (12)
Intensity of Smoking (Packs per Day)				
Never Smoker	49 (14)	19 (15)	164 (53)	69 (51)
< 1	132 (38)	39 (30)	84 (27)	44 (33)
1 - 2	163 (47)	61 (47)	56 (18)	20 (15)
> 2	7 (2)	10 (8)	4 (1)	1 (1)
Pack Years of Smoking**				
Never Smoker	49 (14)	19 (15)	164 (53)	69 (51)
< 10	13 (4)	4 (3)	28 (9)	17 (13)
10 - 19	33 (9)	5 (4)	28 (9)	16 (12)
20 - 29	39 (11)	13 (10)	27 (9)	6 (4)
30 - 39	61 (17)	17 (13)	15 (5)	7 (5)
≥ 40	156 (44)	71 (55)	46 (15)	19 (14)
Environmental Tobacco Smoke Exposure[‡]				
Never lived with smoker	20 (6)	11 (9)	32 (10)	8 (6)
Ever lived with smoker	331 (94)	118 (91)	274 (89)	123 (92)
Unknown	0 (0)	0 (0)	2 (1)	3 (2)

† Limited to those subjects excluded by VS Catalan because of lack of home radon monitoring data. These constituted 20% of the cases and 20% of controls excluded following completion of the personal interview.

● Living control subjects were frequency matched to living cases by age at ascertainment and 'race'; deceased cases were individually matched to controls by race, age, and closest date of death; proxy respondents were used for the latter cases and controls. Ascertainment is defined as diagnosis for cases and interview for controls.

Totals unequal to 100% due to rounding.

* Among those included, significant difference in case and control distribution ($p < 0.05$) (significant predictor of lung cancer).

● Missing value of education for 1 case imputed with the observed control mean.

† Includes 2 subjects (1 case and 1 control) reported to have smoked <100 cigarettes or < 1 year.

§ Smoking refers to cigarettes only; never smokers included 1 case and 1 control reported to have smoked either a pipe or cigars.

◇ Up to year of ascertainment.

‡ From cigarettes only.

Table 4.2

**Covariate Distributions Among Included and Excluded Subjects:
Winnipeg CA**

Covariate	Cases		Controls	
	Frequency (%) Among Included	Frequency (%) Among Excluded [†]	Frequency (%) Among Included	Frequency (%) Among Excluded [†]
Total Number of Subjects Included	701 (100)	37 (100) [†]	724 (100)	14 (100) [†]
Age at Ascertainment ^{••}				
<55	124 (18)	5 (14)	137 (19)	2 (14)
55 - 64	253 (36)	12 (32)	248 (34)	3 (36)
65 - 74	290 (41)	17 (46)	279 (39)	6 (43)
75 - 79	34 (5)	3 (8)	53 (7)	1 (7)
≥80	0 (0)	0 (0)	7 (1)	0 (0)
Calendar Year of Ascertainment ^{••}				
1984	64 (9)	9 (24)	0 (0)	0 (0)
1985	118 (7)	11 (30)	123 (17)	4 (29)
1986	112 (16)	5 (14)	113 (15)	4 (29)
1987	120 (17)	6 (16)	94 (13)	2 (14)
1988	124 (18)	2 (5)	147 (20)	1 (7)
1989	124 (18)	4 (11)	132 (18)	1 (7)
1990	39 (6)	0 (0)	115 (16)	2 (14)
Respondent [*]				
Subject	459 (65)	22 (60)	649 (90)	11 (79)
Proxy	242 (66)	15 (41)	75 (10)	3 (21)
Gender [•]				
Female	247 (34)	13 (35)	246 (34)	4 (29)
Male	464 (66)	24 (65)	478 (66)	10 (71)
Education (Highest level undertaken) ^{•••}				
Primary	103 (15)	5 (14)	54 (7)	2 (14)
Secondary	355 (51)	19 (51)	300 (41)	5 (36)
Post-Secondary	243 (35)	13 (35)	370 (51)	7 (50)
Annual Household Income at Interview ^{•••} (\$ Canadian)				
<\$10,000	45 (6)	3 (8)	36 (5)	1 (7)
\$10,000-\$19,999	184 (26)	15 (41)	132 (18)	5 (36)
\$20,000-\$29,999	228 (33)	8 (22)	192 (27)	3 (21)
\$30,000-\$49,999	190 (27)	9 (24)	240 (33)	5 (36)
≥\$50,000	54 (8)	2 (5)	124 (17)	0 (0)

Covariate	Cases		Controls	
	Frequency (%) Among Included	Frequency (%) Among Excluded [†]	Frequency (%) Among Included	Frequency (%) Among Excluded [†]
Country of Birth**				
Canada	522 (79)	26 (70)	590 (81)	9 (64)
Other	149 (21)	11 (30)	134 (19)	5 (36)
Smoking History**				
Never-Smoker [†]	24 (3)	0 (0)	219 (30)	4 (29)
Cigarette Only Smoker	608 (87)	26 (70)	380 (52)	5 (36)
Pipe and/or Cigar Smoker [†]	1 (0)	1 (3)	23 (3)	0 (0)
Cigarette and Pipe/Cigar Smoker	68 (10)	1 (3)	102 (14)	1 (7)
Unknown	0 (0)	9 (24)	0 (0)	4 (29)
Timing of Cigarette Smoking**				
Never or Pipe/Cigar Smoker	25 (4)	1 (3)	242 (33)	4 (29)
Current Smoker (<14 years since cessation)	616 (88)	26 (70)	296 (41)	5 (36)
Former Smoker (≥ 14 years since cessation)	60 (9)	1 (3)	186 (26)	1 (7)
Unknown	0 (0)	9 (24)	0 (0)	4 (29)
Intensity of Smoking (Packs per Day)				
Never or Pipe/Cigar Smoker	25 (4)	1 (3)	242 (33)	4 (29)
< 1	256 (37)	11 (30)	263 (36)	2 (14)
1 - 2	362 (52)	13 (35)	200 (28)	4 (29)
> 2	58 (8)	1 (3)	19 (3)	0 (0)
Unknown	0 (0)	9 (24)	0 (0)	4 (29)
Pack Years of Smoking** ^o				
Never or Pipe/Cigar Smoker	25 (4)	1 (3)	242 (33)	4 (29)
<10	33 (5)	7 (19)	91 (13)	2 (14)
10 - 19	49 (7)	2 (5)	88 (12)	0 (14)
20 - 29	86 (12)	7 (19)	90 (12)	1 (7)
30 - 39	124 (18)	3 (8)	71 (10)	0 (0)
≥ 40	384 (55)	13 (35)	142 (20)	3 (21)
Unknown	0 (0)	9 (24)	0 (0)	4 (29)
Environmental Tobacco Smoke Exposure** [‡]				
Never lived with smoker	60 (9)	7 (19)	88 (12)	3 (21)
Ever lived with smoker	640 (91)	29 (78)	636 (88)	11 (79)
Unknown ^o	1 (0)	1 (3)	0 (0)	0 (0)

† These subjects were excluded by VS Catalan because of lack of information on smoking (9 cases and 4 controls) or home radon monitoring data (28 cases and 10 controls).

● Controls were frequency matched to subjects by age at ascertainment (frequency) and gender (pairwise). Ascertainment was defined as diagnosis for cases and interview for controls.

o Eight missing values (7 cases, 1 control) for education category were imputed with the subject's income-stratum-specific control mean; 306 missing values (165 cases, and 141 controls) for income were imputed with the control mean for the subject's education-stratum-specific control mean; where both income and education were missing for

- a subject, the respective overall control mean was used.
- * Significant difference in case and control distribution ($p < 0.05$).
 - # Some totals unequal to 100% due to rounding error.
 - b Those born outside Canada were grouped together in subsequent analyses.
 - † Includes subjects reported to have smoked <100 cigarettes or < 1 year.
 - § Included with never-smokers in subsequent analyses.
 - ◇ Up to year of ascertainment.
 - ‡ From cigarettes only.
 - ⊠ Unknown exposure to environmental tobacco smoke coded as unexposed in subsequent analyses.
 - ** Among included subjects, significant difference in case and control distribution ($p < 0.05$) with one-tailed exact test (but not two-tailed test) .

Table 4.3

**Covariate Distributions Among Included and Excluded Subjects:
Missouri USA**

Covariate	Cases		Controls	
	Frequency (%) Among Included	Frequency (%) Among Excluded ¹	Frequency (%) Among Included	Frequency (%) Among Excluded ¹
Number of Subjects	538 (100)	80 (100)	1183 (100)	219 (100)
Age at Ascertainment ² :				
<55	46 (9)	1 (1)	95 (8)	8 (4)
55 - 64	90 (17)	11 (14)	202 (17)	31 (14)
65 - 74	174 (32)	26 (33)	384 (32)	73 (33)
75 - 79	120 (22)	22 (28)	306 (26)	63 (29)
≥80	108 (20)	20 (25)	196 (17)	44 (20)
Calendar Year of Ascertainment*				
1986	85 (16)	16 (20)	0 (0)	0 (0)
1987	112 (21)	14 (18)	0 (0)	0 (0)
1988	147 (27)	16 (20)	192 (16)	39 (18)
1989	129 (24)	19 (24)	193 (16)	45 (21)
1990	64 (12)	15 (19)	506 (43)	86 (40)
1991	1 (0)	0 (0)	292 (25)	49 (22)
Respondent*				
Subject	197 (37)	25 (31)	1183 (100)	219 (100)
Proxy	341 (63)	55 (69)	0 (0)	0 (0)
Gender				
Female	538 (100)	80 (100)	1183 (100)	219 (100)
Male	0 (0)	0 (0)	0 (0)	0 (0)
Education (Highest level undertaken) ³				
Primary	203 (38)	37 (46)	438 (37)	97 (44)
Secondary	225 (42)	32 (40)	437 (37)	75 (34)
Post-Secondary	110 (20)	11 (14)	308 (26)	47 (21)
Annual Household Income at Interview ⁴ *(US\$)				
<\$10,000	210 (39)	40 (50)	362 (31)	93 (42)
\$10,000-\$19,999	162 (30)	26 (33)	427 (36)	78 (36)
\$20,000-\$29,999	90 (17)	8 (10)	210 (18)	32 (15)
\$30,000-\$49,999	47 (9)	3 (4)	119 (10)	13 (6)
≥\$50,000	29 (5)	3 (4)	65 (5)	3 (1)
Race ⁵				
'White'	538 (100)	80 (100)	1183 (100)	219 (100)

Covariate	Cases		Controls	
	Frequency (%) Among Included	Frequency (%) Among Excluded [†]	Frequency (%) Among Included	Frequency (%) Among Excluded [†]
Smoking History ^{□*}				
Never-Smoker [†]	396 (74)	60 (65)	1049 (89)	196 (90)
Cigarette Smoker	142 (26)	20 (25)	130 (11)	23 (11)
Pipe/Cigar Smoker	0 (0)	0 (0)	0 (0)	0 (0)
Mixed Smoker [‡]	0 (0)	0 (0)	2 (0)	0 (0)
Timing of Cigarette Smoking ^{□*}				
Never Smoker	396 (74)	60 (65)	1051 (89)	196 (90)
Current Smoker				
(<14 years since cessation)	0 (0)	0 (0)	0 (0)	0 (0)
Former Smoker				
(≥ 14 years since cessation)	142 (26)	20 (25)	132 (11)	23 (11)
Pack-Years of Smoking ^{§*}				
Never Smoker	396 (74)	60 (65)	1051 (89)	196 (90)
< 10	39 (7)	10 (13)	66 (6)	11 (5)
10 - 19	35 (7)	1 (1)	25 (2)	3 (1)
20 - 29	21 (4)	2 (3)	21 (2)	2 (1)
30 - 39	24 (4)	1 (1)	9 (1)	4 (2)
≥ 40	23 (4)	6 (8)	11 (1)	3 (1)
Intensity of Smoking (Packs per Day)				
Never Smoker	396 (74)	60 (65)	1051 (89)	196 (90)
< 1	71 (13)	8 (10)	80 (7)	14 (0)
1 - 2	65 (12)	11 (14)	48 (4)	9 (4)
> 2	6 (1)	1 (1)	4 (0)	0 (0)
Environmental Tobacco Smoke Exposure [‡]				
Never lived with smoker	279 (52)	34 (43)	581 (49)	100 (46)
Ever lived with smoker	259 (48)	46 (58)	602 (51)	119 (54)

● Controls were frequency matched to cases by age at ascertainment. Ascertainment is defined as diagnosis for cases and interview for controls.

Some totals unequal to 100% due to rounding error.

○ Among those subjects included, fifty-two missing values (for 26 cases and 26 controls) of education category were imputed with the control mean education level specific to the subject's income stratum; Two-hundred four missing values of annual income at interview (for 83 cases and 121 controls) were imputed with the control mean income level specific to the subject's education stratum.

* Significant difference in case and control distribution ($p < 0.05$).

□ Inclusion criteria restricted this study to 'white' women who had not smoked in the 15 years prior to ascertainment.

† Includes 87 subjects (19 cases, 68 controls) reported to have smoked <100 cigarettes or < 1 year.

§ Included with cigarette smokers in subsequent analyses.

◇ Up to year of ascertainment.

‡ From cigarettes only.

A slightly greater loss of direct than proxy respondents occurred in both case and control series in the two earlier studies: direct (live) respondents constituted 52% of cases and 53% of controls excluded in New Jersey where source of information was a matching variable; 59% of cases and 79% of controls excluded in Winnipeg were live respondents; in Missouri where only live controls were recruited, 31% of the excluded cases were live respondents. Since source of information is not a risk factor, it was included as a term in multi-variate models only in the one site on which it was a matching variable (New Jersey), where, not surprisingly, it was an unimportant covariate. Since the quality of smoking information was expected to vary with the source of information, however, an interaction term between type of respondent and smoking (intensity or recency) was fit in both sites where proxies were controls were recruited (Winnipeg and New Jersey) and in the analysis of the pooled data.

In Winnipeg and Missouri, controls were slightly better educated and of higher income than cases. Slight but differential distortion of the education distribution by exclusions was apparent in New Jersey and Winnipeg. It was therefore necessary to control in the analysis for any potential bias arising from selection on this predictor of lung cancer.

In Winnipeg and Missouri, but not in New Jersey, those subjects retained in the analysis tended to be slightly younger than those excluded, both amongst cases and controls. Age was included by default in all regression models.

In New Jersey, non-white race was less prevalent amongst controls than cases, and amongst included than excluded in each outcome group. In Winnipeg, significantly different distributions with respect to country of birth were evident between cases and controls. Selection fractions post-interview formed an odds ratio (OR^1) of one, obviating concern for differential selection assuming similar selection factors operated pre- and post-interview.

¹ Since the odds ratio estimates the rate ratio, the two terms are used interchangeably throughout the text when referring to effect estimates based on logistic regression analyses.

Missouri employed restriction to control for ethnic variation. Since ethnicity was possible selection factor in New Jersey and risk factor in Winnipeg, it was an important control variable.

Cigarette use was less prevalent amongst controls than cases in all sites, with excluded subjects demonstrating similar distributions to those included with respect to timing, intensity, and combined intensity and duration (pack-years) of use. There was no threat of differential attrition of smokers. Nevertheless, intensive efforts were undertaken to account for this most important lung cancer risk factor.

Only a very few subjects, 1 (<1%) case and 23 (3%) controls in Winnipeg, reported smoking only either cigars or pipes. At the time of study conduct such products were not in vogue among the female populations to which the other two studies were restricted. Moreover, even among males, the consumption of pipe tobacco and cigar products has markedly declined over the past century (Bartecchi et al. 1994). In several cohort studies in the US the risk of lung cancer was shown to be much smaller for pipe and cigar smokers than for cigarette smokers, although greater than that for non-smokers (US DHHS 1982). Those Winnipeg subjects reporting smoking of only pipes or cigars were therefore included with never-smokers in subsequent analyses.

Similar proportions of subjects were excluded in case and control groups in all other covariate strata and in all sites with the exception of calendar year of ascertainment. Defined as the year of diagnosis of cases, and the year of interview of the controls, ascertainment was uniformly later amongst controls than cases, by one to two years. This slight delay was a result of sampling procedures in which controls were matched to cases following the diagnosis and identification of the latter. There was a tendency in Winnipeg and Missouri for those excluded to have been ascertained earlier in the conduct of the study than those included; after accounting for the one (Winnipeg) or two (Missouri) calendar year delay

between case and control recruitment, this trend was, however, consistent between cases and controls. Since year of ascertainment was not associated with radon exposure, this design variable posed no threat of bias.

It should be noted that whether the observations made above in regard to differences between those included and excluded following questionnaire completion are generalizable to selection which occurred prior to the questionnaire is unknown. Moreover, no information was available for the majority of subjects excluded in New Jersey. Considering the extensive efforts undertaken to control for the effects of the important measured risk factors, (namely smoking, socio-economic status and ethnicity), however, bias arising from differential attrition is not an important concern in the interpretation of the (adjusted) results. The possibility of confounding or effect modification relies on an association with radon, to be examined later (see section 4.3).

Although the possibility of bias due to unmeasured covariates, such as exposure to other environmental or occupational exposures to lung carcinogens, cannot be ruled out, there is no obvious mechanism by which selection could have occurred on these factors, and no reason to expect them to be related to radon exposure in the study sites under consideration. Unmeasured covariates are therefore assumed to pose no important threat to the validity of the study inferences.

4.2 RADON EXPOSURE HISTORIES

4.2.1 Home Identification and Radon Monitoring

Residential inclusion criteria varied by site. In New Jersey, length of residency in a single home determined subject inclusion, resulting in a relatively homogeneous population of residentially stable persons. Budgetary considerations restricted subjects to those who had

lived in 'index residences'² for a minimum of seven years, and for whom a maximum of three residences covered 14 or more years of the targeted 25-year exposure time window. This resulted in a much smaller number of homes per subject identified for measurement than in the two other sites, but a slightly smaller proportion of missing radon exposure data (considered in detail in section 4.2.3).

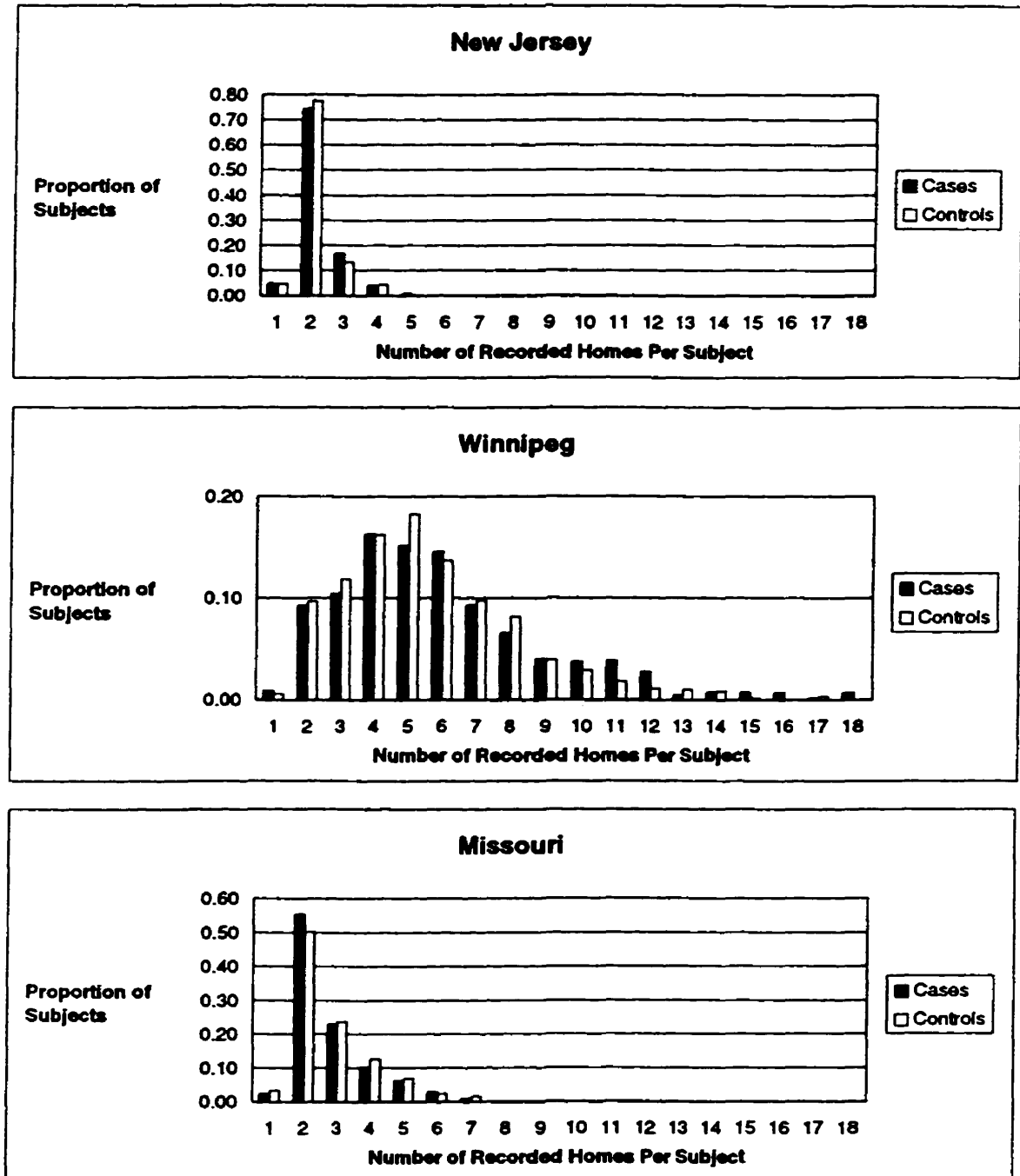
In Winnipeg, all previous homes of greater than 1 year residence were recorded in data files and eligible for radon monitoring, whereas in Missouri, where the same criteria applied to monitoring, only those homes for which radon monitoring was completed were retained in the data files.

The number of homes per subject that were both identified by the subjects and recorded by the investigators in their data files is depicted in Figure 4.1. The average of this distribution (the sum of the product of the number of homes and the proportion of subjects) is 2.2 in New Jersey cases and controls, 5.9 and 5.5 in Winnipeg cases and controls, respectively, and 2.7 and 2.8 in Missouri cases and controls, respectively. Between site differences reflect the aforementioned differences in criteria for home identification and recording. Although it is obvious from Figure 4.1 is that the Winnipeg subjects were highly mobile, the extent to which this was also true in the other two sites is unfortunately not ascertainable from the available data.

² See section 3.3.2.1 for the definition of an "index residence" in New Jersey.

Figure 4.1

Number of Homes Identified and Recorded

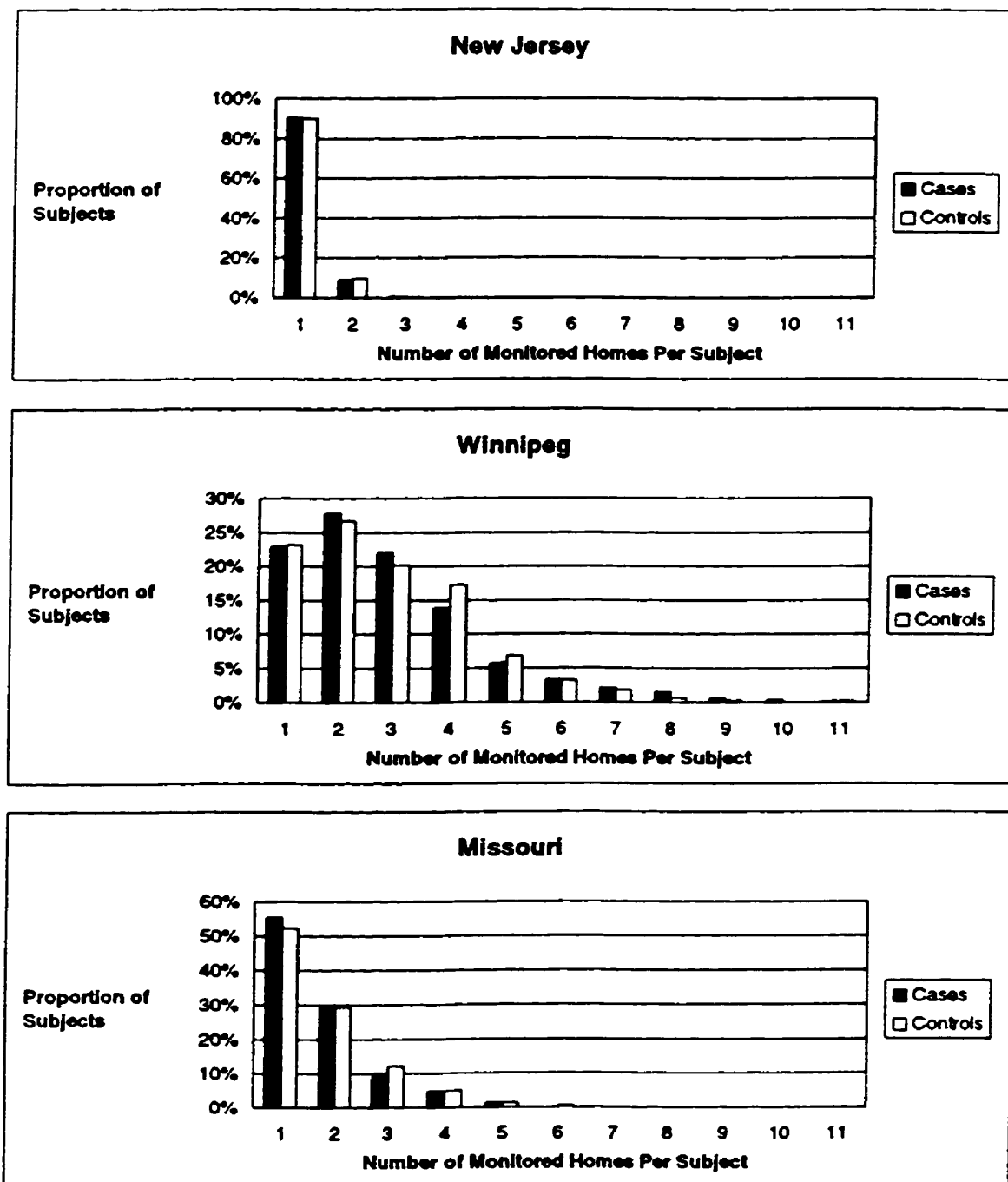


The number of homes in which radon monitoring was completed per subject is depicted by site in Figure 4.2. An average of 1.1 and 2.8 homes per subject were monitored in New Jersey and Winnipeg, respectively, among both case and control series. Missouri cases had an average of 1.1 homes monitored compared to 0.9 in controls, a trivial difference. A comparison of Figures 4.1 and 4.2 reveals a high degree of vigour in pursuit of home measurements, at least in Winnipeg where records were maintained for all identified homes.

Home monitoring estimates were derived exclusively from annual alpha track detector (ATD) monitoring directly from the living area of the respective home in Missouri. In Winnipeg and New Jersey a small proportion of measures (less than 6% in any strata formed by the combination of site, time window, and subject series) were derived by regression from a basement ATD measure in the respective home. In New Jersey an even smaller proportion of measures (less than 2% in any strata formed by time window and subject series) were obtained by short-term charcoal canister measurement. Estimation was not applicable for a fraction of the time window 25 - 50 years prior to ascertainment, for a small number of subjects who were 50 years of age or older at ascertainment (ie. in years which preceded the subjects' birth dates); such periods represented less than 3% of this time window in strata formed by site and subject series. The proportion of data for which estimation was not applicable differed between cases and controls by less than one percent.

Figure 4.2

Number of Homes in Which Radon Monitoring was Completed



4.2.2 HOME RADON LEVELS

Figure 4.3 depicts the frequency distribution of radon concentrations amongst monitored control homes by site. These distributions have a long right tail and are approximately, although not strictly, log-normal³. The central tendency and inter-quartile range (25th and 75th percentiles) are summarized in Table 4.4.

Table 4.4

Central Tendency and Inter-quartile Range (IQR) of the Distribution of Radon Concentration (in Bq/m³) Amongst Monitored Control Homes

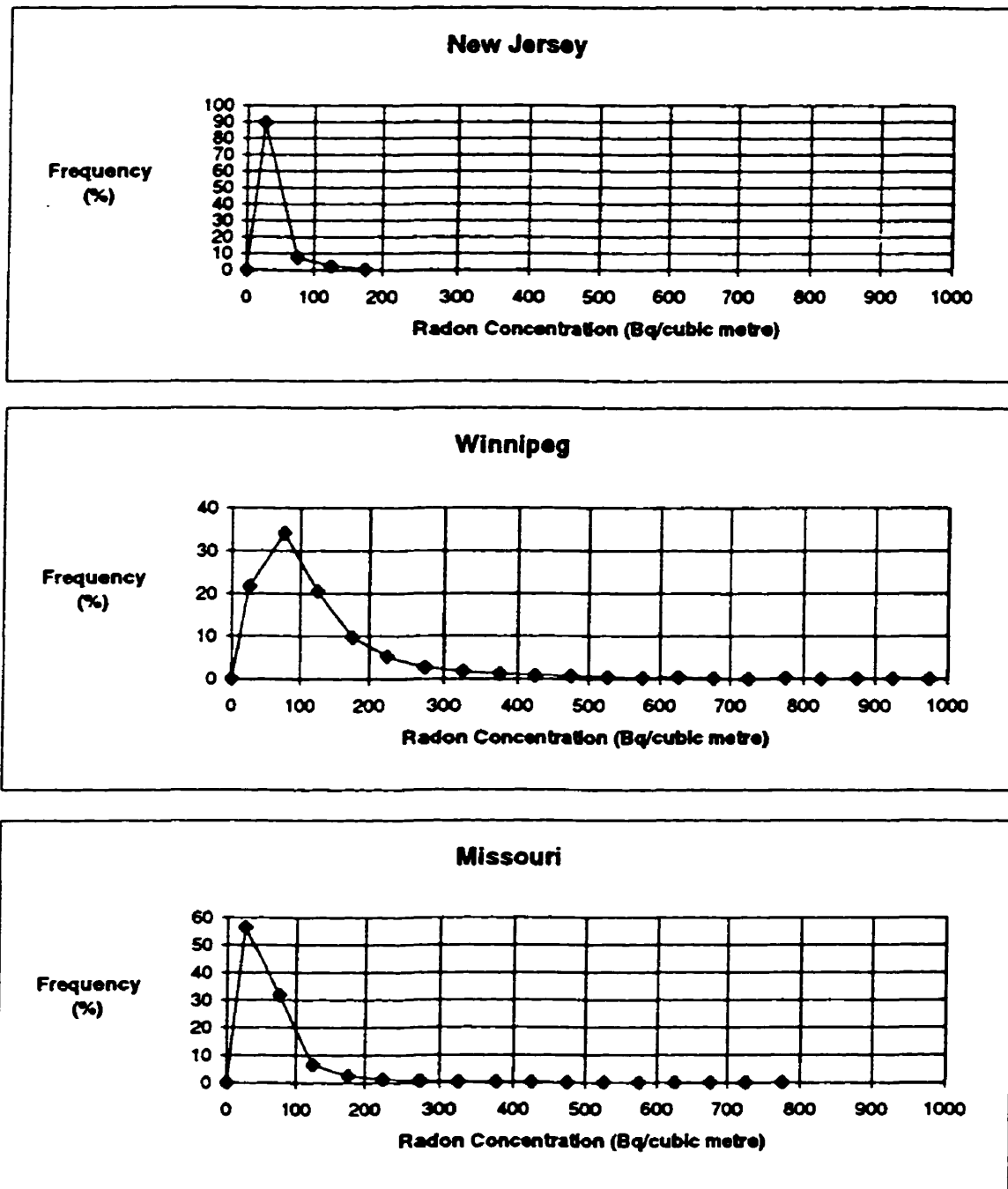
Site	Arithmetic Mean	Median	25% ile	75% ile	IQR (Difference)	Ratio of 75 th to 25 th % ile
New Jersey	27	18	11	36	25	3.3
Winnipeg	121	91	54	144	90	2.7
Missouri	60	42	28	70	42	2.5

Both the means and medians follow the same trend with site, being roughly double in Missouri what they are in New Jersey, and roughly double again in Winnipeg. Similar variability, reflected by the ratio of 75th to 25th percentiles (approximately 3), was observed in each site.

³ A log-normal distribution for the radon concentrations in single family homes in the US has been suggested both Nero et al. (1986).

Figure 4.3

**Frequency Distribution Of Radon Concentration
Among Monitored Control Homes**



4.2.3 IMPUTATION OF GAPS IN RADON EXPOSURE HISTORY

As discussed in chapter 3, the observed control mean imputation (OCMI) method was used to fill in gaps in the exposure history for subjects with incomplete home radon monitoring data in the period 5 to 50 years prior to ascertainment. Weinberg et al. (1996) suggested stratified imputation within strata formed by other available predictors of radon such as residence type, underlying geology, or window-opening behaviour of occupants, since the use of stratum-specific control means would result in greater precision in risk estimates. The only type of available data which was possibly predictive of radon levels was that of residence type (categorized in New Jersey as ranch house, split house, bi-level house, two story house, other single house, apartments, or duplex or row house). In an analysis of variance, only 4% of the variance in home radon concentrations could be explained by dwelling type, and 0.06% by dichotomization of dwelling type into houses versus apartments. This was considered insufficiently informative to justify the use of stratified OCMI. Unstratified OCMI was therefore employed within each study site.

The proportion of measures which were derived from direct ATD monitoring in the living area of the home (not including measures regressed from another floor level) was highest for the time window closest to ascertainment (varying between 75 and 85 percent in the strata formed by site and subject series, 5 - 14 years prior to ascertainment) and declined successively in more historically distant time windows (reaching lows of 25 to 36 percent in the 25 to 50 year time window). Conversely, the proportion of imputed measures increased with time prior to ascertainment. Over all subjects, approximately 17, 26, and 66 percent of the time windows 5 - 14, 15-25, and 25 - 50 years prior to ascertainment were covered by imputed radon measures.

Table 4.5 indicates the proportion of annual radon measures within each time window for which radon estimates were imputed by OCMI, as opposed to having been estimated by home

monitoring (or not applicable).

Table 4.5

**Proportion of Imputed Annual Radon Measures
Within Three Exposure Time Windows**

Site	Time Window in Years Prior to Ascertainment					
	5 - 14 Years		15 - 24 Years		25 - 50 Years	
	Cases	Controls	Cases	Controls	Cases	Controls
New Jersey	0.18	0.15	0.14	0.13	0.71	0.72
Winnipeg	0.23	0.15	0.36	0.27	0.65	0.61
Missouri	0.15	0.16	0.27	0.25	0.70	0.64

Table 4.5 demonstrates a tendency in the Winnipeg site for cases to have had a greater proportion (risk) of missing and subsequently imputed data in each time window than controls; percentage differences of 8% and 9% (corresponding to relative risks of missing data of 1.27 and 1.20) in the 5 - 14 and 15 - 24 year time windows, raised concern for information bias⁴. If on average, case radon levels were in fact higher than those of controls, a greater degree of imputation among cases, would result in bias.

⁴ A difference of 14 % in the most distant (25- 50) year time window in Missouri was considered to be of lesser importance because of the relative lack of informational value in this time window as indicated through sensitivity analyses to be discussed later; also, a part of this difference arose because of differing proportions of subjects less than 50 years of age, leading to time (1.6 % in cases, 2.8 % in controls) for which measurement was not applicable.

4.2.4 RADON EXPOSURE IN THE 5 -50 YEARS PRIOR TO ASCERTAINMENT

Figures 4.4 and 4.5 depict the estimated cumulative frequency distribution of radon among study subjects in the 5 - 50 years prior to ascertainment, expressed as (weighted) cumulative working level months (WLMs), and (weighted) average radon concentration), respectively. Note that both the cumulative WLM and concentration measures were weighted as defined in section 3.6.5 for each of the three time windows (5 - 14 years, 15 - 24, and 25 - 50 years prior to ascertainment). The representation of radon levels in these units is mainly for calibration by the reader to familiar units; studies in underground miners have employed the WLM units, whereas studies among residential populations have customarily employed average concentration in Becquerels per cubic metre (Bq/m³) of air. Since most subjects were alive, and therefore had radon estimated for a full 50 years prior to ascertainment, these units can be considered for practical purposes to differ only by a constant scaling factor, although as developed in section 3.6.5.5, this constant differs slightly for males and females because of home occupancy assumptions. Figures 4.4 and 4.5 therefore not surprisingly show similar curves.

Whether in units of cumulative exposure or average concentration, only slight differences can be observed in the radon exposure distributions for cases versus controls in the default time window. Case exposure slightly exceeded that of controls in New Jersey and Missouri; in Winnipeg, control radon slightly exceeded that of cases. The direction of the difference in Winnipeg is not attributable to the greater degree of imputation (with the annual mean radon level of controls) of missing case radon levels than controls, since radon levels among measured homes was slightly greater among controls than cases prior to imputation. Assuming that there was a real difference in exposure, imputation of the control value for a greater number of missing measures amongst cases (than among controls) would have brought the case distribution closer to that of controls resulting in an apparent odds ratio that deviates less (than in truth) from the null, regardless of the direction of the true deviation.

Figure 4.4

Cumulative Frequency Distribution of Cumulative Radon Exposure in the 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months

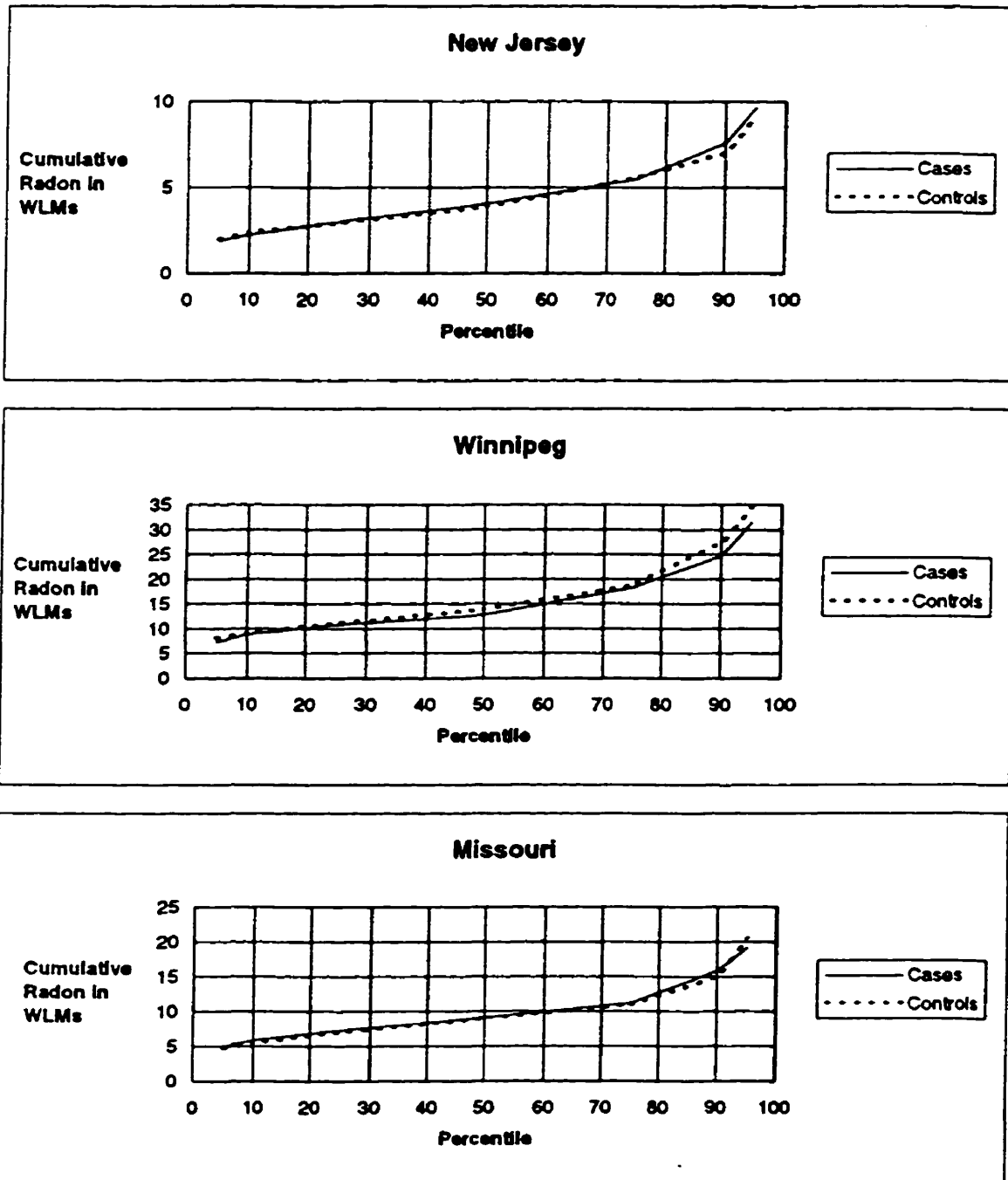
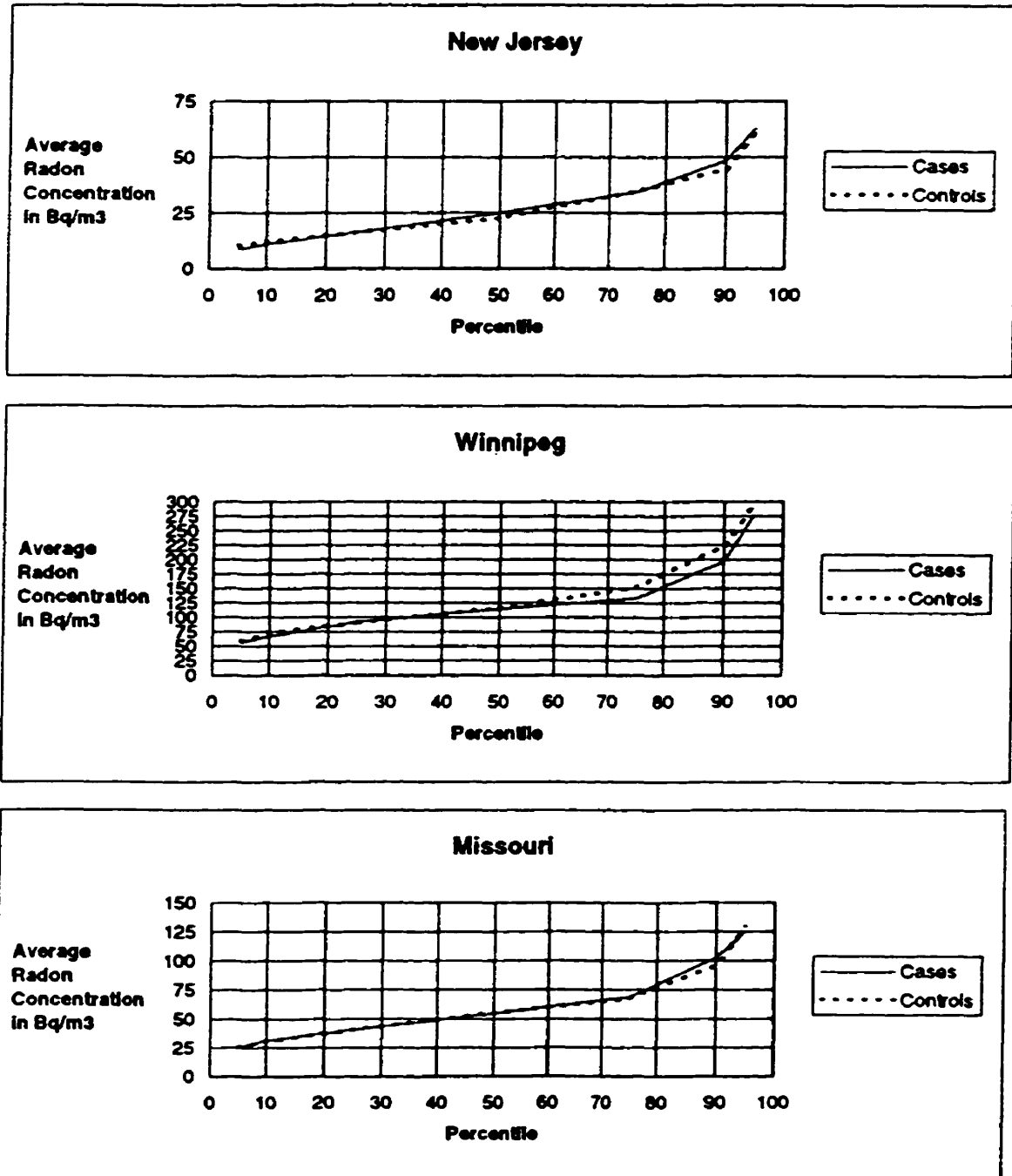


Figure 4.5

**Cumulative Frequency Distribution of Average Radon Concentration
in the 5 - 50 Years Prior to Ascertainment in Weighted Becquerels per Cubic Metre**



Although it may be tempting at this stage to conclude that the case and control radon distributions depicted in Figures 4.4 and 4.5 are so close as to rule out the possibility of detecting a radon effect in these samples, a number of considerations render such a conclusion premature: 1) the pooling of OR estimates representing differing portions of the exposure distribution will increase precision in the estimated measures of effect; 2) the absorbed radiation dose per unit of exposure is affected by the presence of tobacco smoke in the home (as discussed in section 2.4.2.10) necessitating adjustment for active and passive smoke exposure; 3) evidence from an ongoing case-control study suggests that socio-economic status may be associated with radon exposure, necessitating adjustment on this factor on which case and control subject selection appeared to be differential; 4) other risk factors may confound or modify crude differences in radon distribution between cases and controls; and, 5) consideration by default of the widest, most conservative (5 - 50 year) exposure time window may dilute the maximal effect to be observed within the most biologically relevant exposure period; sensitivity analyses will investigate two narrower time windows.

The radon distributions amongst cases across sites were analysed in an attempt to define an appropriate reference group for radon categories for use in an analysis of the combined data. Amongst cases, the minimum radon observation in Winnipeg (not shown) corresponded to the 90th percentile in New Jersey. Clearly it would be impossible to form a reference group that would be appropriate for New Jersey and which would contain any cases whatsoever in Winnipeg. To achieve (almost) 5% of cases in Winnipeg, the upper bound of the reference group would be such that it would include 97% of the New Jersey cases. All of the main analyses were therefore conducted using a continuous variable for radon exposure. In the main analyses (by logistic regression), this assumed a log-linear relationship between radon exposure and lung cancer rate ratio.

Assuming sufficient power, using a continuous variable for exposure will yield a significant

slope whether a true relationship between the logit (log-odds) and the exposure (or covariate) is linear, quadratic, other non-linear, or binary (Hosmer and Lemeshow 1989, page 89). The use of categories of exposure, although reducing the influence of outliers (Lagarde 1992, Kramer 1981), has several drawbacks: 1) loss of information (Zhao and Kolonel 1992); 2) reduced power to detect a monotonic exposure-response relationship (Lubin et al. 1988); 3) sensitivity of results to arbitrary changes in category cut-points (Lagarde 1992, Kramer 1981); and, 4) potential bias in the evaluation of synergy (Lagarde 1992).

4.3 RELATIONSHIPS AMONGST RADON EXPOSURE, OUTCOME AND COVARIATES

Plots of the median and inter-quartile range of radon exposure (in weighted WLMs) in the default (5 - 50 year) exposure time window were created by covariate level and outcome in each site for each candidate covariate. These plots provided three important types of information: 1) the degrees of association between radon exposure and the covariate levels in (the base population of) controls, necessary for a risk factor⁵ to confound an observed relationship between exposure and outcome; 2) the degree of association between the exposure and outcome status taking into account variability in exposure, and the extent to which an association (assuming one is observed) varies with covariate level (ie. 'effect' modification); and, 3) the degree to which such patterns are consistent across sites.

This exercise was conducted before and after introducing a gender-specific occupancy factor. Those plots which are noteworthy are displayed in Figures 4.6 through 4.13. The remainder (with respect to ETS, education, income, age at ascertainment, ethnicity, source of information, year of ascertainment) appear in Appendix D. Those factors unrelated to radon

⁵ The degree of association between covariate levels and outcome status was represented in Tables 4.1 to 4.3 above.

Figure 4.6

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Smoking Recency and Outcome**

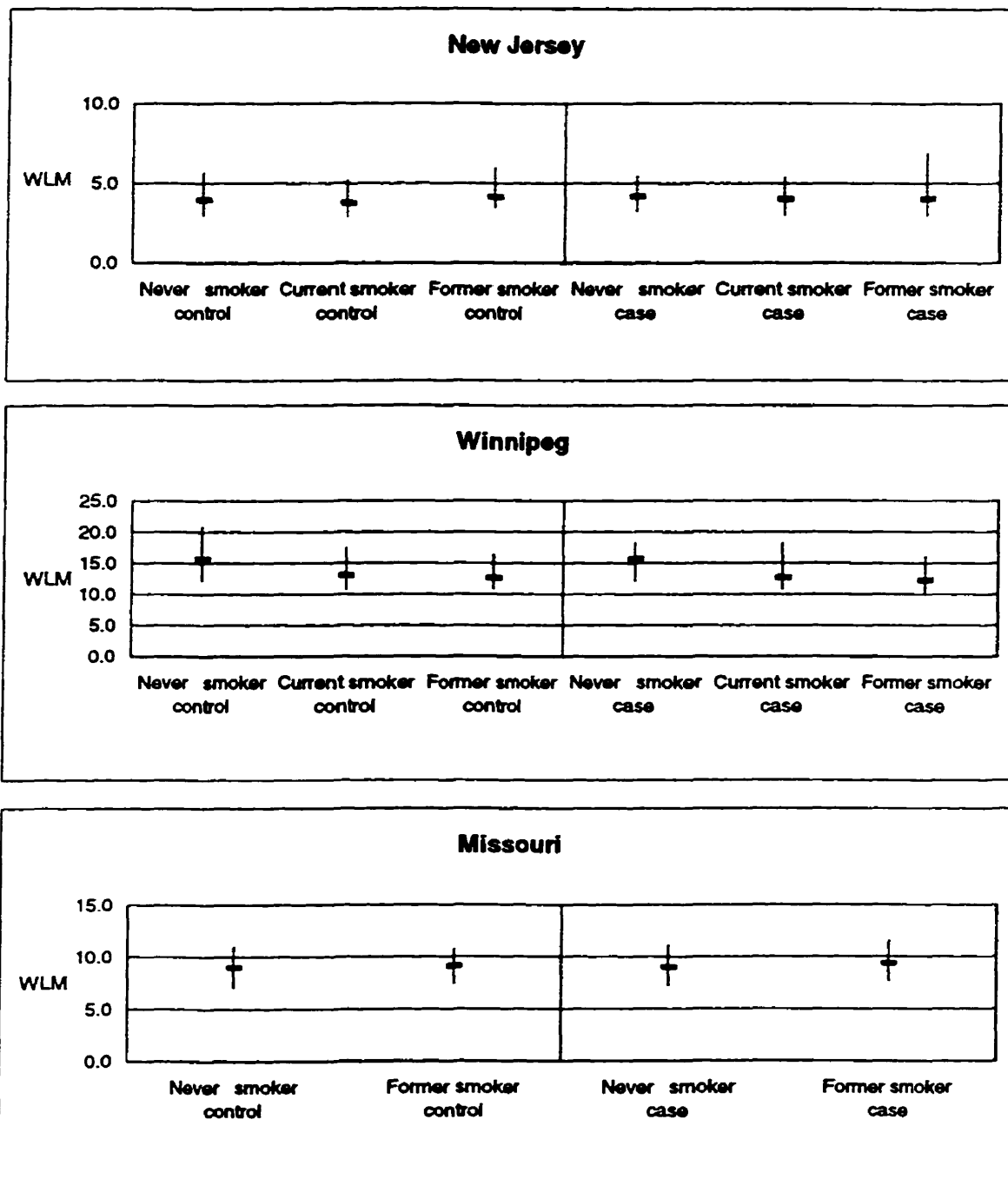


Figure 4.7

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Smoking Intensity (in Packs Per Day) and Outcome**

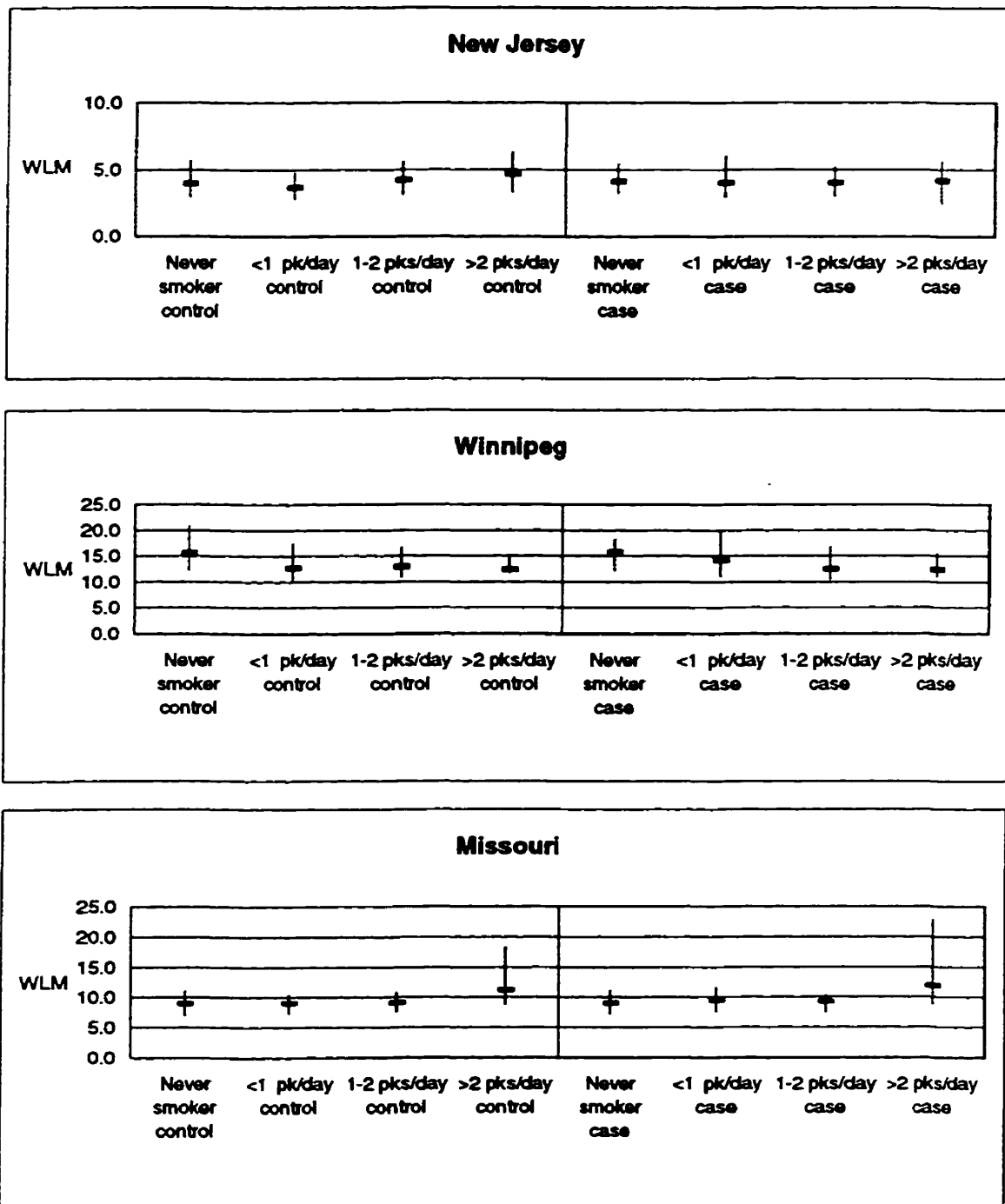


Figure 4.8

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Pack-Years of Smoking and Outcome**

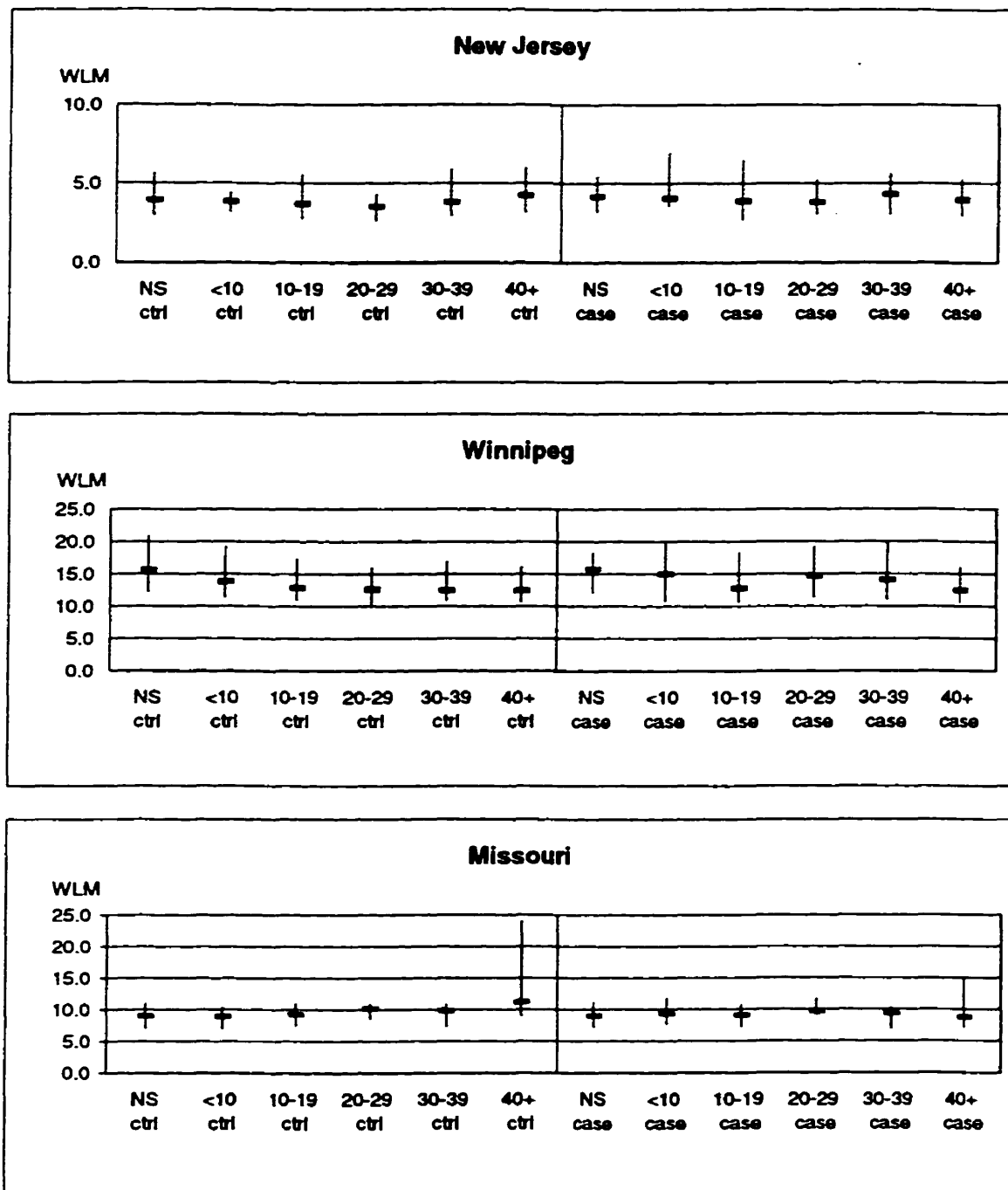


Figure 4.9

Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Smoking Recency and Outcome Assuming Non-Gender-Differential Home Occupancy

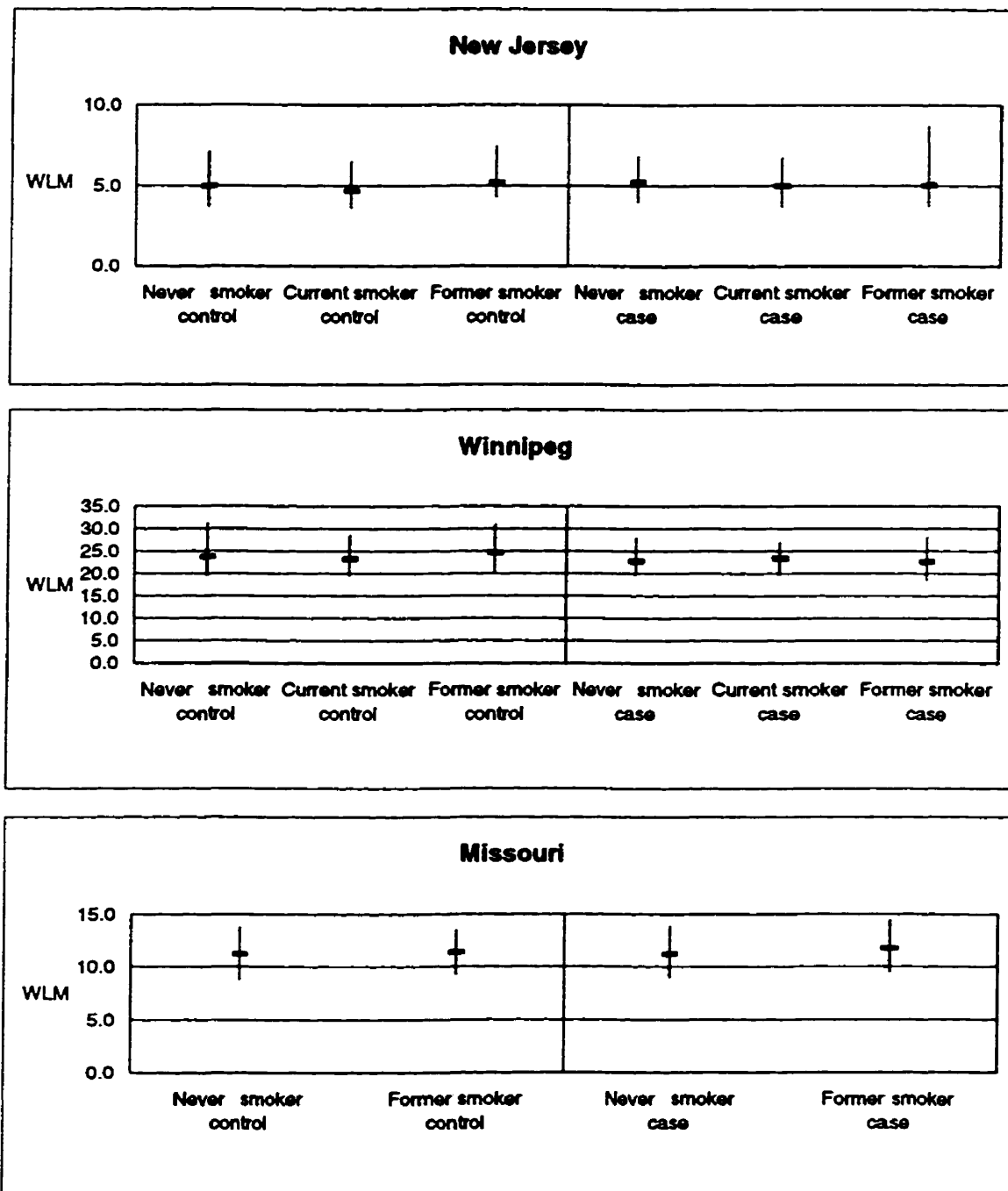


Figure 4.10

Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Smoking Intensity (in Packs Per Day) and Outcome Assuming Non-Gender-Differential Home Occupancy

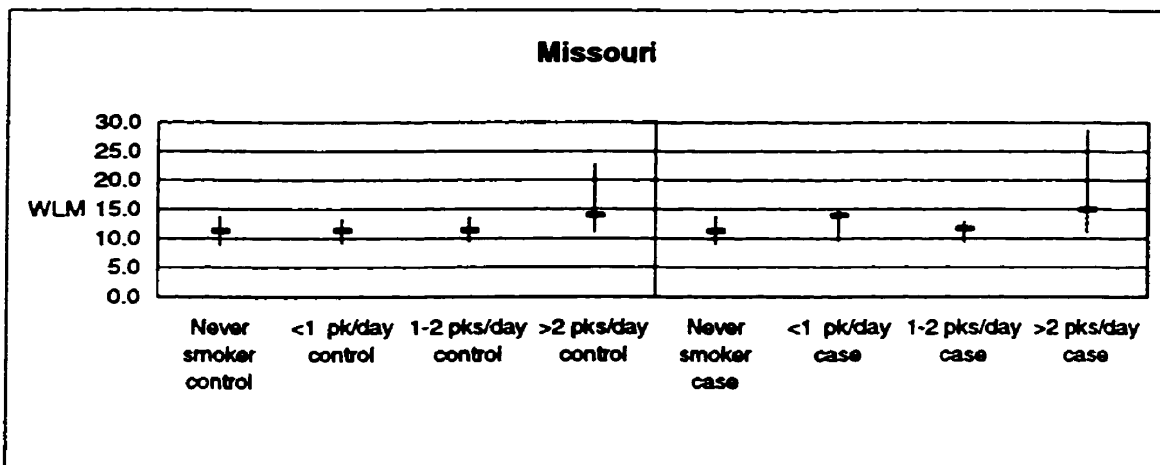
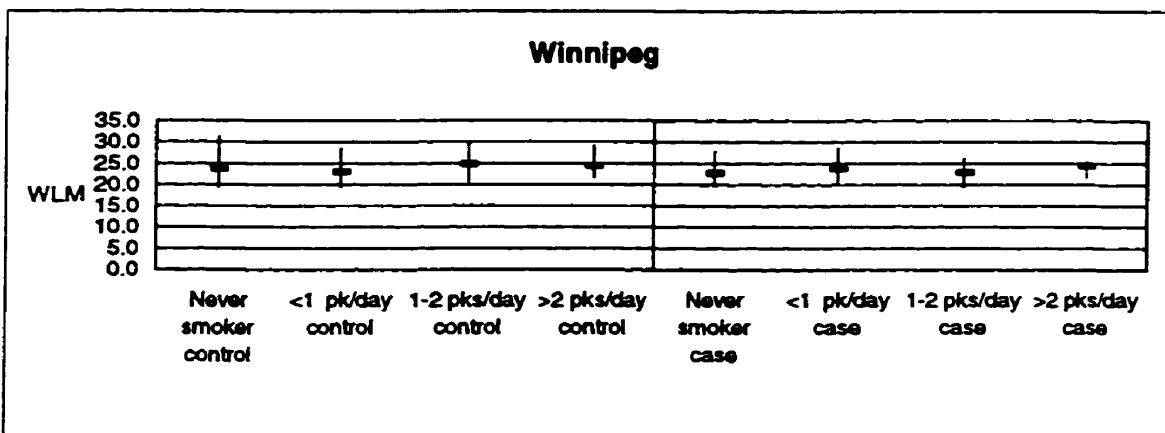
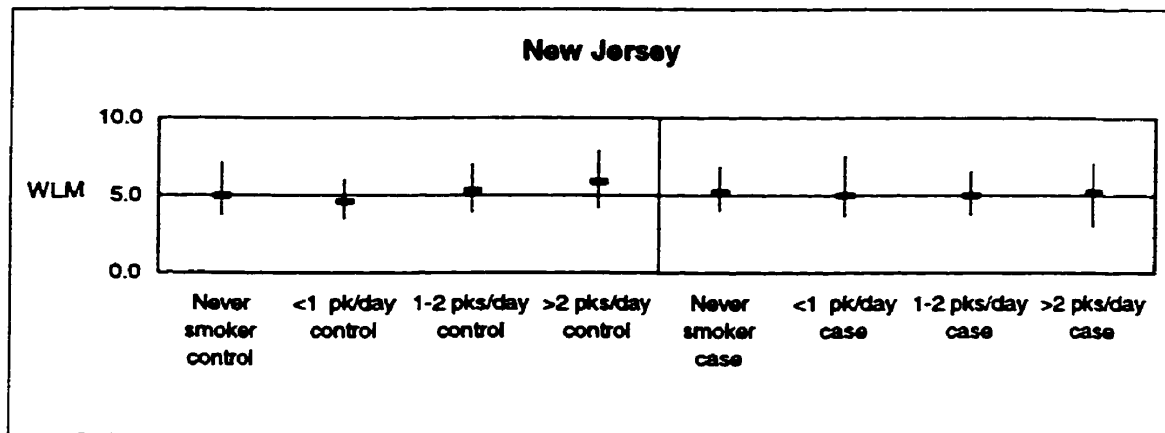


Figure 4.11

Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Pack-Years of Smoking and Outcome Assuming Non-Gender-Differential Home Occupancy

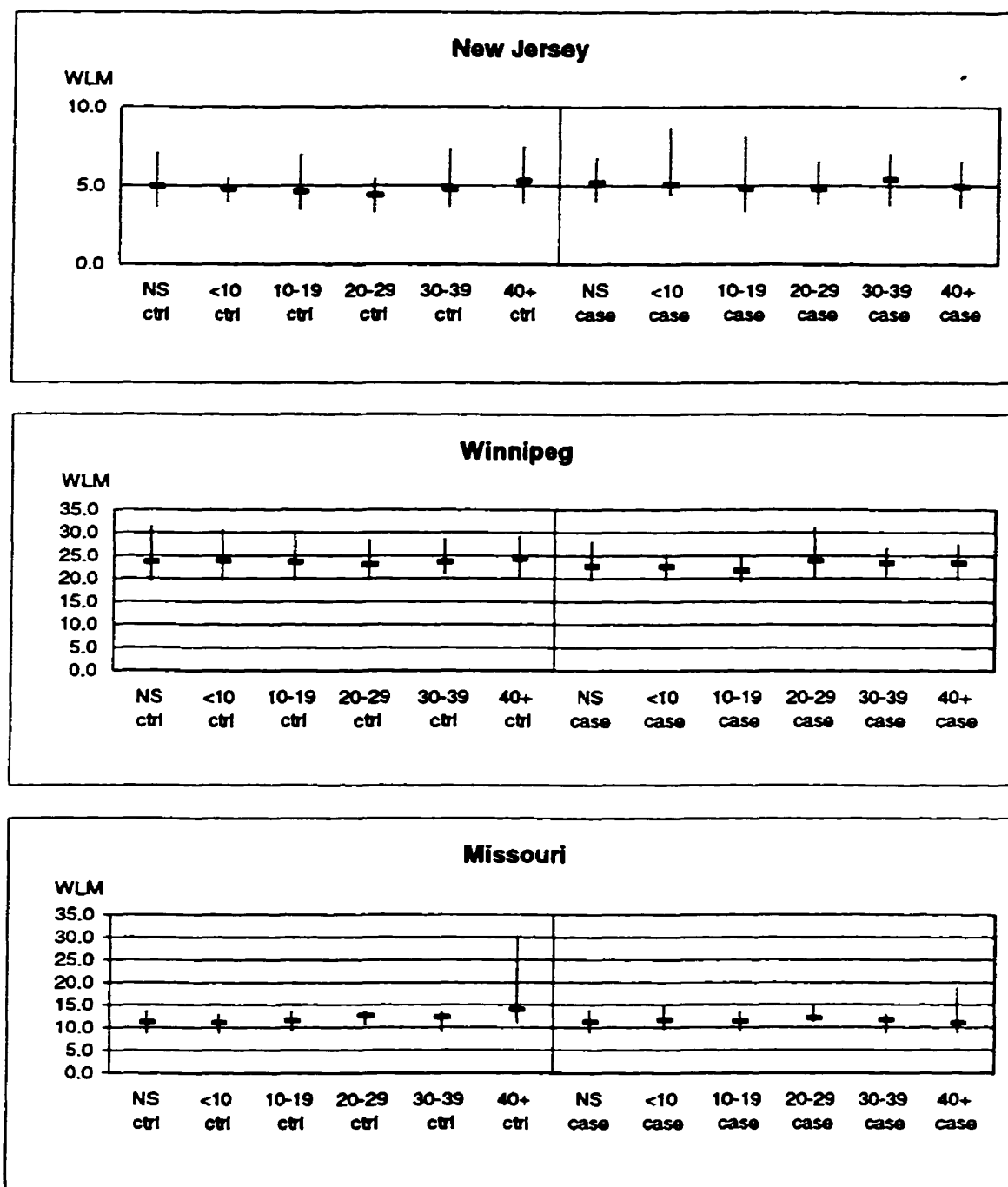


Figure 4.12

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Gender and Outcome**

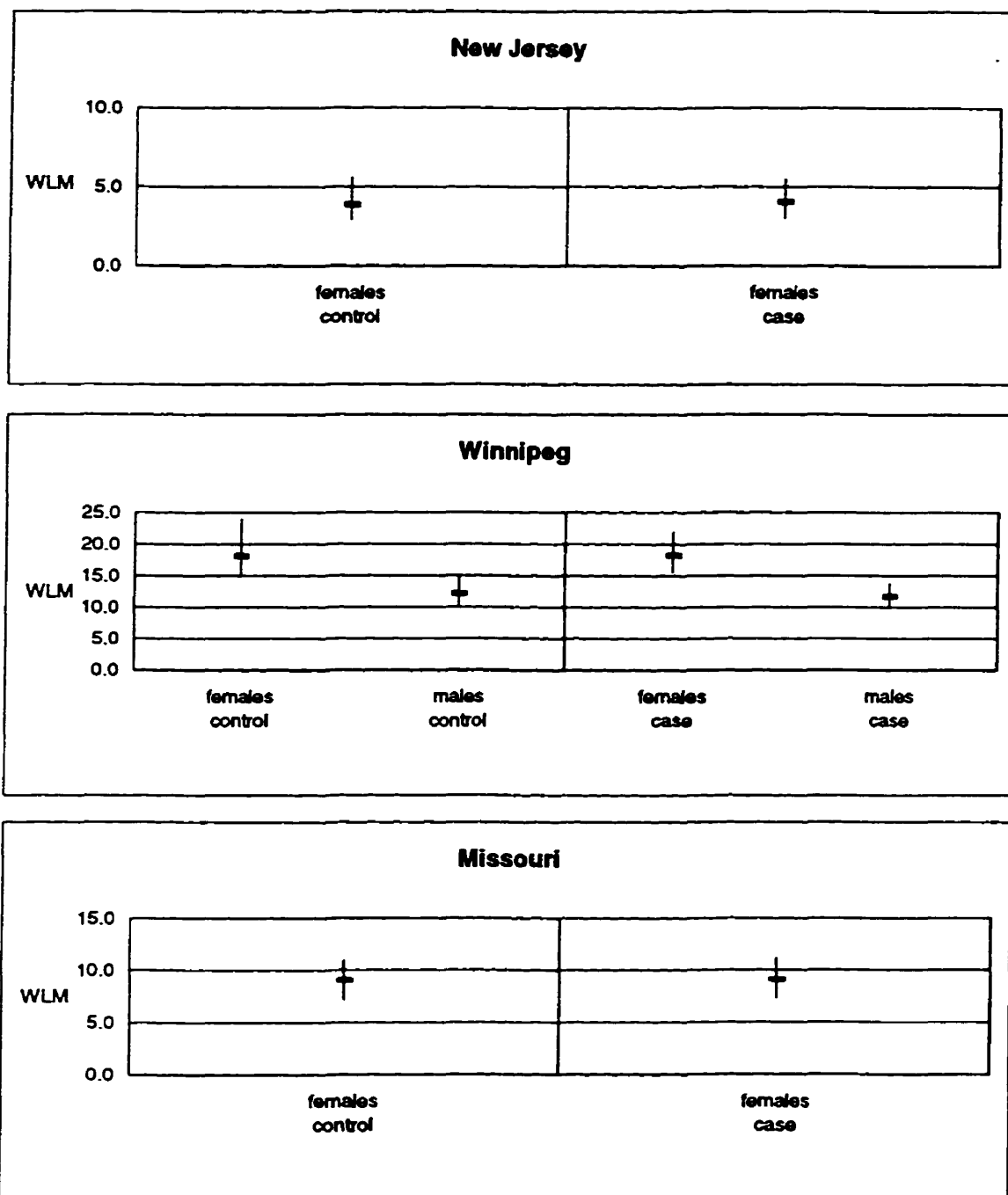
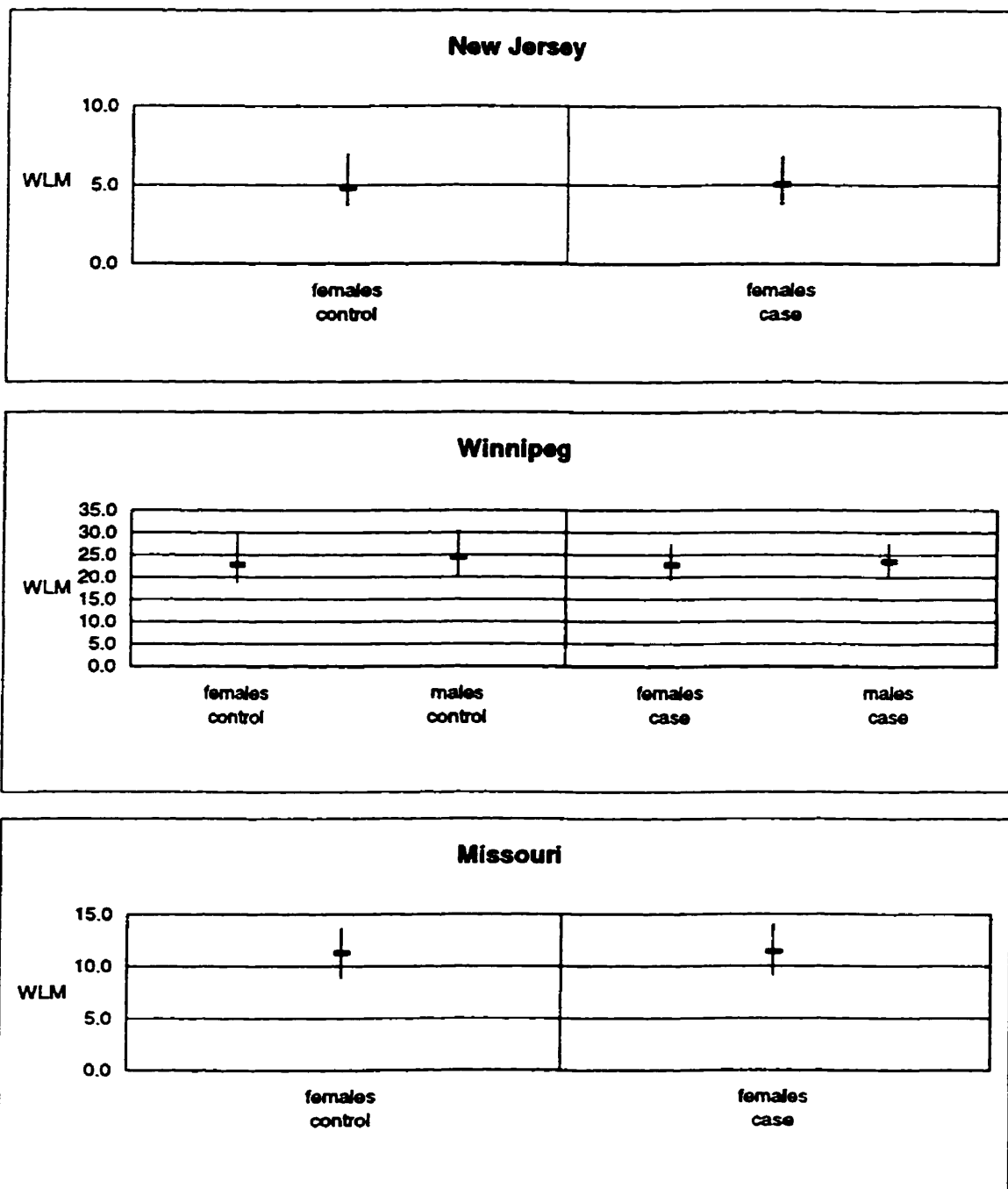


Figure 4.13

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Gender and Outcome Assuming Non-Gender-Differential Home Occupancy**



exposure can be ruled out as potential confounders, but not necessarily as effect modifiers.

Although stratum-specific effects were not discernable from plots of the relationships amongst radon exposure, outcome and covariates, interaction terms for those potential effect modifiers suspected *a priori* were fit during model development.

4.3.1 ACTIVE AND PASSIVE TOBACCO SMOKE EXPOSURE

Figures 4.6 to 4.8 provide radon exposure levels (in WLMs cumulated over the 5 - 50 years prior to ascertainment) in relation to active smoking in terms its recency⁶, intensity⁷, and combined intensity and duration⁸. Considering the variability in radon exposure as indicated by the inter-quartile range bars, there does not appear to be an association between radon exposure and active smoking, except possibly in Winnipeg, where smokers (regardless of their sub-classification) tended to have marginally lower radon levels than never-smokers. By comparing these plots to similar ones (Figures 4.9 to 4.11) created before a gender-specific occupancy factor was introduced, it seems apparent that the slight association between radon and smoking in Winnipeg is through the association of smoking with gender. Exposure to environmental tobacco smoke (Appendix D, Figure D.1) is similarly not associated with exposure to radon gas.

⁶ For this analysis, recency of smoking was considered as a categorical variable including never-cigarette-smokers, current smokers, and former smokers (ie. those who had ceased smoking at least 14 years prior to ascertainment).

⁷ Categories of smoking intensity included never-cigarette-smokers and levels of <1, 1 - 2 or > 2 packs of cigarettes smoked per day (on average) during the period of smoking within the time window 5 - 50 years prior to ascertainment.

⁸ Combined intensity and duration of smoking was represented in pack-years, calculated as the sum (over the years the subject smoked within the time window 5 - 50 years prior to ascertainment) of the annual average number of packs of cigarettes smoked per day.

4.3.2 GENDER

Figures 4.12 and 4.13 provide gender-specific information. In Figure 4.12, cumulative radon exposure (in WLM) is calculated using gender-specific occupancy factors. Figure 4.13 shows radon exposure in WLM assuming 100% occupancy by both males (only represented in Winnipeg) and females over their 5 - 50 year exposure histories. The difference in cumulative radon between Winnipeg males and females apparent in Figure 4.12 is attributable to the assumed gender-specific occupancy factors, since no such difference seen in Figure 4.13 under the assumption of equivalent (100%) home occupancy. No difference was observed between genders in measured home radon concentrations.

All ensuing analyses assumed an occupancy proportion of 80% for females and 50% for males. Since the effect of gender was controlled in the analyses, however, the uncertainty implicit in the former assumption has impact only on the external validity of the scaling of the radon effect in units of working level months.

4.4 BASIC INFORMATION RELATING ACTIVE SMOKING AND LUNG CANCER

Table 4.6 summarizes information related to active smoking amongst ever-cigarette smokers in the three pilot study sites. As expected, in every site, those cases who had reported ever-smoking tended to have initiated smoking at an earlier age (although differences and variability were small), and smoked for a greater duration and at a higher intensity than ever-smoking controls. The observed differences highlight the importance of accounting fully for variation in lung cancer risk arising from exposure to cigarette smoking, prior to modelling any potentially additional risk associated with radon.

Table 4.6

Characterization of Active Smoking Amongst Cigarette-Smoking Subjects

Site	Outcome Status	Total Number of Subjects	Number (%) of Ever-Smokers	Age at Smoking Initiation			
				Mean	Std Dev	Minimum	Maximum
New Jersey ¹	Case	351	302 (86.0)	20.2	6.8	10.0	51.0
	Control	308	144 (46.8)	20.8	6.8	13.0	49.0
Winnipeg ²	Case	701	676 (93.4)	18.8	7.0	7.0	64.0
	Control	724	482 (68.8)	19.4	6.8	7.0	65.0
Missouri ³	Case	538	142 (24.5)	22.1	7.1	13.0	53.0
	Control	1183	132 (11.2)	22.9	7.0	14.0	45.0

Site	Outcome Status	Total Number of Subjects	Number (%) of Ever-Smokers	Duration of Cigarette Smoking (Years) ⁴			
				Mean	Std Dev	Minimum	Maximum
New Jersey ¹	Case	351	302 (86.0)	40.3	10.5	1.0	63.0
	Control	308	144 (46.8)	34.0	14.0	1.0	60.0
Winnipeg ²	Case	701	676 (93.4)	40.4	11.4	1.0	63.0
	Control	724	482 (68.8)	31.0	15.1	1.0	63.0
Missouri ³	Case	538	142 (24.5)	22.9	11.6	1.0	47.0
	Control	1183	132 (11.2)	18.9	11.0	1.0	46.0

Site	Outcome Status	Total Number of Subjects	Number (%) of Ever-Smokers	Pack-Years Of Cigarette Smoking 5 - 50 Years Prior to Ascertainment ⁵			
				Mean	Std Dev	Minimum	Maximum
New Jersey ¹	Case	351	302 (86.0)	39.6	21.3	0.0	144.9
	Control	308	144 (46.8)	29.1	22.6	0.0	151.7
Winnipeg ^{2,5}	Case	701	676 (93.4)	42.5	23.8	0.2	210.0
	Control	724	482 (68.8)	27.8	19.5	0.0	106.2
Missouri ^{3,6}	Case	538	142 (24.5)	22.8	20.1	0.2	102.0
	Control	1183	132 (11.2)	15.2	17.6	0.0	132.0

- 1 For one New Jersey subject, missing values for age at smoking initiation, age at smoking cessation, duration of smoking, years since cessation, and average number of cigarettes smoked per day, were replaced by the respective control mean value; for one New Jersey subject a questionable value of age at smoking initiation was replaced by the control mean value.
- 2 Questionable values (<7) of age at smoking initiation for 17 Winnipeg subjects were replaced with the observed control mean specific to the smoking-history category of the subject.
- 3 For 21 Missouri subjects with missing age at smoking initiation, the observed control mean was imputed.
- 4 Up to year of ascertainment (diagnosis for cases, interview for controls); adjusted for periods of cessation except in Winnipeg where calculated as age at smoking cessation (or age at ascertainment, where cessation age was missing) minus age at smoking initiation.

- 5 An inconsistent value (5) of age of quitting smoking age for one Winnipeg subject who reported smoking cessation at age 20 was replaced with the observed control mean specific to the smoking-history category of the subject. A missing value for one Winnipeg subject of the average number of cigarettes smoked per day was replaced with the observed control mean specific to the smoking-history category of the subject.
- 6 For 2 Missouri subjects (1 of which also had missing start) with missing age of quitting smoking, values were replaced by the control mean. The imputation of missing cessation age did not result in any inconsistency in years since cessation with study entrance criteria which required that subjects had quit smoking a minimum of 15 years prior to interview. For 16 Missouri subjects with missing average number of cigarettes smoked per day, values were replaced by the control mean.
- ☒ Cigarette smokers in Winnipeg exclude 1 case and 23 controls who smoked only pipes or cigars.
- ☐ Inclusion criteria restricted the Missouri study to women who had not smoked in the 15 years prior to diagnosis (cases) or interview (controls).
- ✧ Note that in the analysis, only smoking in the 5 to 50 years prior to ascertainment was considered; the pack-years variable calculated here reflects this restricted time window; a total of 11 cigarette-smoking subjects (1 case and 1 control in New Jersey, 6 controls in Winnipeg, and 3 controls in Missouri) quit smoking more than 50 years prior to ascertainment, and hence had a value of 0 for pack-years within the considered time window.

Using the data from the third and fourth columns of Table 4.6 one may observe comparable crude relative risks (RR) of lung cancer associated with ever-smoking of 2.94 and 2.84 in New Jersey and Winnipeg, respectively. One may also calculate the population attributable risk (PAR), or etiologic fraction⁹ attributable to ever-smoking, as 74% among New Jersey females, and 87% among the mostly (66%) male Winnipeg subjects. In spite of the fact that population attributable risk varies with the prevalence of the exposure, here to tobacco smoking, in the population, even these are crude estimates (unadjusted for other risk factors), are comparable to those provided by a large (1.2 million person) prospective study by the American Cancer Society (Shopland et al., 1991, as cited by Mason, 1994) of 78% and 90% among females and males, respectively. In Montreal, where the prevalence of smoking exceeds that elsewhere in North America, an adjusted estimate of 90% was observed among males in a large case-control study by Siemiatycki et al., (1995).

A lesser RR of 1.46 in Missouri accords with the fact that the RR associated with smoking declines with time since quitting, and all of the Missouri subjects had ceased smoking a minimum of 15 years prior to interview. Only 15% of lung cancer cases in the non-smoking

⁹ $EF = P_c (RR - 1) / RR$ where EF is the etiologic fraction (proportion of cases attributable to the exposure, here smoking), P_c is the proportion of ever-smoking cases and RR (relative risk, an estimate of the rate ratio) is calculated as usual from a two by two table (Rothman 1986).

target population of Missouri were attributable to previous cigarette smoking. A substantially greater proportion of lung cancer risk is therefore attributable to factors other than active smoking (such as radon) in Missouri than in the other two sites where smoking was the dominant etiologic factor. The Missouri study therefore provides a potentially more informative study population with respect to radon.

Data from a case-control¹⁰ study conducted by Siemiatycki et al. (1994 and 1995) in Montreal Canada during a time frame close to that of the radon studies was used to further validate the smoking information in the radon studies. For the present radon studies, Table 4.7 provides estimated odds ratios for lung cancer associated with a cumulative index of smoking intensity and duration analogous to that employed by Siemiatycki et al. (here as six pack-year categories) with nonsmokers as the reference group, and adjusting for the same factors (namely age, socio-economic status, and ethnicity). In the present analyses, two covariates not controlled for by the previous authors were included in the logistic models in sites where they were matching factors: source of information (included in the New Jersey model) and gender (included in the Winnipeg model).

The data from Table 4.7 is also plotted in Figure 4.14. Within each site, an exposure-response trend is evident with the odds ratio of lung cancer increasing with pack-years of smoking. Only a slight deviance from this trend is observed (in the second highest smoking category in New Jersey), though the confidence interval for this data point is wide. The slope of the exposure-response trend differs between sites, however, being greatest in Winnipeg and smallest in Missouri. The confidence intervals estimated in the three sites overlap only within the lowest two of five pack-year categories; overlap amongst sites for the upper three categories is limited to two out of three sites. Restriction to live subjects in all sites did not materially affect these results. Adjustment for radon exposure similarly had no effect on the

¹⁰ Siemiatycki et al.'s study included 857 cases of lung cancer diagnosed between 1979 and 1985.

Table 4.7

Adjusted¹ Odds Ratios for Lung Cancer By Pack-Years^o of Cigarette Smoking

SITE	Never-Smokers		< 10 Pack-Years		10 - 19 Pack-Years		20 - 29 Pack-Years		30 - 39 Pack-Years		≥ 40 Pack-Years	
	Cases	OR	Cases	OR	Cases	OR	Cases	OR	Cases	OR	Cases	OR
	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)
NJ ¹	49	1	14	2	37	6.1	47	7	79	19.6	125	13.5
	164		29	(0.9 - 4.2)	27	(3.2 - 11.6)	29	(3.8 - 12.9)	19	(10.2 - 37.3)	40	(7.9 - 23.1)
WPG ²	25	1	35	4.7	60	8.3	105	16	139	25.8	337	42.6
	242		97	(2.6 - 8.5)	93	(4.8 - 14.3)	90	(9.4 - 27.3)	74	(15.1 - 43.8)	128	(25.6 - 70.8)
MISS ³	396	1	43	1.9	33	2.9	26	4	18	4.5	22	9.2
	1051		64	(1.3 - 2.9)	32	(1.7 - 4.7)	18	(2.2 - 7.4)	11	(2.1 - 9.7)	7	(3.9 - 21.9)

o Pack-years is the product of average number of packs per day and years of smoking. The reference group is never-cigarette smokers: In New Jersey this (includes 2 subjects, 1 case and 1 control who reported pipe or cigar smoking; In Winnipeg this includes 24 subjects - 1 case and 23 controls - who reported pipe or cigar smoking; pack-years is the product of average number of packs per day and years of smoking. In Missouri this includes 2 controls who reported pipe or cigar smoking, and 87 subjects - 19 cases and 68 controls - who reported smoking < 100 cigarettes or < 1 year.

1 Odds ratios were adjusted for age (indicator variables), education (indicator variables), ethnicity (dichotomous; New Jersey and Winnipeg), source of information (New Jersey only), and gender (Winnipeg only).

1 The New Jersey study was restricted to females. Living control subjects were frequency matched to living cases by age at ascertainment and 'race'; deceased cases were individually matched to controls by race, age, and closest date of death; proxy respondents were utilized for the latter cases and controls.

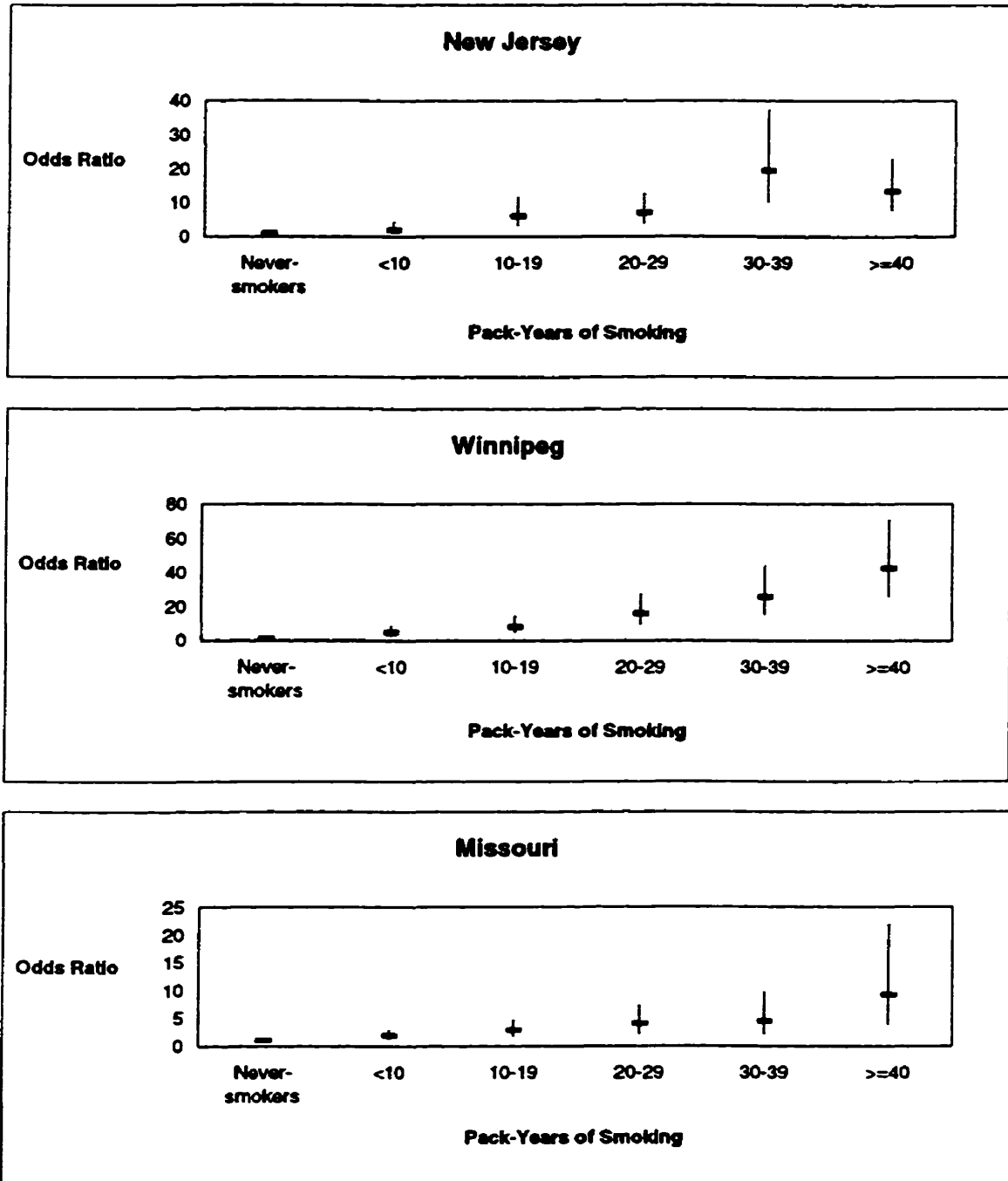
2 The Winnipeg study included both males (66%) and females (34%); Controls were frequency matched to cases by age at ascertainment and gender.

3 The Missouri study was restricted to females of 'white' race. Controls were frequency matched to cases by age at ascertainment (defined as diagnosis for cases and interview for controls). This study was restricted to females who had not smoked in the 15 years prior to interview (corresponding to a minimum of 13 years prior to ascertainment).

Ctrls Controls

Figure 4.14

Adjusted Odds Ratios [95% CI] Associated with Pack-Years of Smoking



differing magnitude of effect of smoking measured in combined units of intensity and duration. The reason for the observed between-site differences are discussed following comparison of the smoking odds ratios from the radon studies to that of Siemiatycki et al. (1994 and 1995).

The results of Siemiatycki et al. (1994 and 1995) are summarized in Table 4.8 for comparison to the data presented in Table 4.7 and Figure 4.14 from the radon studies. The reference group is never-smokers, as before.

Table 4.8
Summary of Odds Ratios for Lung Cancer By Pack-Years of Smoking
Estimated by Siemiatycki et al., (1994 and 1995)

Pack-Years of Smoking	1994 OR (95% CI)^a	1995 OR (95% CI)^b
< 25	5.4 (2.7 - 10.8)	4.4 (2.2 - 8.6)
25 - 50	12.8 (6.6 - 24.7)	9.9 (5.3 - 18.7)
51 - 75	23.0 (11.8 - 45.0)	16.1 (8.5 - 30.8)
> 75	42.2 (21.4 - 83.2)	28.0 (14.5 - 54.0)

a Results (population controls) adjusted for age, SES, ethnicity, and occupational exposures (not significant).

b Results (population controls) adjusted for age, SES, and ethnicity.

The pack-year category of 25 - 50 years with a midpoint of 37.5 most closely approximates that employed in the present analysis of 30 - 40 pack-years with midpoint equivalent to 35. In spite of the between site variability in smoking dose-response between the radon studies, the point estimates generated by the each of the radon studies fit within the confidence intervals estimated by the Montreal cancer study. The Montreal cancer study thus externally validates the radon study smoking data.

I next investigated the variability in the pack-year category-specific odds ratios between

radon study sites. Since the Missouri study was restricted to never or ex-smokers, the cumulative (pack-years) smoking exposure in this population was accumulated more than 13 years prior to ascertainment. Thus, in order to obtain odds ratios pertaining to truly comparable doses of tobacco exposure, the subjects in the other two sites would have to be similarly restricted, ie. to the eligibility criteria in the Missouri study -- to female, ethnically similar subjects (within country), who had not smoked within the 13 years immediately prior to ascertainment. Unfortunately, such a restriction left only 5 cases in Winnipeg's reference group of never-cigarette-smokers, and too few cases in the pack-year categories in both Winnipeg and New Jersey to obtain stable odds ratio estimates.

Restriction to non-recent smokers, arguably most important criteria, was, however possible. These results are provided in Table 4.9. Note the very small numbers of subjects remaining in New Jersey, and the correspondingly unstable odds ratios and large confidence intervals. Among those who had not smoked within 13 years of ascertainment, the confidence intervals for lung cancer odds ratio associated with pack-years of smoking are consistent between site, overlapping at all pack-year category levels. Controlling for recency of smoking in this way resulted in comparable dose-response relationships for combined smoking duration and intensity and lung cancer odds.

Table 4.9

**Adjusted¹ Odds Ratios for Lung Cancer By Pack-Years^o of Cigarette Smoking
Among Subjects Who Did Not Smoke In the 13 Years Prior to Ascertainment**

SITE	Never-Smokers		< 10 Pack-Years		10 - 19 Pack-Years		20 - 29 Pack-Years		30 - 39 Pack-Years		≥ 40 Pack-Years	
	Cases	OR	Cases	OR	Cases	OR	Cases	OR	Cases	OR	Cases	OR
	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)
NJ ¹	49	1	7	1.7	1	1	7	18.6	2	9.9	1	1.8
	164		19	(0.6 - 4.7)	7	(0.1 - 8.9)	2	(3.4 - 103.2)	11	(0.8 - 118.8)	2	(0.2 - 21.7)
WPG ²	25	1	17	3.8	8	3.2	12	6.8	15	15.1	8	8.5
	242		76	(1.8 - 7.9)	40	(1.2 - 8.7)	35	(2.7 - 17.2)	18	(5.7 - 40.2)	17	(2.9 - 25.3)
MISS ³	396	1	43	1.9	33	2.9	26	4	18	4.5	22	9.2
	1051		64	(1.3 - 2.9)	32	(1.7 - 4.7)	18	(2.2 - 7.4)	11	(2.1 - 9.7)	7	(3.9 - 21.9)

◊ Pack-years is the product of average number of packs per day and years of smoking. The reference group is never-cigarette smokers; In New Jersey this (includes 2 subjects, 1 case and 1 control who reported pipe or cigar smoking; In Winnipeg this includes 24 subjects - 1 case and 23 controls - who reported pipe or cigar smoking; pack-years is the product of average number of packs per day and years of smoking. In Missouri this includes 2 controls who reported pipe or cigar smoking, and 87 subjects - 19 cases and 68 controls - who reported smoking < 100 cigarettes or < 1 year.

† Odds ratios were adjusted for age (indicator variables), education (indicator variables), ethnicity (dichotomous; New Jersey and Winnipeg), source of information (New Jersey only), and gender (Winnipeg only).

1 The New Jersey study was restricted to females. Living control subjects were frequency matched to living cases by age at ascertainment and 'race'; deceased cases were individually matched to controls by race, age, and closest date of death; proxy respondents were utilized for the latter cases and controls.

2 The Winnipeg study included both males (66%) and females (34%); Controls were frequency matched to cases by age at ascertainment and gender.

3 The Missouri study was restricted to females of 'white' race. Controls were frequency matched to cases by age at ascertainment (defined as diagnosis for cases and interview for controls). This study was restricted to females who had not smoked in the 15 years prior to interview (corresponding to a minimum of 13 years prior to ascertainment).

Numerous prospective and retrospective studies have demonstrated a decline in lung cancer rate ratios for smoking with increasing time since smoking cessation (see Tables 3 and 4 of DHHS 1990). A set of analyses was undertaken to determine whether recency of smoking, (where subjects who had smoked within 13 years of ascertainment were categorized as 'current smokers', and compared with those quitting at varying time intervals) might offer a comparable index of smoking effects across site. As detailed in Table 4.10 and depicted in Figure 4.15, classification of smokers by recency of smoking provided highly comparable lung cancer odds ratios between site, as evidenced by the large overlap of confidence intervals among all sites within categories. The point estimate of effect of current versus never smoking remained much higher (almost double) in Winnipeg than in New Jersey, however. To control for this discrepancy, site-specific terms for recency of tobacco exposure were fit during analysis of the pooled data.

Table 4.10

Adjusted¹ Odds Ratios for Lung Cancer By Recency^o of Cigarette Smoking

SITE	Never-Smokers		Quit Smoking ≥ 36 Years Prior to Ascertainment		Quit Smoking 23 - 35 Years Prior to Ascertainment		Quit Smoking 14 - 22 Years Prior to Ascertainment		Quit Smoking ≤ 13 Years Prior to Ascertainment	
	Cases	OR	Cases	OR	Cases	OR	Cases	OR	Cases	OR
	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)	Ctrls	(95% CI)
NJ ¹	49	1	26	1.3	5	4.3	11	2.4	284	11.2
	164		6	(0.2 - 6.7)	5	(1.2 - 16.2)	20	(1.0 - 5.5)	113	(7.2 - 17.5)
WPG ²	25	1	434	1.3	26	4.1	30	5.3	616	23.3
	242		34	(0.4 - 4.0)	80	(2.2 - 7.8)	72	(2.9 - 9.9)	296	(14.8 - 36.7)
MISS ³	396	1	43	2	33	3	26	3.6	0	--
	1051		27	(1.1 - 3.7)	49	(2.0 - 4.5)	56	(2.5 - 5.2)	0	

◇ The reference group is never-cigarette smokers: In New Jersey this (includes 2 subjects, 1 case and 1 control who reported pipe or cigar smoking; In Winnipeg this includes 25 subjects - 2 cases and 23 controls - who reported pipe or cigar smoking; pack-years is the product of average number of packs per day and years of smoking. In Missouri this includes 2 controls who reported pipe or cigar smoking, and 87 subjects - 19 cases and 68 controls - who reported smoking < 100 cigarettes or < 1 year.

† Odds ratios were adjusted for age (indicator variables), education (indicator variables), ethnicity (dichotomous; New Jersey and Winnipeg), source of information (New Jersey only), and gender (Winnipeg only).

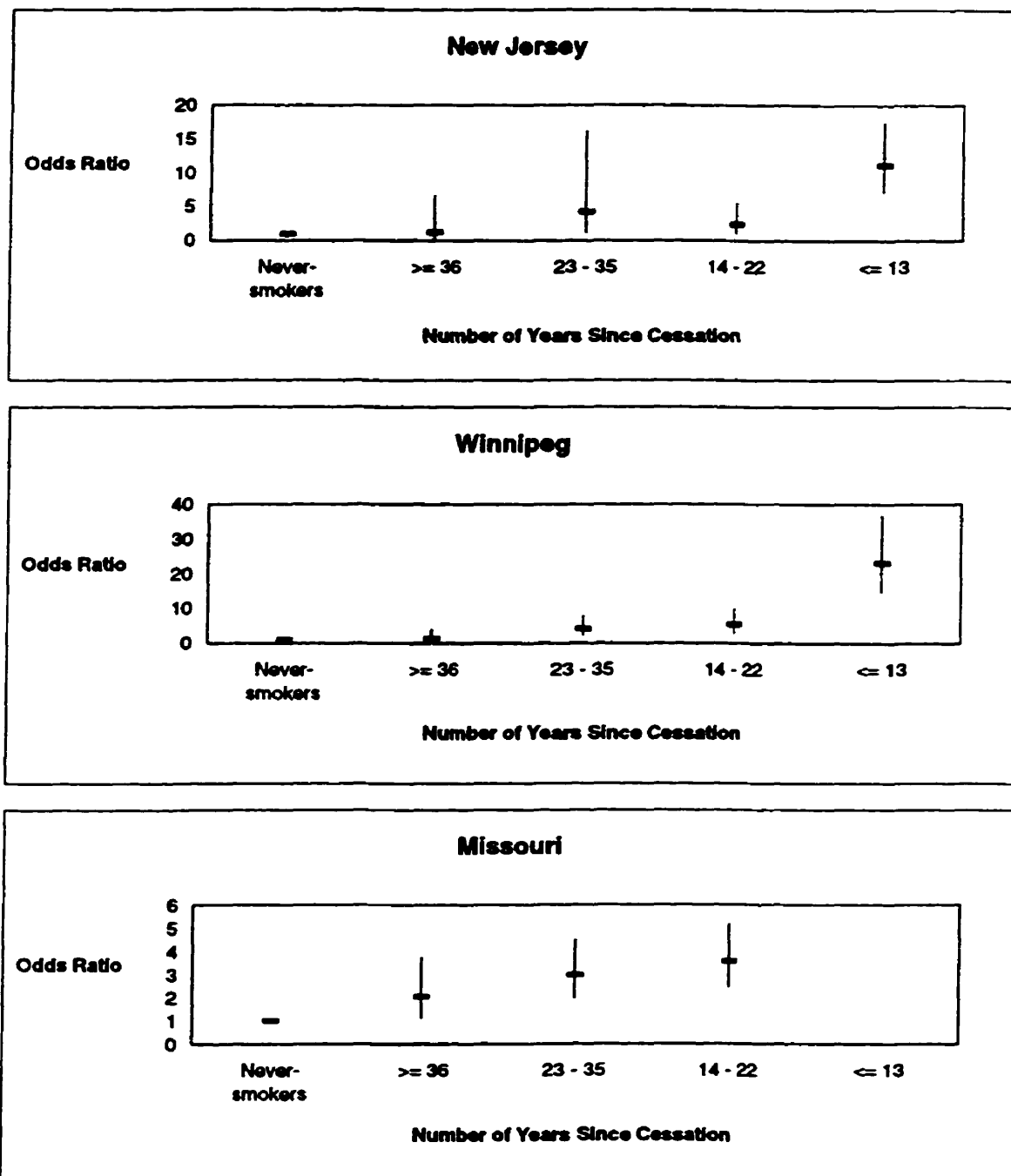
1 The New Jersey study was restricted to females of 'white' race. Living control subjects were frequency matched to living cases by age at ascertainment and 'race'; deceased cases were individually matched to controls by race, age, and closest date of death; proxy respondents were utilized for the latter cases and controls.

2 The Winnipeg study included both males and females; Controls were frequency matched to cases by age at ascertainment and gender.

3 Missouri controls were frequency matched to cases by age at ascertainment (defined as diagnosis for cases and interview for controls). This study was restricted to females who had not smoked in the 15 years prior to interview (corresponding to a minimum of 13 years prior to ascertainment).

Figure 4.15

Adjusted Odds Ratios [95% CI] Associated with Recency of Smoking



The above data are shown without simultaneous adjustment for pack-years and recency of smoking not only because these dimensions of smoking were correlated¹¹ (the value of Pearson correlation coefficient ranged from 0.6 to 0.7 in the three studies), but because the primary interest was the demonstration and between-site consistency of dose-response trends.

Previous radon researchers have used timing of smoking as a co-control variable in a logistic regression model already containing intensity of smoking. Analyses (not shown) which included terms for both smoking recency and smoking intensity [as cigarettes/day, $\log(\text{cigarettes/day} + 6)$, or combined intensity and duration as pack-years] resulted in highly significant odds ratios for each adjusted for the other. As outlined in section 3.6.3.1.1, in the development of a common logistic model which would account as completely as possible for variability in lung cancer odds attributable to smoking, four *a priori* alternatives were fit, with recency as an optional variable. Based on my empirical analyses, I also fit a fifth alternative, including terms for smoking intensity and recency and an interaction between the two.

4.5 DEVELOPMENT OF A COMMON LOGISTIC MODEL

4.5.1 OVERVIEW

Because of the large number of models fit and compared both within and across sites, the discussion of model development will be condensed. For those with an interest in greater detail, Appendix B specifies each of the more than 70 distinct models evaluated, numbered sequentially, within the four main model types defined *a priori*.

¹¹ Pearson correlation coefficients between the recency variable and the components of the other 3 main smoking models in the three sites (New Jersey, Winnipeg, and Missouri, respectively) were as follows: Log (intensity): 0.86, 0.81, 0.87; Intensity: 0.72, 0.66, 0.75; Log(duration*): -0.18, -0.06, -0.07; and, Duration: 0.94, 0.88, 0.94.

The purpose of model comparison was to find a single common model (according to changes in model deviance, the Schwartz Criterion, and the Akaike Information Criterion) which included tobacco-related risk factors for lung cancer, other suspected risk factors which were important in at least one site, and necessary design variables, which would provide the best possible fit across the three sites. Most models were fit on both the full set of subjects (as defined in section 3.3.2), and in samples restricted to the criteria defined in section 3.6.4.1.1¹² relating to the regularity of smoking. The first 42 models were fit without a term for radon exposure and included four alternative types of models for the representation of active smoking exposure (as defined in section 3.6.4.1.1):

- A Combined Intensity and Duration (in pack-years)
- B Log(intensity) and Log(duration*)
- C Intensity and Duration
- D Log(intensity) and Age

These model types were developed before conducting the empirical analyses (discussed in the previous section) for the purpose of establishing comparability with each other and a reference study. Recency of smoking was, at the outset of model development, an optional variable.

4.5.2 BEST-FITTING MODELS WITHIN SITE

Model fitting began in the Winnipeg site. This site was chosen as an initial reference because

¹²

Additional subject restrictions included: 1) Age < 80 years; 2) Smoking Intensity $\leq$ 40 Cigarettes/Day; and 3) 'Regular' smoking history. The third criterion was not applied to the Winnipeg study, since 97% of Winnipeg subjects were 'irregular' smokers, that is the data suggested that they had quit smoking more than once between their initial start date and the date of interview. The restricted samples included 274 cases and 253 controls in New Jersey, 643 cases and 698 controls in Winnipeg, and 415 cases and 960 controls in Missouri.

the original analyses included the fewest model covariates, and had shown no evidence for a radon effect; it was therefore thought that this would be the site in which improvements in overall model fit would have the most potential for impact on the accuracy and precision of radon coefficients. Initial models developed in Winnipeg included models A1 to A3, B1 to B22, and C1 to C3. These models were then analysed in the other two sites and compared for relative fit within and across sites. Subsequent models were fit in all three sites in parallel. Within all three sites, the best fitting models among the A, B and C models, were of the B type. In Winnipeg and Missouri, the same best fitting models¹³ [B1, which included recency as a significant term, in addition to log(intensity) and log(duration)] were selected by both the Akaike and Schwartz criteria. In New Jersey models B1 and B2 provided the best fit according to the Akaike and Schwartz criteria, respectively. Winnipeg was the only site for which the log(duration) term was significant, however. In the less well fitting C models, smoking duration was significant in Winnipeg, regardless of whether or not (a dichotomous term for) smoking recency was included; in New Jersey, smoking duration was significant only in models which did not also include recency; since at this stage of analysis, recency was a dichotomous term which discriminated between those subjects on the non-current smoking restriction for the Missouri study, this term was not relevant in Missouri, where duration was significant after adjustment for smoking intensity (in B models), but log(duration*) was not a significant predictor after adjustment for Log(intensity) (in C models).

As an alternative to the B models, D models were next evaluated. These replaced the log(duration) term, which was important only in Winnipeg, with an age term. The latter fulfilled the criterion of including all matching variables. Following the empirical analyses discussed in section 4.4, an ordinal recency term which could be included (with one less level) in Missouri was also added to maximize cross-site comparability of the smoking coefficients. The main effects D models (D1 to D5) provided fits comparable to the B

¹³ Excluding those models which included a term for year of ascertainment, a predictor of outcome which was not a risk factor or confounder.

models in all three sites. Inclusion of recency substantially improved model fit in Winnipeg, but improved model fit only marginally in New Jersey, and only among restricted subjects in Missouri. The D models were subsequently refined through the investigation of interactions between $\log(\text{intensity})$ and recency, gender, age, and source of information. A term for radon exposure (either in cumulative units of WLMs or concentration units of Bq/m^3), was then introduced. Radon interaction terms were also investigated (as outlined in section 3.6.7).

Part of model refinement involved a re-investigation of a term for smoking duration. Models (D20 and D21) containing terms for $\log(\text{intensity})$, recency and age demonstrated a superior fit in all three sites and by all fit criteria (except for the deviance criteria in Missouri) than models (D30 and 31) containing $\log(\text{dose})$ and $\log(\text{duration})$. The adequacy of common covariate coefficients across site was demonstrated for both types of models by the observation of a superior fit of an analysis of the combined data (constraining covariate coefficients to be constant across site) compared with the summed deviances of the three site-specific models (allowing coefficients to vary), taking into account degrees of freedom.

Table 4.11 summarizes the best-fitting models specific to each site (D models) and in an analyses of the pooled data (E models). The term 'optional' refers to any variables selected based on fit and/or importance; only those finally selected are shown. Optional variables were of three types: 1) terms related to active smoking for which extensive modelling was undertaken to determine the best of four types of models across sites (see Appendix B for details), 2) all potential modifiers of either the smoking or radon effects, and 3) family income. As indicated in section 3.6.6, default variables (other than site and radon) included age, gender, ethnicity, source of information (subject or proxy), education, and environmental tobacco smoke.

Table 4.11

Summary of Best-Fitting Logistic Models By Site

First Order Terms	Second Order Terms (All Optional)	Type	Included Terms			
Default:			NJ	Wpg	Miss	Pooled
Age		O (5)	✓*	✓	✓	✓*
Gender		D	R	✓*	R	✓*
Ethnicity		D	✓	✓*	R	S*
Education		O (3)	✓	✓*	✓	✓
Environmental Tobacco Smoke		O (5)	✓	✓	✓	✓
Source of Information (Proxy)		D	✓	X	R	S
Site						✓
Weighted Radon (Cumulative or Concentration)		C	✓	✓	✓	S
	Age x Radon (Cumulative or Concentration)	C				S
Optional:						
Recency of Smoking		O (5)	✓	✓	(R)	S
	Recency x Log(Intensity)	C	✓	✓	(R)	S
	Recency x Cumulative Radon	C		✓	(R)	
Log (Intensity) of Smoking		C		✓^	✓	✓
	Log (Intensity) x Proxy	C	✓	✓^	R	✓
	Log (Intensity) x Age	C	✓^			S
Age at smoking initiation		O (5)		✓		✓
Household income		O (5)	—	✓	✓	S

D Dichotomous Variable.

O Ordinal Variable (number of levels).

C Continuous Variable.

✓ Control variable in this site (implies a p value < 0.2 for optional second order terms).

X Source of information was not adjusted for in Winnipeg, since it was not a risk factor or matching variable.

* Default variable with p value < 0.05.

A Indicates a variable whose importance was limited to the all-subjects analysis.

R Indicates a variable for which adjustment was not applicable due to restriction on this factor.

(R) Indicates partial restriction which did not prevent adjustment on this factor.

S Indicates a term for which at least one site-specific difference was significant at a p-value < 0.20.

— Income information was not available in New Jersey; in the analysis of the combined data a fixed value corresponding to the mean amongst the controls from the other two sites combined was assumed for New Jersey subjects.

The final column of Table 4.11 indicates with an S the terms which varied (by the admittedly liberal type I error rate of 0.2) by study site. Although the adequacy of common covariates was demonstrated in terms of overall model deviance, the use of site-specific terms in an analysis of the combined data enabled maximum precision in the control of risk factors (e.g. family income) and allowance for effect modifiers the importance of which may vary by study because of differences in the target populations (e.g. recency of smoking). In addition, two terms (ethnicity and proxy respondent) on which matching was carried out in separate studies proved to be inappropriate for common adjustment (as will be discussed in section 4.6.1).

In New Jersey and Missouri, the same site-specific models were predicted regardless of the whether radon was expressed cumulatively or as average concentration. In Winnipeg, a positive interaction between radon exposure and recency of smoking on the odds of lung cancer demonstrated a p-value less than the criterion of 0.2 in the cumulative but not the concentration analysis. Although type I error is a possible explanation, only those interactions suspected *a priori* were investigated. Such an observation was consistent with prior evidence indicating a possible modifying effect of tobacco smoke exposure and its timing relative to that of radon. That such an effect was observed only in Winnipeg may have been due to the size of this study. The limitation of this finding to the cumulative radon analysis which accounted for the greater home occupancy of females was also consistent with a true effect. Separate analyses in two recency strata (current and ex-smokers, ie. using the Missouri restriction criterion) were therefore carried out.

In the analysis of the combined data, a three-way interaction between age, cumulative radon, and site achieved a p-value less than the *a priori* criterion of 0.2. The diminution by age (greater or less than age 55) of the effect of radon on lung cancer rate ratio was greater in New Jersey ($p < 0.01$) than in the other two sites. The possibility that this observation resulted from the decreasing weighting of earlier exposures was considered, but this would

have had the opposite effect on risk estimates. This observation was, however, consistent with those among miner studies, which showed a sharp decline in the ERR after 65 (Lubin et al., 1994a), and residential studies in which the modifying effect of age on radon-associated risk was explored. Since a substantial percentage of the pilot study subjects were over 65 at ascertainment (in New Jersey 52% of cases and 48% of controls; in Winnipeg, 50% of cases and 47% of controls; and in Missouri, 74% of cases and 75% of controls), a stronger effect at younger ages would be diluted in an overall estimate. The effect of radon was therefore investigated within age strata.

4.5.3 RESTRICTION OF SUBJECTS

Within each study site, analyses conducted on subject samples restricted to those which demonstrated homogeneity in exposure-response for smoking in the analyses by Doll and Peto (1978) had lower deviances than those conducted on the complete samples of subjects. Small, if any, differences were observed in the constellation of covariates in the best-fitting models in these two sets of samples. No differences were observed between samples in Missouri. In the other two sites all between-sample distinctions were such that the variable in question was important only when the analysis was conducted on the full sample of subjects, suggesting that the differences relied on increased power to detect a given effect rather than indicating heterogeneity in effects.

According to Breslow and Day (1980), although it would have been standard procedure in the past to adjust for smoking effects by restricting the case control sets to subjects who were homogeneous for smoking according to some pre-described definition, the loss of data in proportion to the stringency of control) is wasteful. The remainder of the results, provided for complete subject samples, follow their recommendation of the use of modelling as a more satisfactory technique.

4.5.4 STRATIFIED ANALYSES

Two separate stratified analyses were completed, one by age and one by recency. In the recency-specific analyses, age was treated as a dichotomous variable. In the age-specific analyses, recency was treated as an ordinal variable with five levels (except in Missouri where only four levels were applicable), and an interaction of recency with (log) intensity of smoking was allowed for.

Table 4.12 summarizes the models commonly employed across sites for stratification (separate analyses by strata) with respect to age at ascertainment and recency of smoking, and compares the covariate constellation to that in the original published analyses in each site.

The stratified common analysis included a term for every covariate included in the analyses documented by the primary investigators with the exception of occupational exposures. Investigated in the primary studies in New Jersey and Winnipeg, occupational exposures were included as a covariate only in New Jersey (where socio-economic covariates were excluded from the final model). Original to the common analysis is the investigation and modelling of recency and age at ascertainment as modifiers of the effect of smoking intensity on lung cancer odds, and control for age at smoking initiation, household income, and ethnicity. All of the latter factors were important independent risk factors for lung cancer in at least one site. Effects of the covariates are discussed in section 4.6.1.

Figure 4.16 plots the adjusted odds ratios (and 95% confidence limits) associated with an increment of 10 weighted WLMs radon exposure within strata specific to age, and within strata specific to recency of smoking. The values of 10 weighted WLMs and 100 weighted Bq/m³ were selected as reference points for expression of the results (analysed in continuous unit); 10 weighted WLM was between the median (9.8) and the mean (11.5) values for cumulative radon exposure in the 5 - 50 years prior to ascertainment, among all pilot study

subjects combined; 100 Bq/m³ was a typical index value in primary radon case-control studies, and exceeded that of the median (62.8) or mean (83.1) weighted average radon concentration (again, 5 - 50 years prior to ascertainment) among the present study subjects combined. Approximately 10.9% of US homes have radon concentrations in excess of 100 Bq/m³ (NRC 1998). Unless otherwise noted, results were similar for exposures expressed per weighted WLMs and weighted Bq/m³ average concentration.

Ex-smokers were defined as those who had quit smoking at least 13 years prior to ascertainment; all ever-smoking Missouri subjects were included in this group. Current smokers were those who had smoked within 13 years of ascertainment. Never-smokers formed the reference group in both recency strata. Comparing the rate ratio for radon exposure in ex-smokers with that of current smokers, no consistent pattern was discernable even among point estimates. Similarly, no difference was discernable between age strata. Large confidence intervals included the null value of one within smoking-specific and age-specific strata, even after combining the site-specific estimates. Given the inadequacy of stratum-specific data, all further radon analyses were unstratified.

Table 4.12

Covariate Control in Original and Stratified¹ Common Analyses

Factor	Type of Variable and Number of Levels						
	Original Analysis			Common Analysis			
	NJ ¹	Wpg ²	Miss ³	NJ	Wpg	Miss	Pooled
Age	U	S ¹	S 3 O (6)	S 2 ¹ O (5)	S 2 O (4)	S 2 O (5)	S 2 O (5)
Gender	R	S ¹	R	R	D	R	D
Ethnicity			R	D	D	R	D
Source of Information (Proxy)	D		R	D	X	R	D x site
Education		U	U [*]	O (3)	O (3)	O (3)	O (3)
Environmental Tobacco Smoke			U [*]	O (5)	O (5)	O (5)	O (5)
Recency of Smoking	U		S 2 U [*]	S 2 O (5) [†]	S 2 O (5)	R O (4)	S 2 O (5)
Intensity of Smoking	S 4 U	D	Pack- Years [*]	log C	log C	log C	log C
Intensity of Smoking x Proxy	U			C	C	R	C x site
Age at smoking initiation				O (5)	O (5)	O (5)	O (5)
Family Income	—			—	O (5)	O (5)	O (5) x site
Occupational Exposures	U		—			—	

1 Schoenberg et al., 1990

2 Létourneau et al., 1994

3 Alavanja et al., 1994; note that results were provided in two ways: 1) stratified on age or smoking status (never vs. ex-smokers) and 2) in an overall model with age and one or two of several other control variables (never vs. ex-smoker, pack-years of smoking, years since quit smoking, previous lung disease, passive smoking, saturated fat consumption, or education).

U Indicates a control variable whose data type was unspecified.

• Indicates a control variable not included in the main results but in separate analyses as described in note 3 above.

S Refers to a factor on which strata were formed for separate analyses; the number of strata are indicated.

h Two separate stratified analyses were completed, one by age and one by recency. In the recency-stratified analysis, age was treated as a dichotomous variable, and the interaction (not shown) between age and (log) intensity of smoking was also controlled. In the age-stratified model, recency was treated as an ordinal variable with 5 levels; an interaction term between recency and (log) intensity of smoking was also included.

† Age and gender were strata for conditional logistic regression in Winnipeg; the paper reported unconditional results since radon results did not differ between conditional and unconditional analyses.

D Dichotomous variable.

O Ordinal variable (number of levels).

C Continuous variable.

R Indicates a variable for which adjustment was not applicable due to restriction on this factor.

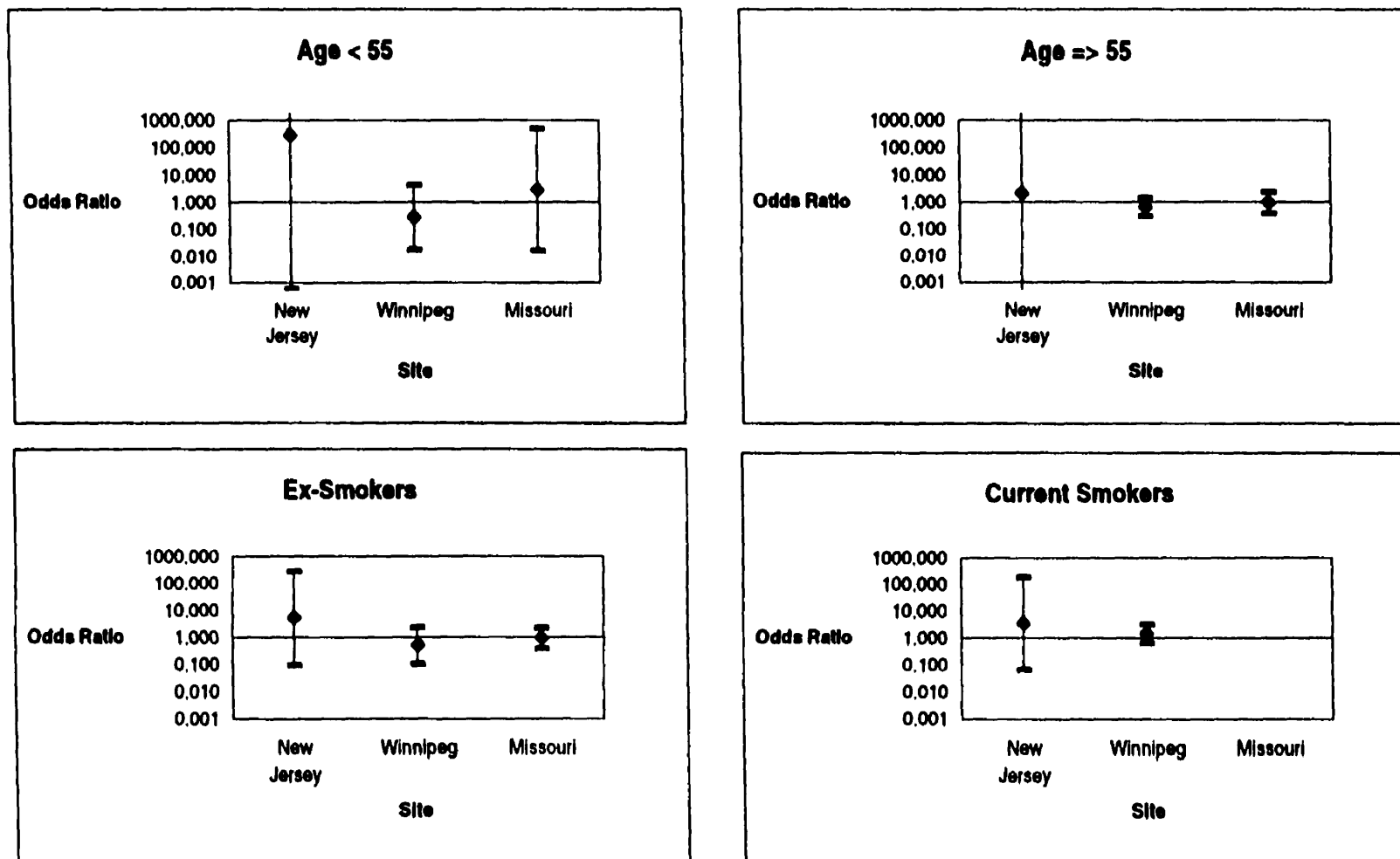
— Indicates a factor not ascertained in the original study.

X Source of information was not adjusted for in Winnipeg, since it was not a risk factor or matching variable.

x site Indicates a covariate for which site-specific adjustment was made.

Figure 4.16

Adjusted Odds Ratios [95% CI] Associated With An Increment of 10 Weighted WLMs Radon Exposure Within Age and Smoking Strata



4.5.5 THE FINAL COMMON LOGISTIC MODEL

The final common model is delineated in Table 4.13. Figure 4.17 plots adjusted (from the final model) and the crude odds ratios in radon units of 10 weighted WLMs and 100 weighted Bq/m³. A slight (but non-significant) increase is observable in the estimated radon odds ratios after adjustment in Winnipeg. Pershagen et al., (1992) also observed what they termed 'negative confounding' in his Stockholm study, the radon effect becoming more apparent after adjustment. Confounding *per se* is not suggested in the present analyses, as no association was observed between radon and any risk factors in any site. The slight change in radon estimates after adjustment in the Winnipeg site therefore suggests the presence of (and possible of selection with respect to) a strong risk factor (such as smoking, education, or interview). The odds ratios associated with radon were consistent between the three sites; confidence intervals included the null value of one.

Table 4.13

Covariates Included in the Final Common Logistic Model

First Order Terms	Second Order Terms	Type	Included Terms			
			NJ	Wpg	Miss	Pooled
Age		O(5)	✓*	✓	✓	✓
Gender		D	R	✓*	R	✓*
Ethnicity		D	✓	✓*	R	S*
Source of Information (Proxy)		D	✓	×	R	S*
Education		O(3)	✓	✓*	✓	✓*
Environmental Tobacco Smoke		O(5)	✓	✓	✓	✓
Recency of Smoking		O(5)	✓*	✓*	(R)	S*
	Recency x Log(Intensity)	C	✓*	✓*	(R)	✓*
Log (Intensity) of Smoking		C	✓	✓*	✓*	✓*
	Log (Intensity) x Proxy	C	✓*	✓*	R	✓*
Age at Smoking Initiation		O(5)	✓*	✓*	✓	✓*
Household Income		O(5)	—	✓*	✓*	S*

D Dichotomous Variable.

O Ordinal Variable (number of levels).

C Continuous Variable.

✓ Control variable in this site (implies a p value < 0.2 for optional variables).

* Variable for which the added-last Chi-squared p-value was < 0.05.

R Indicates a variable for which adjustment was not applicable due to restriction on this factor.

(R) Indicates partial restriction which did not prevent adjustment on this factor. The recency variable was coded as 0 for non-smokers, 1 for subjects who had smoked in the 13 years prior to ascertainment, and 2 through 4 for those who quit smoking between 14 and 36 years prior to ascertainment. The Missouri study excluded those who would have been coded as level 1 for the recency variable. Thus four of the 5 levels of the recency variable remained applicable among Missouri subjects.

S Indicates a term for which site-specific adjustment was allowed in the analysis of the combined data.

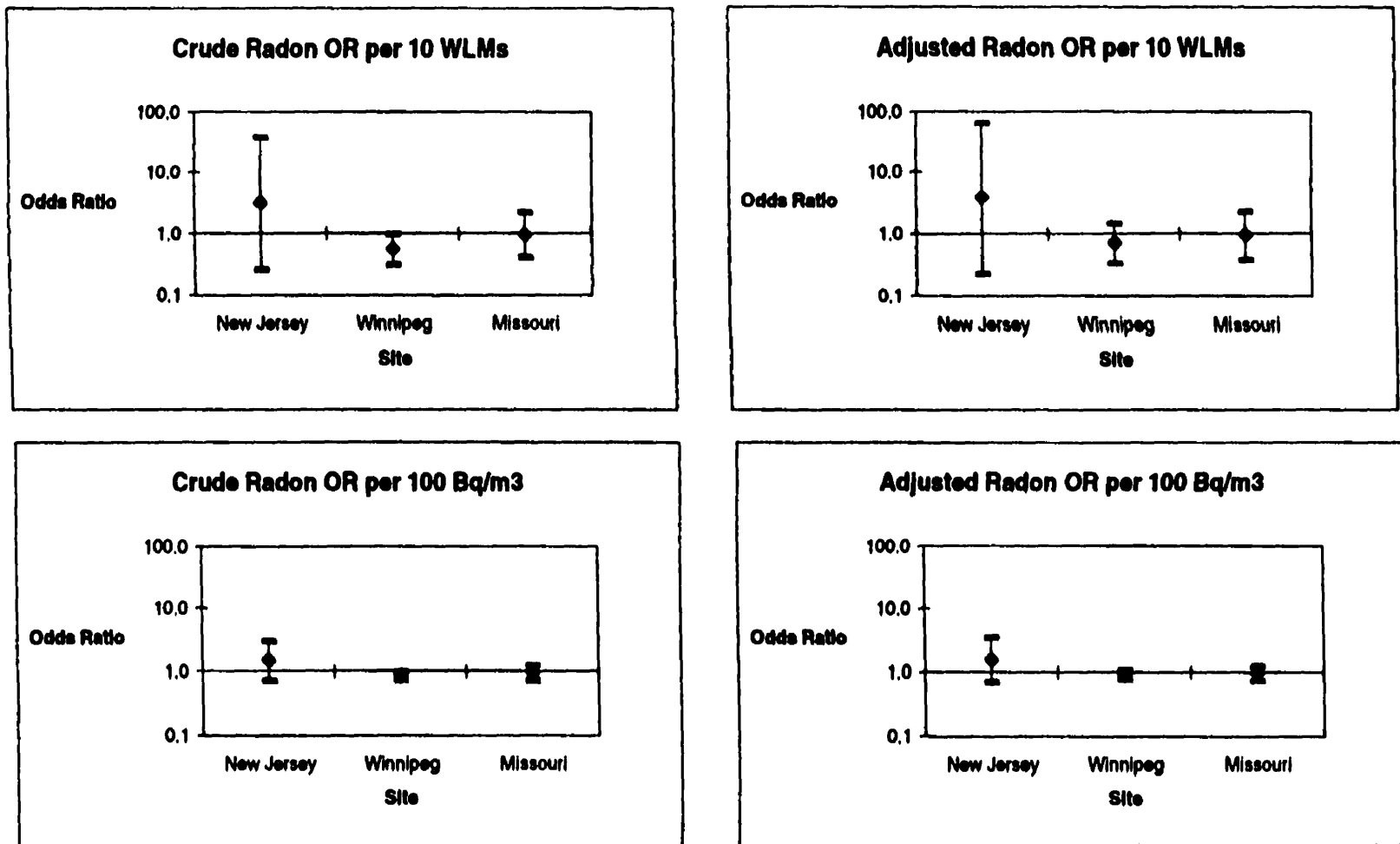
— Income information was not available in New Jersey; in the analysis of the combined data a fixed value corresponding to the mean amongst the controls from the other two sites combined was assumed for New Jersey subjects.

×

Source of information was not adjusted for in Winnipeg, since it was not a risk factor or matching variable.

Figure 4.17

**Crude and Adjusted Odds Ratios [95% CI] Associated with an Increment of
10 Weighted WLMs and 100 Weighted Bq/m³ Radon Exposure**



4.6 RESULTS OF THE FINAL ANALYSES

4.6.1 PREDICTORS OF LUNG CANCER

The analysis of the combined primary data from the published North American studies of lung cancer and residential radon confirmed or revealed important results concerning a number of covariates, namely (in order of importance), active smoking, family income, education, ethnicity, and source of information. Apart from the latter, which is discussed as a 'spurious confounder', these were found to be independent risk factors for lung cancer. The associated results discussed in the following sections are those derived from the final "common model" in a logistic regression analysis of the primary data from all three sites combined. This model included terms for site, smoking intensity (as log of the average number of cigarettes smoked per day during the period smoked with the 5 - 50 years prior to ascertainment plus six) , smoking recency, the interactions of smoking intensity with smoking recency and source of information¹⁴, age at smoking initiation, gender, ethnicity, income, education, age at ascertainment, environmental tobacco smoke exposure (ETS), and site-specific radon exposure (in the 5 - 50 years prior to ascertainment). Negligible (if any) differences were observed between the covariate coefficients reported herein, based on an analysis in which radon is expressed in cumulative exposure units, and one in which radon exposure is in units of average concentration.

4.6.1.1 Active Smoking

In a careful review of lung cancer and cigarette smoking, Wu-Williams and Samet (1994) concluded that 'it is unlikely that any one model will satisfactorily describe lung cancer incidence in relation to all aspects, including current and past exposure and dose and duration

¹⁴ The main effect of source of information was also included in the New Jersey site.

of exposure'. These authors cited Thomas (1988), who warned that 'because many competing models will give essentially the same fits to the data, it is important not to conclude that a well-fitting model represents the truth'. For those readers who are nonetheless interested in the specifics of the best fitting common logistic model (with covariates as specified in Table 4.13) for the combined data from the three included studies, Table 4.14 summarizes the adjusted regression coefficients (and standard errors) for the smoking-related covariates.

Table 4.14
Adjusted Regression Coefficients For Smoking-Related Covariates
in the Final Common Logistic Model

Variable	Type of Variable (Number of Levels)	Regression Coefficient (SE)
Log(Intensity)	C	0.497 (0.146)
Log(Intensity) x Recency	C	0.403 (0.072)
Log(Intensity) x Proxy Response	C	-0.879 (0.113)
Recency (Winnipeg)	O (5)	0.561 (0.065)
Recency x New Jersey	O (5)	-0.179 (0.078)
Recency x Missouri	O (4)	-0.289 (0.098)
Age at Smoking Initiation	O (5)	0.059 (0.045)
Environmental Tobacco Smoke	O (5)	0.005 (0.024)

Intensity number of cigarettes smoked per day (on average during period of smoking within the 5 - 50 years prior to ascertainment) plus six

C Indicates a continuous variable

O Indicates an ordinal level with the number of levels indicated

The investigation of five alternative models for the control of intensity and duration of cigarette smoking in the present analysis provided unusually thorough treatment of this lung cancer risk factor in the context of radon research. A particular requirement of the common model was to provide adequate fits to data from two studies which reflected the natural

prevalence of smoking in lung cancer cases and the general population (New Jersey and Winnipeg), as well as from a study which restricted its lung cancer cases and base population to persons who had not smoked within the 15 years prior to their interview (Missouri). The importance of a smoking-related variable which could account for this difference in target population, and ultimately for site-specific terms to allow for remaining differences, rendered "recency" of smoking a more important variable than it may be in other study contexts, especially without adjustment for duration. By the same token, however, duration of smoking was not sufficiently predictive in all sites to be included in a common model, even without adjustment for recency. Recency of smoking (as an ordinal variable) was, in any case, highly correlated (0.88 - 0.94) with duration of smoking in these populations, which varied little in age at smoking initiation and were matched on age. That the recency of cigarette smoking modified the effect of smoking intensity on lung cancer relative risk confirmed a similar observation by Lubin et al., (1984) in their re-analysis of data from a large European case-control study.

Expression of active smoking exposure as $\log(\text{cigarettes/day} + 6)$ in a model controlling for age, provided a better fit to the data in the three sites than did three main types of alternative models. These alternatives included models in which active smoking was expressed as $\log(\text{cigarettes/day} + 6)$ and $\log(\text{age} - \text{average age at smoking initiation})$, as cigarettes/day and smoking duration (as age at cessation minus age at initiation), and in pack-years (a unit which combines intensity and duration). The better fit of curvilinear compared with a linear form for smoking intensity confirmed the findings of Doll and Peto (1978).

Doll and Peto (1978) fit separate exponents for intensity and duration of smoking to explain smoking rates. Their best-fitting model is expressed in equations 3.4 through 3.6. Analogously, the present analysis fit separate coefficients which explained rate ratios in a log scale (equation 3.7). Doll and Peto modified the exponent of the intensity term in their model of incidence rates, and observed adequate fit for any exponent greater than two, if a

suitable background (constant) was first added to the dose (in cigarettes/day). (In the present logistic analyses, the addition of a constant to smoking intensity in the log form was also necessary to avoid undefined values for non-smokers, whose average intensity was zero.) In their best-fitting model with exponent two, this constant took a value of six. Using this same constant to account for background risk in a logistic model, coefficients (and standard errors) of 1.6 (0.12) , 2.2 (0.13) , and 1.3 (0.16) were observed per unit $\log(\text{cigarettes/day} + 6)$ in a comparable model [B2, which represented smoking as $\log(\text{intensity})$ and $\log(\text{duration}^*)$ but did not include recency] in New Jersey, Winnipeg and Missouri, respectively. (Although the coefficient in the final combined model was reduced to 0.6, a number of interaction terms involving the $\log(\text{intensity})$ variable obscure any direct comparison). Doll and Peto's estimate was roughly equivalent to an odds ratio per unit $\log(\text{intensity})$ of 7.4. The 95% confidence intervals for the odds ratios from the comparable model in the present analysis were [3.9,6.2], [7.0, 11.6], and [2.7, 5.0], for New Jersey, Winnipeg, and Missouri, respectively. Although the lack of an estimate of standard error from Doll and Peto's study prevents a strict evaluation of the comparability of this data, the confidence intervals for the studied sites were fairly consistent. The exposure-response slope in Missouri was lower than elsewhere since the doses in this study were less recent, being related to tobacco exposure more than 15 years prior to interview.

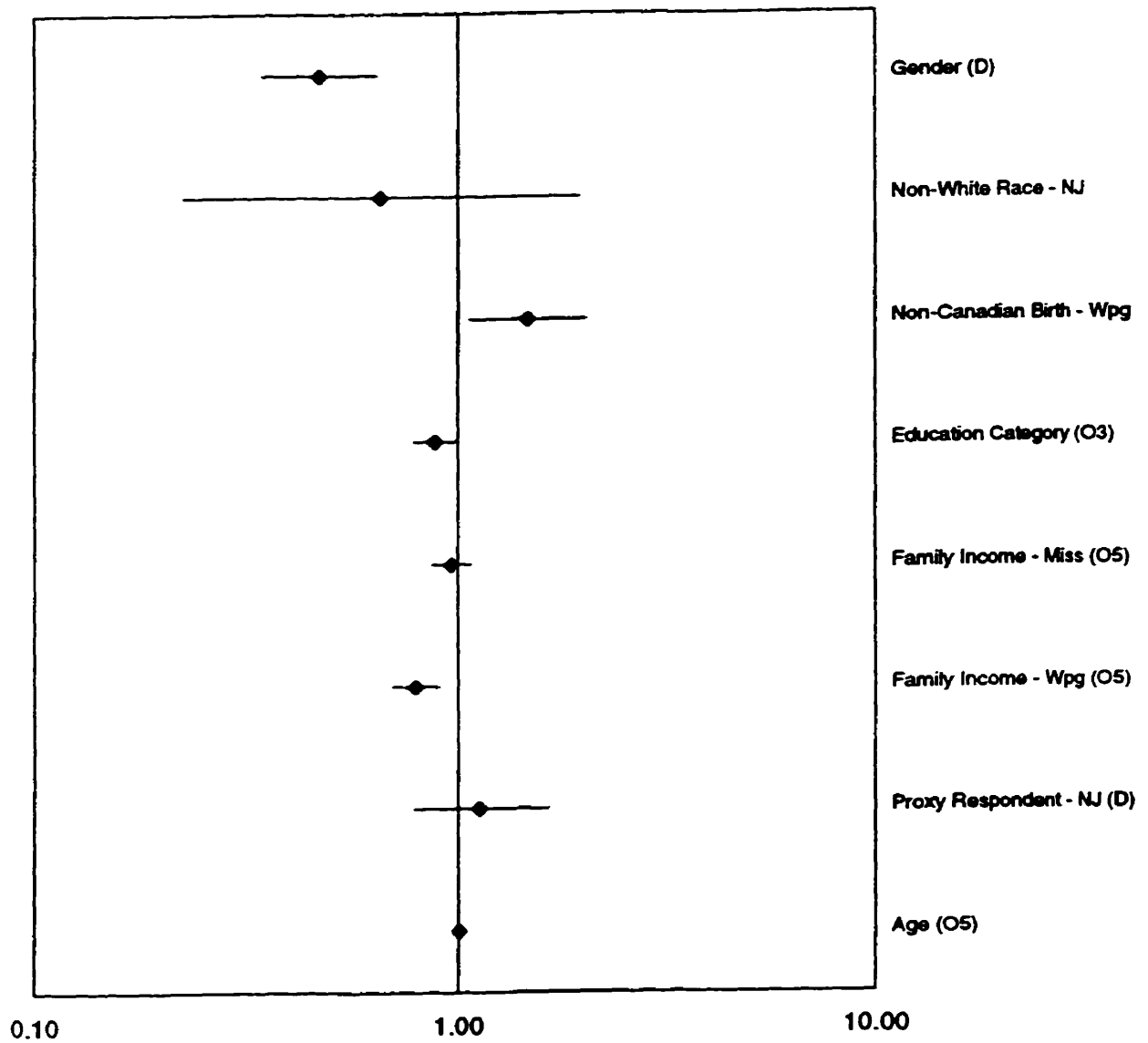
The comparability of the above results reduces concern about the assumption of the same value for background as in Doll and Peto's analysis and more generally validates the smoking exposure data collected in the radon studies.

4.6.1.2 Lung Cancer Predictors Other Than Smoking

Figure 4.18 provides estimated odds ratios for covariates other than those related to cigarette smoking. Since gender, race, proxy respondent, and age were matching variables, their associated odds ratios are not interpreted. The other factors are discussed below.

Figure 4.18

Adjusted Odds Ratios [95% CI] for Covariates Other Than Smoking



**Estimated Odds Ratio
Based on an Analysis of the Combined Data**

4.6.1.2.1 *Socio-Economic Correlates*

Level of education and family income were both significant and independent predictors of lung cancer. The correlation between their logistic regression coefficients were -0.2 or less, indicating that these variables represent fundamentally different phenomena.

Based on site-specific analyses and adjusting for all other common covariates, the adjusted lung cancer odds ratio associated with family income was similar in the two sites for which data was available, being 0.83 [0.73, 0.95] in Winnipeg and 0.89 [0.80, 0.96] in Missouri per increment in income level on a five point ordinal scale (in country-specific currency). Compared with persons whose reported annual family income was less than \$10,000, those with incomes greater than \$ 50,000 had a 53% lower lung cancer rate in Winnipeg, and a 37% lower lung cancer rate Missouri. A slight (but non-significant) distortion of the site-specific estimates for family income was observed in the analysis of the combined data (with the use of a fixed value for this missing variable in New Jersey), in which an incremental increase in family income was associated with a lung odds ratio of 0.79 [0.70, 0.90] in Winnipeg, and 0.96 [0.87, 1.07] in Missouri.

The effect of education (again, after adjustment for all other common covariates) was consistent across the three sites; a 12% [1%, 22%] decrease in lung cancer rate was associated with an incremental increase of each education level on a three point ordinal scale (including elementary, secondary, and post-secondary education). Compared with persons who had only elementary education, those with (at least some) post-secondary education had a 23% lower rate of lung cancer.

4.6.1.2.2 *Ethnicity*

Although the use of racial or ethnic data is perhaps a contentious issue, the use of this type

of data in the design of the US studies, to restrict (in Missouri), or match (New Jersey) subjects necessitated attention to this factor (in spite of the lack of interpretability of its coefficient in the latter site). Evidence in the Winnipeg study that country of birth differentiated cases from controls necessitated consideration of this factor as a potentially important lung cancer predictor. Although recognizing the definitions and classifications recently promulgated by the US Office of Management and Budget for the terms 'race' (as American Indian or Alaskan Native, Asian, Hawaiian Native or other Pacific Islander, African American, and White), and ethnicity (as Hispanic or Latino origin versus non-Hispanic origin) (Wallman 1998), I use the term "ethnicity" to more loosely represent the fundamental distinctions made by the principal investigators in the original studies, that is, between 'white' and other race in the US studies, and country of birth (Canadian or other) in the Winnipeg study. Separate terms (formed by an interaction between the respective ethnicity factor and site) were, however, employed in the analysis of the combined data.

Persons born outside Canada had a 44% [7%, 99%] higher rate of lung cancer than those born in Canada after adjustment for all other common covariates. Although considered as a covariate by Letourneau et al. (1994), since inclusion of country of birth did not affect the radon estimates in Winnipeg, this term was not retained in their final model. Although country of birth could reflect other influences (e.g. ability to find employment in their field of expertise, or carcinogenic occupational exposures), the importance of this term even after adjustment for education level and income, to which (according to Sexton et al., 1993) ethnic differences in cancer differences have previously been attributed, suggests that at least part of the effect may be due to genetic susceptibility. Moreover, ethnicity was included (in addition to tobacco smoke, socioeconomic status and age) by Siemiatycki et al., (1994 and 1995), as a predictor in lung cancer models.

Sexton et al. (1993) described race-related differences in lung function and postulated genetic factors as one possible explanation, although indicating that it is unknown the extent to

which differences in lung function may arise from exposures to airborne pollutants such as ETS and allergens in house dust and the use of gas-fired cooking stoves. Greater genetic susceptibility, or the potential for other unknown risk factors were also explanations offered by Schwartz and Swanson (1997) for their observation of higher rates of lung cancer among African Americans than 'whites' within each level of smoking exposure even after adjustment for age, education, and number of years since smoking cessation. These authors cited four lines of evidence supporting the hypothesis of differential genetic susceptibility to lung cancer by race: 1) differences in the frequency of various cytochrome P450 and phase II bio-degradation enzymes involved in carcinogen metabolism; 2) a slow acetylator (a detoxification process) genotype (at the NAT2 gene locus) recently described in African Americans; 3) a novel restriction fragment polymorphism in a specific gene (CYP1A1) associated with an ERR of 2.6 for adenocarcinoma of the lung, exclusively identified in African Americans; and 4) mutagen sensitivity which is also reportedly associated with a lung cancer RR of 3.7 among African Americans.

The inclusion of "ethnicity" as a covariate in the logistic model assumes a multiplicative relationship between radon exposure and "ethnicity" This representation does not preclude its potential role as a mediator of the relationship between radon exposure and absorbed radiation dose. The possibility of a sub- or supra-multiplicative relation is one that should be investigated in the Phase II analysis.

4.6.2 RATE RATIO OF LUNG CANCER AS A FUNCTION OF RADON EXPOSURE IN THE DEFAULT EXPOSURE TIME WINDOW

Figures 4.19 and 4.20 provide the adjusted radon odds ratios from the final common model associated with an increment of 10 WLMs and 100 Bq/m³. Results are displayed for each site and for the three sites combined (both assuming random and fixed effects). For each site, three estimates are displayed: 1) the first is the estimated odds ratio (and 95% CI) for

lung cancer per increment of radon exposure based on the parallel analysis, that is, an analysis using the common final model but in separate within-site analyses. In the parallel analyses, all covariate coefficients are allowed to vary between sites; 2) The second confidence interval represents the stage one 'combined' estimate of the site-specific odds ratio for radon. This stage one estimate was obtained from an analysis in which the data from all three sites were combined, and the majority of covariates¹⁵ were constrained to be constant across site. 3) The third site-specific estimate shown is the stage-two estimate of the radon odds ratio. As described in the section 3.7.1, the stage two or 'shrunk' estimate is a weighted average of the site-specific and the final combined (pooled) estimate obtained by the 'two stage method' for random effects. The stage-two estimate of the within-site effect thus takes into account between site-variability.

One may observe a very slight (and non-significant) difference between the parallel and stage one estimates in the New Jersey site. The marginal increase in the estimated radon odds ratio for the New Jersey site based on an analysis of the combined data may be attributable to bias arising from imputation of a fixed value for family income on which information was not collected in this site. A less complete overlap in covariate constellations is to be expected between the seven studies to be combined in the phase II North American pooling. Although a more sophisticated imputation method could be employed, because of the large number of terms for which site-specific adjustment was important in the analysis of data from only three sites, (4 out of 12 covariates), the combination of estimates derived from site-specific analyses seems a more advisable alternative in the next phase.

¹⁵

As summarized in Table 4.13, family income, ethnicity, source of information (as a main effect), and recency of smoking, were treated separately by the use of interaction terms with site, as was the main exposure (radon).

Figure 4.19

**Adjusted Odds Ratios [95% CI] Associated with an Increment of 10 Weighted
WLMs Radon Exposure in the 5 - 50 Years Prior to Ascertainment:
Site-Specific and Pooled Results**

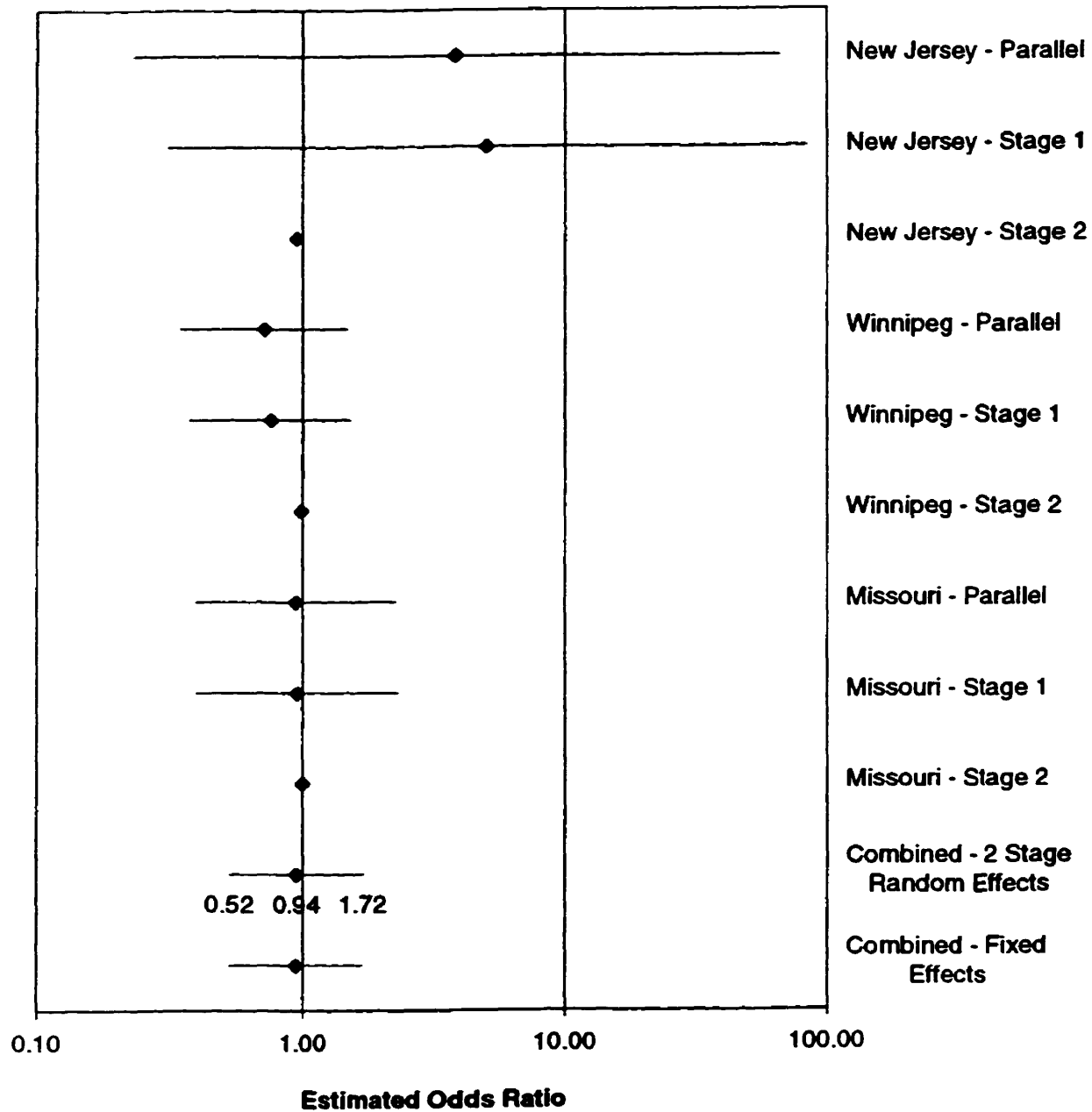
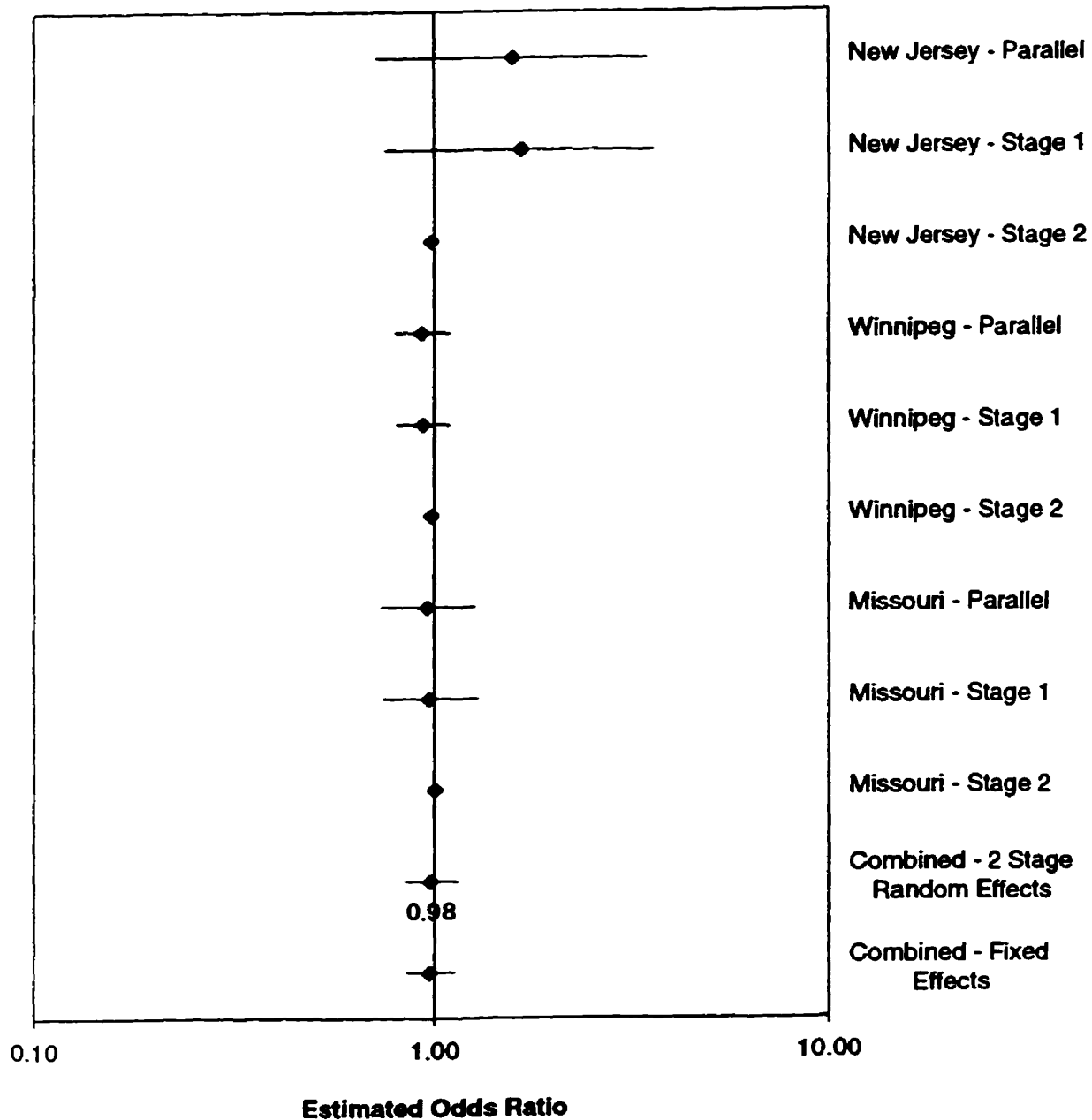


Figure 4.20

**Adjusted Odds Ratios [95% CI] Associated with an Increment of 100 Weighted Bq/m³ Radon Exposure in the 5 - 50 Years Prior to Ascertainment:
Site-Specific and Pooled Results**



The combined odds ratio associated with an increment of 10 weighted WLMs of radon exposure in the 5 - 50 years prior to ascertainment was 0.94 [0.52, 1.72], with a confidence interval that included the null value of 1.0. Between study variability in the estimated radon odds ratio was negligible ($\text{Tau} = 0.09$) as is evident from the similarity between the two-stage and fixed effects.

The combined odds ratio associated with an increment of 100 weighted Bq/m³ of average radon concentration was 0.98 [0.84, 1.14], with no between site variability.

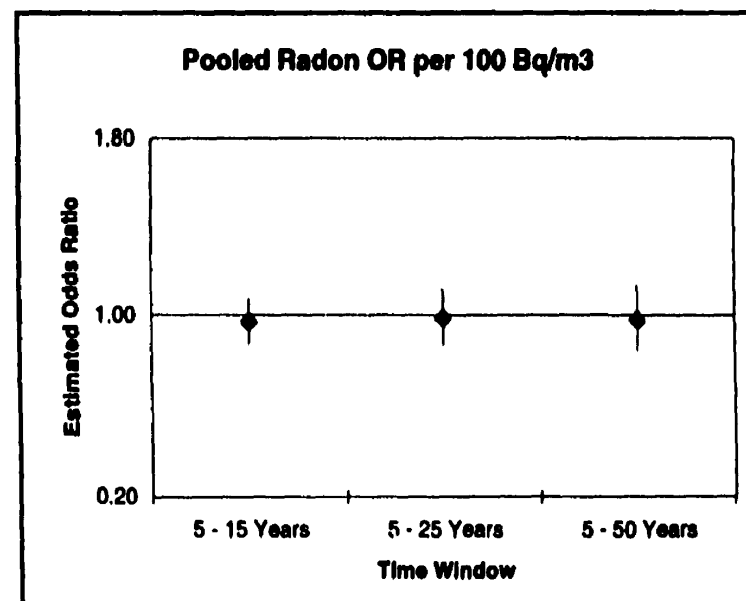
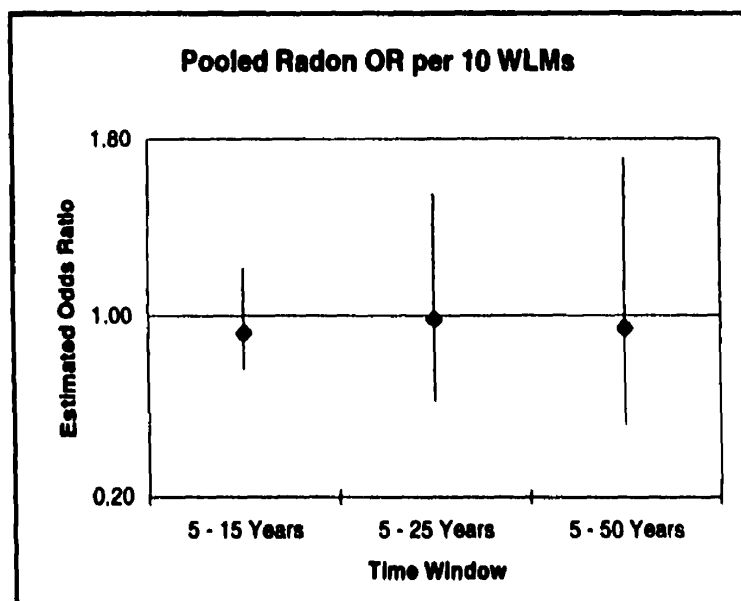
Compared with the separate (parallel) analyses, the analysis of the combined data reduced uncertainty in the estimate of the radon odds ratio.

4.6.3 SENSITIVITY OF THE RADON RESULTS TO TIME WINDOW DEFINITION

Figure 4.21 provides estimated odds ratios associated with increments of 10 WLMs cumulative radon exposure and 100 Bq/m³ average radon concentration, for three distinct combinations of time windows: 5 - 15, 15 - 25, and 5- 50 (the default scenario) years prior to ascertainment, based on an analysis of the combined data. The weighting of radon exposure under these scenarios was defined in section 3.6.5. Estimates of radon effect were most precise within the time window 5 - 14 years prior to ascertainment, but the odds ratio point estimates were highest in the time window 5 - 25 years prior to ascertainment. According to Rothman (1981), 'the maximum point estimate of effect selects the most appropriate assumption about the empirical induction period, and simultaneously offers an estimate of the maximum effect unobscured by an inappropriate assumption'. This statement presumes that the exposure is causally related to a positive effect, however. The increasing uncertainty in the more historically distant time windows may simply reflect increasing uncertainty in radon exposure, arising from increasingly frequent missing radon data as time prior to ascertainment increases.

Figure 4.21

**Sensitivity of the Pooled Results to Time Window Definition: Adjusted Odds Ratios [95% CI]
Associated with Increments of 10 Weighted WLMs and 100 Weighted Bq/m³ Radon Exposure**



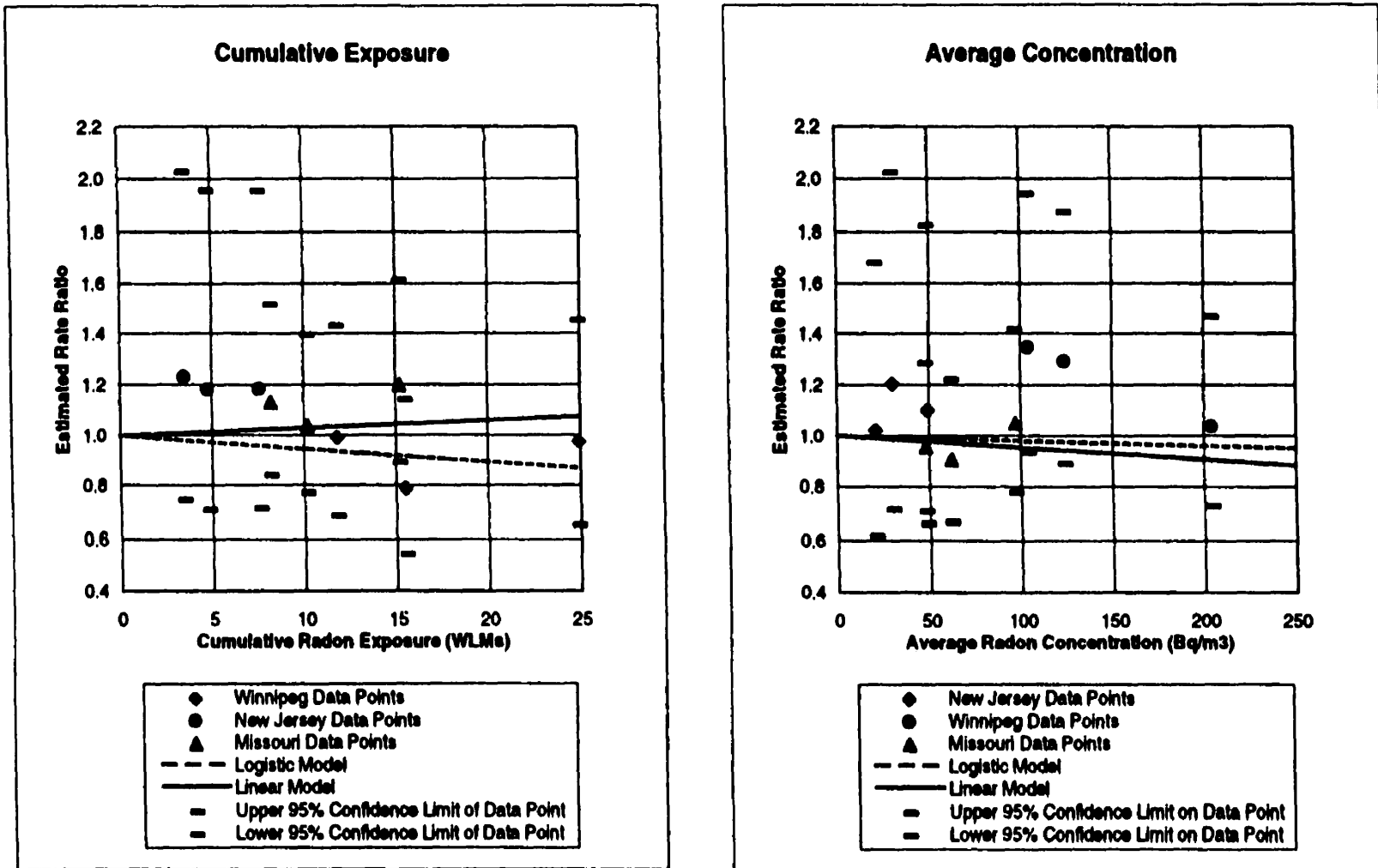
Considering the non-significance of the variations observed, their lack of impact on the interpretation of the results, and the lack of theoretical evidence to suggest that exposures received earlier are less important than those received later, it seems prudent to retain the default assumption of the largest time window until the sensitivity analyses can be replicated with more precision during the phase II analysis of all seven North American studies.

4.6.4 SENSITIVITY OF THE POOLED RESULTS TO MODEL FORM

Figure 4.22 illustrates for the default (5-50 year) time window, sensitivity of the radon rate ratio estimates (in both cumulative and average concentration exposure units) to model form. The fitted exposure-response curves are combined (two-stage) estimates of the RR per unit effective radon exposure modelled as a continuous variable. These fitted curves are based on either an exponential (logistic) or linear (ERR) model. The exposure-response relationship predicted by the final logistic model is denoted by a dashed line, and that from the linear excess relative risk model by a continuous line. The data points on these plots are point estimates of the rate ratio from parallel analyses by site in which radon exposure was treated as a categorical variable (quartiles within site) and adjusted for all covariates in the final common model. Upper and lower 95% confidence limits for the site-specific (categorical) data points are also plotted. (Confidence intervals for selected increments of continuous exposure, ie. at points on the predicted curves, are provided later, in Table 4.15.) Although the category-specific odds ratios were not strictly scaled to zero exposure (reference levels vary between site and are never zero), given the width of confidence intervals, scaling of these ORs to either the multiplicative or additive scale would not affect the conclusions. The two fitted curves and the site-specific data points were all adjusted for the common set of covariates.

Figure 4.22

Sensitivity of the Pooled Results to Model Form:
Adjusted Rate Ratios for Radon Exposure in the 5 - 50 Years Prior to Ascertainment



As expected, given the low levels of exposure and effect observed in this study, the two predicted curves (shown in Figure 4.22) are very close together, and when statistical uncertainty is considered, provide indistinguishable fits to the data points.

Table 4.15 provides both confidence intervals and point estimates for the adjusted rate ratios associated with selected increments of radon exposure (treated as a continuous variable, and estimated in the 5 - 50 years prior to ascertainment) based on the exponential and linear models.

Table 4.15
Adjusted Rate Ratios [95% CI] Associated with Selected Increments of
Radon Exposure Based on Exponential and Linear Models

Model Form	Weighted Cumulative Radon Exposure		Weighted Average Radon Concentration	
	1 WLM	10 WLM	100 Bq/m ³	150 Bq/m ³
Linear	1.00 [0.96, 1.04]	1.03 [0.63, 1.43]	0.95 [0.86, 1.05]	0.93 [0.79, 1.07]
Exponential	0.99 [0.94, 1.06]	0.94 [0.52, 1.72]	0.98 [0.84, 1.14]	0.97 [0.77, 1.21]

Both the linear and exponential models provide estimated rate ratios for radon (expressed either in units of cumulative or average concentration of exposure) which are at once consistent with no effect, a small positive effect, or a small negative effect. A slightly narrower confidence interval was associated with the linear than the exponential model.

5 DISCUSSION

5.1 NOVEL ASPECTS OF THE ANALYSIS

The present analyses of the combined primary data from three North American case-control studies of residential radon and lung cancer provide the most valid and precise estimates to date of the effect of residential radon exposure on the rate of lung cancer. Although Lubin et al., (1994) analysed the combined primary data from three (in Shenyang China by Blot et al., 1990; New Jersey by Schoenberg et al., 1990; and Stockholm Sweden, by Pershagen et al., 1992) of the first four large published studies conducted in a residential setting, this earlier analysis had three main limitations: 1) it utilized two methods of radon exposure assessment which are known to induce bias in effect estimates and associated confidence intervals; the first method, time weighting average radon exposure in measured residences only, implicitly assumed zero exposure for missing periods; the second method cumulated radon exposure over a uniform exposure time window, but imputed missing radon levels by outcome-status-specific mean radon levels, accentuating differences between case and control series; 2) it included a study (conducted in China) which may have been irrevocably confounded by poorly measured and unusually high levels of indoor air pollution arising from heating and cooking sources; and, 3) it assumed a fixed effect of radon across sites (a relatively unimportant factor).

The present analysis employed a standardized and median-unbiased method [the observed control mean imputation (OCMI) method] to fill gaps in subjects' residential radon exposure history prior to calculating average or cumulative radon exposure in three time windows, and evaluating three alternative window weighting schemes. The use of a random effects model explicitly accounted for possible heterogeneity in the effect of radon exposure across study sites. Included in the present analysis is the one North American site that was included in

Lubin et al.'s 1994 combining effort, and two more recent North American studies (in Winnipeg Canada and Missouri USA, both published in 1994).

Comparability of site-specific estimates was promoted by: 1) developing a common statistical model, employing a common constellation of covariates across site and allowing for differences in the strength or direction of their coefficients in the analysis of the combined data; 2) utilizing a common method of imputation for missing exposure data; and, 3) evaluating the predicted exposure-response relationship per unit increase of radon exposure treated as a continuous variable.

Two sets of sensitivity analyses investigated questions not specifically examined previously in the context of residential radon exposure and lung cancer risk. The first evaluated the sensitivity of the results to three different exposure time windows which were defined *a priori* based on previous cohort studies of radon exposure in a mining environment. The second evaluated the fit of the data to two forms of exposure-response models: exponential, and linear (excess relative risk). Although Perhagen et al.(1992, and 1994) previously reported results for both types of models in their two Swedish studies of residential radon, the relative fit of the two model forms was not, to my knowledge, examined.

Another novel aspect of this analysis was the investigation of the potential for differential exposure ascertainment which can lead to 'information bias' (Kramer 1988). As indicated by Weinberg et al. (1996), "if the rate of missingness is markedly different for cases and controls, the error structure will be differential, and some bias will be expected, even under OCMP" and a linear model. Since a differential rate of missingness was observed in Winnipeg, (based on data given in Table 4.5, cases had a 27% greater probability of missing data than controls in the 5 - 15 year ETW, and a 20% greater probability in the 15 - 25 year ETW), the radon effect in this site would have been biased. As the mean radon exposure among control homes was slightly greater than that among case homes, the bias due to

imputation of the control mean would have resulted in a less negative estimate. The slight positive bias in the Winnipeg estimate was probably offset by a slight negative bias in New Jersey, due to imputation of a fixed value of income in this site.

Following restriction of subjects to those with a minimum of 75% complete radon data (thereby including only 31% of 738 cases), the original authors of the Winnipeg study (Létourneau et al., 1994) estimated an odds ratio equivalent to 1.08 [0.82, 1.48] per increment of 10 WLM in the time window 5 - 30 years prior to ascertainment. This compares to a present estimate of 0.83 [0.43, 1.61] for an increment of 10 WLM in the 5 - 25 years prior to ascertainment, using OCMI to impute missing data. Although the reduction in power attendant with a complete-subjects analyses resulted in Weinberg et al.'s (1996) inference that this method is inferior to OCMI under a linear model, the above data suggests that under an exponential model, particularly where there is differential exposure ascertainment, restriction of subjects may actually reduce uncertainty relative to OCMI. Given the increased sample size that will be available in the Phase II analysis, restriction of subjects to those with comparable completeness of radon histories may prove worthwhile.

5.2 FINDINGS REGARDING RADON EXPOSURE

5.2.1 ESTIMATION OF RADON EXPOSURE

Calculation of cumulative radon exposure in working level months (WLMs) involved the assumption of an equilibrium fraction (fixed at 0.5) and a home occupancy fraction. As described in section 3.6.5, home occupancy was assumed to be 80% for females, and 50% for males. In reality, home occupancy is known to vary both between subjects (by gender and cohort) and within subjects (by age and season), and probably varies somewhat between sites based on climatic conditions (mostly affecting weekend home occupancy). Subsequent to

the analyses, an ongoing case-control study in Iowa (which features the most sophisticated exposure estimation to date) estimated a mean home occupancy among females which varied by age, ranging from a low of 69.4% among those 50 - 59, to a high of 81.6% among those over 80 years of age (Field et al. 1998), with an overall average for their population of 72.1%. (The latter estimate closely agreed with the finding of the 1996 National Human Activity Pattern Survey that US women spend an average of 72.5% of their time in their residence.) The Iowa subjects were recruited between 1993 and 1996, a decade later than in the studies analysed herein; as female home occupancy would be expected to have been greater in the earlier decade corresponding the present studies, assuming similar age-distributions at recruitment, the 80% estimate employed herein for females may have been fairly accurate as an average value for the default exposure period between the 1930's and 1980's. Unfortunately, given the absence of an observed effect using either cumulative or average concentration to express radon exposure, it is not possible to make any conclusions about the validity or utility of the gender-differential assumptions with respect to occupancy factor. The uncertainty implicit in the gender-specific home occupancy assumption has impact only on the external validity (comparability to the miner studies) of the results in units of WLMs, however.

The use of *a priori* weights based for radon exposure in the time windows 5 - 14, 15 - 24, and 25 - 50 years prior to ascertainment, based on the BEIR VI analysis (NRC 1988), ensures direct comparability (within the constraints of occupancy uncertainty) with the more precise results from occupational studies. In addition, the adoption of *a priori* weights for radon exposure based on the miner studies enables the present analysis to build on the greater power of the occupational results to accurately determine the importance of exposure in each time window, particularly as there is no reason to believe that the etiological relevance of exposures received at differing times prior to exposure would vary across exposure settings (mines versus homes), independent of variations in subject age at exposure. Given the observed lack of sensitivity of the present results to the weighting factors (BEIR VI based,

versus two other alternatives which effectively limit the ETW), however, it may be advantageous in the next, Phase II analysis, which will adding four data-sets to the three analysed herein, to complement the results generated using the *a priori* weighting scheme with results in which separate weights for each time window are derived from the data.

5.2.2 MAIN RESULTS

Category-specific results (shown in Figure 4.31 for radon expressed as average concentration, and Figure 4.32 for radon expressed in cumulative WLMs), provided no evidence of a trend in any site taking into account statistical uncertainty. Combined interval estimates of rate ratio per unit exposure (either per 100 Bq/m³ weighted average concentration or per 10 cumulative WLMs) were centred around the null value of 1, regardless of model form (exponential or linear). No significant between site heterogeneity was observed in the radon effect. The confidence interval for the combined logistic estimate was only marginally wider under a random effects model than under a fixed effects assumption. The estimated rate ratio (RR) per increment of 10 weighted WLMs of radon over the 5 - 50 years prior to ascertainment, was 0.94 [0.52, 1.72] assuming an exponential model, and 1.03 [0.63, 1.43] assuming a linear model. Per increment of 100 Bq/m³ weighted average radon concentration over the 5 - 50 years prior to ascertainment, the RR estimates were 0.98 [0.84, 1.14] under an exponential model, and 0.95 [0.86, 1.05] under a linear model. Statistical model form did not materially affect the observed results, nor did modification of the time window weighting scheme to successively exclude the more historically distant exposure periods in which the proportion of imputed data was greater. These results indicate that the absence of a effect of radon was not attributable to either incorrect model form or imprecise specification of the exposure time window. Other uncertainties which could have resulted in false negative results are discussed in section 5.4 below.

5.2.3 CONSISTENCY OF THE RESULTS WITH ESTIMATES BASED ON MINER COHORTS

Estimates of ERR per WLM from the analyses by Lubin et al., (1994a) of 11 miner cohorts ranged from 0.2% to 5.1%, with a combined estimate of 0.5%, (without considering heterogeneity). The confidence interval for the current estimate of ERR per WLM based on the three published North American case-controls studies of residential exposure [-6%, 6%] includes the point estimates from the miner cohorts. In Sweden, and Stockholm, Sweden, residentially based estimates of ERR/WLM were 1.3% (derived from Pershagen et al. 1994) and 2.7% (Perschagen et al. 1992), respectively.

The BEIR VI committee (NRC 1998) estimated lifetime relative risks of lung cancer for lifetime indoor exposure to radon in units of average exposure concentration in eight separate strata [with respect to gender, ever-smoking status, and the two preferred ERR models which accounted for exposure rate effects]. These estimates differed most between ever and never-smokers, and assumed a sub-multiplicative relationship between smoking and radon. The estimated relative risks among ever-smokers ranged from 1.21 to 1.35, and among never-smokers from 1.52 to 1.82 for an average lifetime exposure of 100 Bq/m³, compared to the present overall estimates of 0.93 [0.79, 1.07] based on the three North American sites, for exposure in the 5 - 50 years prior to ascertainment, and assuming the linear ERR model. The studies included herein recruited inadequate numbers of never-smoking cases (14%, 74%, 3% in New Jersey, Missouri, and Winnipeg, respectively) to form meaningful strata across sites by smoking status. Stratification was, however, completed by recency of smoking (ie. among current versus ex-smokers, and adjusting for smoking intensity in reference to never-smokers) but was relatively uninformative due to large confidence intervals; the combined (across sites, assuming random effects) radon odds ratio was 0.87 [0.41, 1.85] among ex-smokers and 1.54 [0.69, 3.42] among current smokers. The Phase II analysis will benefit from increased precision to investigate both smoking- and gender-specific estimates effects

of radon, by virtue of the four additional (ongoing) studies from which data is expected to become available in the near future.

Occupational mining data may overestimate the level of risk associated with radon exposure because of confounding by concomitant exposures to other carcinogens present in mines. These agents may include: arsenic (Taylor et al. 1989; Xiang-Zhen et al., 1993), silica (Amandus et al. 1991), diesel and blasting fumes (IARC 1989), and, of most recent concern, mycotoxins (Scram et al. 1993; Venitt and Biggs 1995). Lubin et al. (1994) observed that adjustment for other mine exposures, such as arsenic, consistently reduced the estimates of radon-related risk. The neglect or imprecise measurement of one or more of these factors would therefore be expected to result in an overstatement of risk. Although it is possible that the radon effects observed in all 11 miner cohorts are over-estimates, the extent of bias would be expected to vary across mining sites, as does the presence of these agents. The variable presence of other mine carcinogens may account, at least in part, for significant but unexplained heterogeneity amongst site-specific mining estimates. Given the substantial experimental evidence from cellular mutagenicity studies and research in several animal species (IARC 1988, NRC 1994, Gilbert et al., 1996), and significant exposure-response trends observed within miner cohorts, however, variable degrees of (residual) confounding in the latter occupational studies, although adding uncertainty to the associated numerical estimates of risk, is insufficient to rule out a causal inference.

5.3 META-ANALYSIS VERSUS ANALYSIS OF THE COMBINED PRIMARY DATA

5.3.1 COMPARABILITY OF RESULTS

The question of whether a meta-analysis yields results different from that of an analysis of

the combined primary data is an important one in view of the current expansion of the epidemiologic practice of meta-analysis. This question will be addressed from both a theoretical and an empirical perspective. As cited by Olkin and Sampson (1998), Jeng et al. (1995) and Stewart and Parmar (1993) have shown for particular examples that empirical results can differ in conclusions depending on the pool of studies selected or the effect measures employed. According to Steinberg et al. (1997), when the pool of studies is essentially identical and the same effect measures employed, the use of summary level results in a meta-analysis yields results which agree closely to an analysis of the primary data. Olkin and Sampson (1998) demonstrate mathematically that a two way analysis of variance model, adding a term for site to the treatment indicator to analyse the combined primary data from several studies, results in exactly the same estimate of treatment effect (even in the unbalanced case) as a meta-analytic synthesis of the estimates derived from the separate one way ANOVA's. Their abstract concludes that "because a meta-analysis of individual patient data can be considerably more costly in terms of data retrieval than a meta-analysis of summary data, this equivalence provides for cost-efficient analysis". Olkin and Sampson's fundamental point was that meta-analysis suffices to determine the efficacy of 'treatments' versus controls, when those 'treatments' have continuous outcomes, and the studies are homogeneous.

Several points warrant emphasis, however: 1) The potential foci for meta-analyses is, however, broader than continuous outcomes among the 'patients' from randomized clinical trials (RCTs), which their methods and terminology subsume; 2) Olkin and Sampson indicate, their results assume no interaction; a single site variable accounts for random variation between (fixed) site effects. Set aside was the issue of random effects. In the present analysis an interaction between site and the exposure effect allowed for potential differences in exposure effect between site; several other between-site differences were allowed; 3) by contrast with least squares estimation in an ANOVA of continuous experimental outcome data, maximum likelihood estimation in the logistic regression model

employed to analyse dichotomous outcome data from case-control studies requires the inclusion of all strong risk factors in order to provide an unbiased estimate of exposure effect (Gail et al., 1994); 4) a case-control study is vulnerable to differential exposure ascertainment, which must be investigated; 5) Olkin and Sampson's results rely on homogeneity of the original studies; several analytic features differed amongst the original radon publications; as an example, gaps in radon exposure histories were dealt with in an inconsistent and bias-inducing manner; 6) as also noted by Olkin and Sampson, analyses of pooled primary data permit other study features to be explored; in the present analysis, the sensitivity of results to exposure time window definition and statistical model form was investigated.

How do the results of the present analysis of primary data compare with results based only on the published secondary data ? Lubin and Boice's (1997) meta-analysis of all eight internationally published case-control studies estimated an overall RR of 1.14 [1.0, 1.3] for an increment of 150 Bq/m³ in average radon concentration, based on an exponential model. The authors of the latter analysis derived site-specific estimates of radon effect by fitting a linear model to the logarithms of the originally reported odds ratios for radon concentration exposure category means or medians. I combined the secondary effect estimates thus derived and published by Lubin and Boice (1997) for New Jersey, Winnipeg and Missouri, under both a fixed and random effects assumption (by the methods detailed in section 3.7.1). This meta-analysis of the three studies yielded fixed and random effects RR estimates of 1.03 [0.93, 1.13], and 1.15 [0.89, 1.48] respectively, per increment of 150 Bq/m³ average radon concentration. The present (exponential) analysis of the primary data from the same three studies yielded fixed and random effects RR estimates of 0.97 [0.77, 1.21], and 0.96 [0.78, 1.17], respectively, for the same increment in average radon concentration (during the 5 - 50 years prior to ascertainment using common time window weights based on those derived by the BEIR VI committee). Importantly, whereas the meta-analysis of the original summary data indicated significant between-study variability, as reflected in the difference between the

fixed and random effects estimates, no such variability was observed when the primary data was re-analysed according to standard criteria. The meta-analysis thus evidenced spurious heterogeneity in radon effect between sites. Moreover, the first thesis objective of improving comparability of site-specific risk estimates through a standard approach to the analysis was achieved.

5.3.2 ADVANTAGES OF PRIMARY OVER SECONDARY DATA

In summary, an analysis of the combined primary data from several similar investigations has several advantages over a meta-analysis of the reported summary-level results of the same investigations. Access to the primary data enables the collaborating investigator to:

- i. Define common criteria for the eligibility of subjects (as was depicted in Figures 3.1 through 3.3);
- ii. Analyse the exposure data in its original continuous units, avoiding the problems of defining arbitrary categories, particularly when appropriate-sized reference groups can not be formed across sites;
- iii. Undertake a standardized assessment of the relative potential for selection and information bias;
- iv. Define a common exposure time window, and examine the sensitivity of results to varying assumptions about the etiologically relevant period of exposure;
- v. Employ a common method for imputing missing data;
- vi. Ensure that a common set of *a priori* risk factors and effect modifiers are examined and treated according to consistent criteria;
- vii. Provide and combine modifier-specific estimates of effect, especially since original

investigators may not have had the advantage of recently published theoretical or empirical evidence as to potential modifiers, or may have had inadequate data within strata to detect interactions;

- viii. Examine the sensitivity of the results alternative statistical model forms); and,
- ix. Answer all or many of the inevitable what-if questions that arise when reviewing primary studies that took different analytic approaches, did not report all of the covariates fitted.

In short, an analysis of the combined primary data from a number of similar observational investigations has the potential for greater internal and external validity than a meta-analysis of the reported summary-level results from the same investigations.

The analysis of the primary data enabled detection of differential exposure ascertainment as a potential source of attenuating bias in the Winnipeg site. The original paper reported comparisons of the relative completeness of case and control exposure measurement only within the select set of subjects who had at least 75% of data available within the 5 - 15 or 5 - 30 years prior to ascertainment.

The undertaking of a thorough systematic analysis aimed at deriving a standard model of general validity revealed or confirmed broader findings with respect to lung cancer risk factors than was possible in the original analyses (see Table 4.13). Particularly in the Winnipeg site, factors which were observed to be significant lung cancer predictors (e.g. ethnicity and family income) or important matching variables (e.g. gender) in the present study were not included in the final model reported by the original authors. The absence of such covariates in the original Winnipeg analysis may be attributable to a difference in analytic criteria and/or to more precise control of tobacco smoking. The original investigators used a 'material' change in radon estimate as their criterion for selecting the

most parsimonious model to report. Given the non-significance of radon, a fairly simple model sufficed. The present analysis took a more conservative approach, seeking first to build an optimal model for the prediction of lung cancer risk before considering the possible additional effect of radon. With respect to treatment of tobacco smoke exposure, whereas the original Winnipeg analysis adjusted only for ever-never smoking status, the present analysis included log(intensity), recency, the interaction between these two terms, and age at initiation, (all of which were considered relevant only within the period 5 - 50 years prior to ascertainment).

5.3.3 MAIN LIMITATIONS OF REVIEW METHODS

In the context of the most common focus for meta-analysis, the randomized clinical trial, Thacker (1988) cautioned that like all review methods, meta-analysis can be limited by sampling bias, inadequate data, and biased outcome interpretation. With respect to the latter, the analogous concern in a retrospective case-control study is for exposure information bias. The same concerns are present in an analysis of combined primary data, although the availability of the originally collected data provides a means for the standardized assessment and control of measured selection factors. Each of the above threats to the internal and external validity of quantitative research syntheses will be briefly examined.

Sampling or publication bias can affect both meta-analyses and re-analyses of primary data. As mentioned in section 3.2, I am aware of only one study of residential radon exposure which was unpublished; this small study was conducted in Florida, in a highly mobile population that had low radon levels; the data was considered of little information value. It is possible, but unlikely, that other studies have been completed but unpublished; roughly half of the studies published to date have interpreted their results as negative. Perhaps because of the large potential public health impact, there is no evidence of any reluctance for authors to submit or editors to reject "negative" studies. A publication bias does not appear

to exist. The Phase II analysis of all seven North American studies, (on completion of the final four), will, nonetheless, provide conclusions with greater generalizability.

Inadequate data to form complete radon home histories is a substantial obstacle in studies of residential radon exposure. The potential for exposure ascertainment bias to have resulted in a biased estimate of effect in Winnipeg has been mentioned previously. Although measured covariates (see Appendix D) such as education and income did not appear to be correlated with measured radon exposure, if socio-economic correlates are *truly* associated with radon exposure, but are better measured than radon exposure, radon slopes would be attenuated. Moreover, the lack of data on income in New Jersey, and the use of an assumed fixed income value for this site in an analysis of the combined data slightly attenuated the radon estimate in this site. The use of site-specific models to derive radon slopes for combination is recommended in part to avoid this problem in the Phase II analysis. Inadequate data in strata formed by suspected modifiers of a putative radon effect, namely smoking status, gender, and age, is a limitation which could not be overcome in the present analysis, and which is anticipated to be only marginally alleviated in the Phase II combination of all seven North American studies.

An issue which has not been previously mentioned in the analysis of the combined primary data from case-control studies is how to deal with differences between sites in the ratios of cases and controls. This is an issue only if the data from all sites are pooled in one analysis as was done herein. Although an offset could be used, whose parameter value is known and need not be estimated during the regression, because interactions with the site variable were employed (to estimate site-specific coefficients for radon and some covariates) the main effect of site was simply retained in the model. Based on simulations conducted by JA Hanley (personal communication), the inclusion of site as a main effect rather than using an offset does introduce a small amount of noise into the data. The combination of radon estimates derived from site-specific analyses, as recommended for the Phase II analysis, will

eliminate this source of random error.

5.4 UNCERTAINTIES SPECIFIC TO RESIDENTIAL RADON

Sources of uncertainties and variability affecting the interpretation of case-control studies of residential radon and lung cancer were discussed in detail in Chapter 2. Table 5.1 provides a hierarchical classification of uncertainties in the assessment of lung cancer risk as a function of residential radon exposure, identifies those uncertainties which were addressed (denoted by $\sqrt{}$) or not (denoted by $\times$) in the thesis, and indicates by subjective grading (by the number of marks), the relative importance of each. For those issues not addressed, I have also indicated the anticipated direction of their impact on current estimates of lung cancer risk as a function of residential radon exposure.

Table 5.1

**Sources of Uncertainty in the Assessment of Lung Cancer Risk
as a Function of Residential Radon Exposure**

Source of Uncertainty	Addressed in Thesis ?	Direction of Effect on RR
Statistical Uncertainty	√√√√	
Model Uncertainty:		
Etiologically relevant exposure time window	√	
Form of the exposure-response relationship		
Form of the statistical model relating radon and lung cancer	√	
Modifiers of the relationship between exposure and dose:		
Physical Factors:		
Activity-weighted particle size distribution & radon equilibrium		XX 1
Biologic modifiers or their correlates ¹	√	
Adequate treatment of lung cancer predictors other than radon:		
Cigarette smoking: intensity, recency, and age at initiation	√√	
Gender	√	
Ethnicity	√	
Income	√	
Education	√√	
Age	√	
Environmental tobacco smoke exposure	√	
Other environmental exposures (occupation/air pollution)		X ?
Nutritional factors		X ?
Histologic specificity ?		X 1
Exposure Assessment Uncertainty:		
A. Errors related to measurement of radon indoors:		
Measurement error in the α -particle track device		X 1
Spatial uncertainties:		
Measurement at one fixed location in a room		XX 1
Measurements limited to one or two rooms only		XX 1
Temporal uncertainties:		
Mis-specification in the duration of monitoring		X 1
Mis-specification in the duration of residence		X 1
Use of contemporary measurements to characterize past levels, including:		
Structural changes to homes		XX 1
Meteorological variability		X 1
Differences in ventilation habits of the subject versus the current resident		X +/-
B. Sources of error in the estimation of total exposure:		
Model uncertainties:		
Method of dealing with gaps in exposure measurement	√√√√	
Exclusion of exposures outside the subject's home		XXX 1
Uncertainty in percentage occupancy	√√√√	
Inter and intra-individual variation in percentage occupancy		XXXX 1

1 Correlates of biologic mediators which were investigated include: age; gender; tobacco smoke.

- × Indicates a factor not addressed; the relative number of marks indicates the relative importance of the factor.
✓ Indicates an addressed factor; the relative number of marks indicates the relative importance of the factor.

The literature review in chapter two concluded with a quote from Lubin (1995a) who judged that "the degree of consistency between studies reported to date is compatible with the consequences of population mobility, mis-specification of cumulative radon-progeny exposure, and the inability to measure radon levels accurately in past residences". Population mobility is not included as a separate uncertainty in Table 5.1; its effects are to increase the probability of missing exposure data, and to even out the radon exposure in the population, thus reducing the probability of detecting differences in exposure. In closing my literature review I added to these exposure-related uncertainties other sources of uncertainty which further support Lubin's assertion. These related to the empirical induction period (what I have since termed the etiologically relevant exposure time window), statistical model form, covariate control, differences among study populations, and the predominance of, and inability to accurately separate out the effects of, active and passive smoking. To the extent that was practical, I have specifically addressed these "model uncertainties" and comparability issues herein. Unmeasured covariates, such as nutritional status and other environmental or occupational agents (e.g. air pollution, discussed in sections 2.5.2.6 and 3.6.4.7) are a remaining source of concern. Such unmeasured exposures to occupational or environmental carcinogens would only affect radon risk estimates if such exposures had a large relative effect (which is not supported by previous radon studies), or were correlated or synergistic with that of radon. The possibility of histologic specificity for lung cancer was not investigated, as the *a priori* evidence evaluated in section 2.5.2.8 was considered weak. If radon did produce histologically specific lung cancers, then the effect estimated from all lung cancers would be an underestimate.

One of the largest sources of uncertainty, statistical uncertainty, has been not only reduced but cautiously represented through the random effects combining of statistical information from three study sites. It should be noted that sources of error relating to exposure

assessment, in addition to attenuating estimates of radon effect, will result in an underestimation of their variability. Although the present analysis has been unable to address what I have termed type “A” exposure uncertainties, that is, those related to the actual measurement of radon indoors, (relatively small dosimeter error, and moderate spatial and temporal uncertainties), two of the four large type “B” exposure uncertainties, those related to the estimation of total exposure, have been addressed. A major improvement over the primary studies has been the use of a preferred and uniform method for dealing with gaps in the exposure history which does not induce bias in the estimate of exposure-effect relationship. Calculation of cumulative exposure in WLMs included a point estimate of home occupancy. Inability to account for inter and intra-individual variation in percentage home occupancy, and the absence of time-dependent data on home radon equilibrium (an exposure-dose modifier) or outside radon exposure add a fairly large element of uncertainty and attenuating bias not reflected in estimated confidence intervals.

Model uncertainties with respect to the empirical induction period (or appropriate exposure time window), and form of the statistical model (exponential versus linear), have been ruled out as being responsible for the absence of an observed effect of radon. The treatment of covariates (including potential effect modifiers, potential confounders, and predictors of lung cancer other than radon), was thorough. While differences in the effect of recency of smoking resulted from the selection in one site of subjects who had been non-smokers in the 15 years prior to interview, and site-specific recency terms were included in the final model, these differences were non-significant. The absence of a confounding effect of smoking or any of the other covariates was evidenced by the lack of material difference between crude and adjusted results in the US sites, as well as the consistency of covariate coefficients across sites. Net negative confounding was suggested in Winnipeg, where a significantly negative crude odds ratio was shifted positively, becoming uncertain in direction following adjustment. The lack of net confounding in the other sites and the lack of association of radon levels with any of the covariates, suggests that differential selection of case and control

series (prior to interview) on a strong but measured lung cancer predictor, may have been responsible for the difference between crude and adjusted results in Winnipeg.

Whether an effect of radon might be specific to subgroups defined on the basis of potential modifiers suspected *a priori*, (including age at ascertainment, gender, smoking intensity, smoking recency, and environmental tobacco smoke exposure), was investigated. Although modification by recency and age was suggested, stratum-specific radon estimates demonstrated large and overlapping confidence intervals. These factors should be further investigated in the Phase II analysis, where precision to detect a radon effect, and potential modifiers thereof, will be improved.

While the large number of remaining uncertainties highlighted by Table 5.1 may tempt the reader to minimize the utility of the present undertaking, a large number of factors have been addressed, and a substantial reduction in uncertainty already accomplished. Simulation analyses conducted by Lubin et al. (1995) indicated that more than 7000 cases would be required to detect an effect as large as that observed in the BEIR IV analysis of data among miners with 90% power, assuming an alpha level of 0.05, an average mobility of three residential moves, and 'realistic' exposure measurement error. Given the availability of only 1590 cases, it is unsurprising that the present analysis has not replicated the effect seen in miners. The Phase II analysis, adding North American data from 3337 cases and 3928 controls will achieve a total sample size of 4927 cases and 6143 controls, coming close to the minimum requirements projected by Lubin et al. A number of important recommendations for this next analysis are provided below.

A European combining effort is in the planning stages, and lags approximately one year behind the North American endeavour. Although the number of subjects to be available from European studies is unknown at this time, it is hoped that the combined data to become available for an ultimate world-wide analysis will prove sufficient to make a conclusive

statement about the level of risk associated with residential exposure to radon. A summary and conclusions of the results of the present pilot analysis are provided in Chapter 6.

5.5 RECOMMENDATIONS FOR THE PHASE II ANALYSIS OF NORTH AMERICAN CASE-CONTROL STUDIES OF RESIDENTIAL RADON AND LUNG CANCER

The Phase II analysis will involve the addition of data from four studies currently in progress in Connecticut/Utah/South Idaho (1474 cases), Iowa (448 cases), Missouri (Phase II: 731 cases), and New Jersey (Phase II: 684 cases). The first of these recruited both males and females, and as in the Winnipeg study, controls were matched to cases by gender. The Connecticut/Utah/South Idaho study over-sampled non-smoking cases (all of the never-smoking cases and half of the ever-smoking cases were recruited). The latter three studies were restricted to females and did not place any restrictions on subject smoking status. Age at ascertainment was a standard frequency matching variable. The Phase II analysis will also enable the incorporation (in some sites) of unique data on modifiers of radon exposure-dose relationships, such as ventilation habits (i.e. sleeping with an open window, in Utah/South Idaho), activity-weighted particle size distribution, and equilibrium fraction (in Iowa). In consideration of the forgoing, the following recommendations are made for the Phase II analysis of the 7 North American case-control studies of residential radon and lung cancer:

1. The potential for selection bias and information bias should be carefully investigated in each site. Selection factors should be included in site-specific models. Estimates of combined effect should be provided including and excluding sites in which information bias is suspected.
2. Rather than deriving site-specific radon estimates from an analysis in which all of the

data are combined across sites (using interaction terms between site and radon), site-specific radon estimates should simply be derived from site-specific logistic regression analyses conducted in parallel.

3. Site-specific radon estimates should be combined using the 2 stage method for random effects.
4. A common core set of covariates should be included in each site-specific model, and augmented by covariates which may be available only in a subset of sites. The core set of *a priori* covariates should include smoking intensity (and its interaction with smoking recency and source of information), smoking recency, age, gender, ethnicity, education, and income (where available). The main effect of source of information should only be included in sites where it was a matching variable. Where such information is available, the importance of nutritional, occupational and other environmental exposures, after adjustment for education and income, should be investigated. Third order interactions between smoking intensity and recency with age and gender should also be considered.
5. Gender-specific estimates of radon effect should be derived by combining estimates of effect for males from the Winnipeg and Connecticut/Utah/South Idaho and for females from all sites. These estimates should also be adjusted (at minimum) for age, education, income, ethnicity, smoking intensity, smoking recency, and the interaction between the latter two variables.
6. Smoking-specific (never smokers, non-recent smokers, and current smokers) estimates of radon effect should be derived by combining estimates of effect from among the never-smokers in Missouri and Connecticut/Utah/South Idaho, from current smokers in all sites, and from ex-smokers in all sites. All estimates should be adjusted for age, gender, education, income, and ethnicity (at minimum). The estimates within smokers should be adjusted for intensity of smoking (at minimum). (There will likely be

insufficient data to simultaneously stratify by gender and smoking).

7. The potential for modification of the radon effect by other environmental exposures (air pollution and occupational carcinogen exposures), ventilation habits, and “ethnicity” should also be investigated, where possible.
8. Where available, the following factors should be embedded in the year-specific calculation of radon exposure: equilibrium fraction, activity-weighted particle distribution, home occupancy fraction, outside radon concentration, ventilation habits, spatial adjustments.
9. Missing radon data should be imputed using the OCMI method.
10. Given the widely divergent levels of minimum radon exposures present in the pilot analysis, a single common reference category will remain impractical; radon exposure should continue to be treated as a continuous variable in increments of 10 WLM and 100 Bq/m³.
11. The time window sensitivity analysis should be repeated, but should include only those subjects with a maximum of 10% imputed radon data within each time window (5-14, 15-24, and 25-50 years prior to ascertainment). The resulting site-specific radon estimates should then be combined across sites.
12. The time window sensitivity analysis (above) should be augmented with a separate analysis in which coefficients for each time window are fit to the data.
13. The possibility of histologic specificity of radon-induced lung cancer should be investigated.

6 SUMMARY AND CONCLUSIONS

6.1 SUMMARY

Radon is a gaseous intermediary in the radioactive decay chain of uranium-238. Because of the ubiquity of uranium-238 in the earth's crust, radon gas is present at varying levels as a naturally occurring contaminant of indoor air. One of a very small set of 'known human carcinogens', radon gas is the second most important cause of lung cancer. Based on US extrapolations from occupational epidemiological studies in underground miners it has been estimated that as many as 10-14 % of lung cancer deaths may be attributable to radon exposure (NRC 1998).

Because of the many sources of uncertainty involved in extrapolations of exposure-response relationships from mines to homes, case-control studies have been undertaken to directly estimate the risks associated with residential exposure to radon. Eight case-control studies involving estimation of residential exposure to radon gas have been published to date. Their individual results, although not inconsistent with earlier estimates from studies of miners, were not only statistically imprecise, but of questionable comparability due to differences in exposure time window, reference levels for exposure categories, methods used to impute radon estimates for the substantial proportions of missing exposure data, and investigation and treatment of important covariates (potential confounders and possible effect modifiers). Moreover, uncertainty remained concerning the appropriate form (linear versus logistic) of the exposure-response model. A re-analysis of the combined raw data from several studies was proposed to address these issues.

Following an international meeting of radon researchers in 1995, officials at Health Canada secured the cooperation of researchers from across North America to provide their original

data to a pooling exercise. This thesis constitutes the first of two pooling phases and utilizes primary data from three sites. Individual-level subject data was received from Winnipeg Canada, and New Jersey and Missouri USA, and compiled into a common format. Gaps in exposure history were imputed using the observed control mean imputation method. Estimated arithmetic mean (standard error) cumulative exposures for the period 5 to 50 years prior to subject ascertainment were 3 (0.03), 6 (0.07), and 12 (0.16) working level months for New Jersey, Missouri, and Winnipeg, respectively.

Parallel logistic regression analyses, employing over seventy distinct models and including four approaches for the control of tobacco smoke exposure, were undertaken to identify a single common multivariate model for the prediction of lung cancer risk which would adjust for all matching variables and risk factors determined to be important in at least one site. Radon exposure was expressed both in cumulative Working Level Months (WLMs) and as average concentration in Becquerels per cubic metre, with separate weighting factors (based on the BEIR VI analysis) applied to exposure in each of three distinct time windows (5 - 14 years, 15 - 24 years, and 25 - 50 years) prior to subject ascertainment. Three sets of weighting factors were compared. In addition to weighted radon exposure, the final logistic model included main effect terms for age, gender, ethnicity, education, family income, source of information (in New Jersey only), (log) cigarette smoking intensity, recency of cigarette smoking, age at initiation of cigarette smoking, and environmental tobacco smoke exposure. Two interactions with (log) smoking intensity were also included: 1) with recency of smoking; and, 2) with source of information.

Next, a combined analysis utilized a random-effects model to obtain the most realistic assessment to date of the effects of residential radon exposure. The combined logistic model allowed for site-specific differences in several factors: 1) radon exposure in the 5 - 50 years prior to ascertainment; 2) recency of smoking; 3) ethnicity; 4) family income; and, 5) source of information. The estimated rate ratio [and 95% confidence interval] was 0.94 [0.52, 1.72]

per increment of 10 WLMs cumulative radon exposure, and 0.98 [0.84, 1.14] per increment of 100 Bq/m³ weighted average radon concentration. No important differences were observed between site in the estimated rate ratio of lung cancer associated with radon exposure.

Exposure time windows 5 - 15, 5 - 25, and 5 - 50 years prior to ascertainment did not significantly differ in information value. Final estimates were, however, based on the default scenario, which included radon exposure information available from 5 - 50 years prior to ascertainment.

Sensitivity of the results to statistical model form (exponential logistic versus additive excess relative risk) was also investigated. The two alternative models provided indistinguishable fits to the data. Using a linear excess relative risk model, the estimated rate ratio [and 95% confidence interval] was 1.03 [0.63, 1.43] per increment of 10 WLMs cumulative radon exposure, and 0.95 [0.86, 1.05] per increment of 100 Bq/m³ weighted average radon concentration.

Twelve major recommendations were made regarding the conduct of Phase II of the North American pooling exercise, which will augment the present data-set with that from four ongoing studies, and provide improved precision to estimate lung cancer risk as a function of residential radon exposure.

6.2 CONCLUSIONS

There is ample evidence (as reviewed in Chapter 2) to support a causal relationship between lung cancer and alpha radiation arising from radon exposure at the levels typically experienced by occupationally exposed miners. It is plausible, however, that at the low doses typical of residential exposures, DNA repair mechanisms may be not only sufficient to repair

damage arising from alpha radiation, but may actually be sufficiently induced (stimulated) over their basal level by such exposure so as to provide broad spectrum protection from lung cancer induction by a variety of carcinogens. This would be analogous to the effect of an immunization, but with a broader spectrum response. Such a mechanism may explain why low levels of alcohol reduce cancer risk, by inducing the P450 mixed function oxidase system in the liver, which tends to detoxify rather than bioactivate other pro-carcinogens. Given the wide confidence intervals which span the null value of one estimated in the present analysis, and the many sources of uncertainty (summarized in section 5.4) which render these underestimates, a definitive conclusion is not yet warranted as to the magnitude of effect of residential levels of radon exposure on lung cancer. Because of the widespread population exposure to residential radon, however, it would be prudent to retain current guidelines for home radon exposure, based as these are on the occupational studies, even though the latter may overstate residential risk.

Seven main conclusions may be drawn as a result of this pilot study; these are outlined below:

1. Analysis of the combined primary data according to standard analytic criteria improved comparability of the site-specific estimates of radon effect; between study differences in the rate ratio of lung cancer as a function of radon were negligible or zero.
2. Analysis of the combined data reduced uncertainty in the estimation of radon effect over the results of the individual studies.
3. While not inconsistent with the estimates of effect observed in the cohort studies among miners, the confidence intervals for estimates of relative effect based on the combined data from the three published North American studies of residential radon were centred around the null value of 1.

4. Although present estimates of effect contain sources of exposure uncertainty which attenuate these estimates, the largest rate ratios consistent with the present analysis are 1.72 per increment of 10 weighted WLMs cumulative radon exposure, and 1.14 per increment of 100 Bq/m³ weighted average radon concentration, based on radon exposure estimated in the 5 - 50 years prior to ascertainment.
5. Imprecise specification of the exposure time window does not explain the absence of an observed effect of residential radon.
6. At the observed level of radon exposure, the exponential and linear (excess relative risk) models provided indistinguishable fits to the data.
7. Supplementation of the current data from 3 published North American studies with that from 4 others ongoing will increase statistical power to detect an effect of residential radon on lung cancer.

7 REFERENCES

- Abate T, 1992. Radon research typifies challenges in risk assessment. *The Scientist*, February 17: 13 - 14.
- Alavanja MCR, Brownson RC, Lubin JH, et al. 1994a. Residential radon exposure and lung cancer among nonsmoking women *Journal of the National Cancer Institute*, 86; 24: 1829 - 1837.
- Amandus HE, Shy C, Wing S, et al. 1991. Silicosis and lung cancer in North Carolina dusty trades workers. *American Journal of Industrial Medicine*, 20: 57- 70.
- Archer VE, 1995. Re: "Case-control study of residential radon and lung cancer in Winnipeg Manitoba Canada". *American Journal of Epidemiology* 142; 8: 884 - 885.
- Archer VE, 1997. Re: Residential radon and lung cancer in Sweden. *Health Physics*, 73; 2: 394 - 395.
- Armstrong BG, 1990. The effects of measurement errors on relative risk regressions. *American Journal of Epidemiology*, 132: 1176 - 1184.
- Auvinen A, 1996. Indoor radon and risk of lung cancer: a nested case-control study in Finland. *Journal of the National Cancer Institute*, 88: 966 - 972.
- Axelsson O, Sundell L, 1978. Mining, lung cancer and smoking. *Scandinavian Journal of Work and Environmental Health*, 4: 46 - 52.
- Axelsson O, 1980. Interaction between smoking and exposure to radon daughters. IN: W.

- Rom and V. Archer (eds.), *Health Implications of New Energy Technologies*. Society for Occupational and Environmental Health, Washington DC Page 23.
- Bale WF, 1980. Memorandum to the files, March 14, 1951: Hazards associated with radon and thoron. *Health Physics*, 38: 1062- 1066.
- Bartecchi CE, MacKenzie TD, Schrier RW, 1994. The human costs of tobacco use (first of two parts). *The New England Journal of Medicine*, 330; 13: 907 - 912.
- Bartlett DT 1985. What's in a name ? That which we call CR-39. *NRPB Radiation Protection Bulletin*, 64 20 21.
- Bartsch H, Hollstein M, Mustonen R, et al. 1996. Screening for putative radon-specific p53 mutation hotspot in German uranium miners. *The Lancet*, 346: 59.
- Becher H, 1992. The concept of residual confounding in regression models and some applications. *Statistics in Medicine*, 11; 1747 - 1758.
- Beral V, 1995. 'The practice of meta-analysis': discussion. *Meta-analysis of observational studies: a case study of work in progress. Journal of Clinical Epidemiology*, 48; 1: 165 - 166.
- Berlin JA, Longnecker LM, Greenland S, 1993. Meta-analysis of epidemiologic dose-response data. *Epidemiology*, 4: 218 - 228.
- Bierma TJ, Croke KG, Swartzman D 1989. Accuracy and precision of home radon monitoring and the effectiveness of EPA monitoring guidelines. *Journal of the American Pollution Control Association*, 39: 953 - 959.

- Blot WJ, Xu ZY, Boice JD, et. al. 1990. Indoor radon and lung cancer in China. *Journal of the National Cancer Institute*, 82; 12: 1025 - 1030.
- Bodansky D, 1987. Overview of the Indoor Radon Problem. IN: Bodansky D, Robkin MA, & Stadler DR, (eds.) *Indoor Radon and Its Hazards*. University of Washington Press. Seattle. Pages: 3 - 16.
- Bourbonnais R, Siemiatycki J. In Press Social class and risk of ten types of cancer in Montreal Canada.
- Brenner H, Loomis D, 1994. Varied forms of bias due to nondifferential error in measuring exposure. *Epidemiology*, 5: 510- 517.
- Breslow NE, Day NE, 1980. *Statistical Methods in Cancer Research, Volume 1 - The Analysis of Case-Control Studies*. International Agency for Research on Cancer, Publication 32, Lyon FR .
- Busigin A, van der Booren AW, Babcock JC, et al. 1981. The nature of unattached RaA (218Po) particles *Health Physics*, 40: 333.
- Cardis, E Gilbert, ES Carpenter, L et al. 1995. Combined Analysis of Cancer Mortality Among Nuclear Workers in Canada, the United Kingdom, and the United States of America. International Agency for Research on Cancer, Technical Report No. 25. Lyon FR.
- Chambers DB, Lowe LM, 1987. An assessment of the potential health risks form exposure to radon daughters. Report Prepared for Health and Welfare Canada by SENES Consultants Ltd., Toronto.

- Clayton D, 1991. The Eurodem collaborative re-analysis of case-control studies of Alzheimer's disease: some methodological considerations. *International Journal of Epidemiology*, 20; Suppl 2: S62 - S64.
- Cohen BL, 1989. Radon-222 levels in low income households. *Health Physics*, 56: 349 - 353.
- Cohen BL, 1991. Variations of radon levels in US homes correlated with house characteristics, location, and socioeconomic factors. *Health Physics*, 60: 631 - 642.
- Cole LA 1993. *Element of Risk: The Politics of Radon*. American Association for the Advancement of Science Press.
- Cook DJ, Sacket DL, Spitzer WO, 1995. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation of Meta-Analysis. *Journal of Clinical Epidemiology*, 48; 1: 167 - 171.
- Cross FT, 1992. A review of experimental animal radon health effects data. IN: Chapman, JD, Dewey, WC, Whitmore, GF (eds.) *Radiation Research: A twentieth-century perspective*. Volume II. Academic Press, San Diego CA. 476 481.
- Cross FT, Dagle GE, Gies RA, et al. 1992. Experimental animal studies of radon and cigarette smoke. IN: Cross, FT (ed.), *Indoor Radon and Lung Cancer: Reality or Myth ? Twenty-Ninth Hanford Symposium on Health and the Environment*, Richland WA., October 1990. Battelle Press, Columbus OH Pages: 821 - 844.
- Cross FT, 1994. Residential radon risks from the perspective of experimental animal studies. *American Journal of Epidemiology* 140: 333- 339.

- Damber L, Larsson LG, 1982. Combined effects of mining and smoking in the causation of lung carcinoma- case-control study in northern Sweden. *Acta Radiologica Oncology*, 21: 305- 313.
- De Villers AJ, Windish JP, 1964. Lung cancer in a flourospar mining community. *British Journal of Industrial Medicine*, 21: 94 - 109.
- De Villers AJ, Windish JP, Brent F de N, et al. 1971. Mortality experience of the community and of flourospar mining employees at St. Lawrence Newfoundland. *Occupational Health Reviews*, 22: 1 - 15.
- Doll R, Hill AB, 1950. Smoking and carcinoma of the lung: a preliminary report. *British Medical Journal* 2: 739 - 748.
- Doll R, 1971. The age distribution of cancer: implications for models of carcinogenesis. *Journal of the Royal Statistical Society A*, 132: 133 - 166.
- Doll R, Peto R, 1976. Mortality in relation to smoking: 20 years' observation on male British doctors. *British Medical Journal* 2: 1525 - 1536.
- Doll R, Peto R, 1978. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and life-long non-smokers. *Journal of Epidemiology and Community Health*, 32: 303 - 313.
- Doll R, Gray R, Hafner B, et al. 1980. Mortality in relation to smoking: 22 years' observations on female British Doctors. *British Medical Journal* i: 967 : 971.
- Dosemeci M, Wacholder S, Lubin JH, 1990. Does nondifferential misclassification of

- exposure always bias a true effect toward the null value ? American Journal of Epidemiology 132; 4: 746- 748.
- Duport P, Madelaine G, Zettwoog P, et al. 1977. Experimental study of the penetration of radioactive aerosols into the respiratory tract (Fr.). Chemosphere 1: 35- 40.
- Eddy D, Hasselblad V, Shachter R, 1992. The Confidence Profile Method. Academic Press.
- Edling C, 1982. Lung cancer and smoking in a group of iron ore miners. American Journal of Industrial Medicine 3: 191- 199.
- Edling G, Axelson O, 1983. Quantitative aspects of radon daughter exposure and lung cancer in underground miners. British Journal of Industrial Medicine 40: 182- 187.
- Farland WH, 1992. Similarities and differences between adults and children: Future directions and research needs. IN: Guzelian, PS, Henry, CJ, and Olin, SS, (eds.), Similarities and Differences Between Children and Adults: Implications for Risk Assessment,. International Life Sciences Institute Press, Washington DC. Pages: 273 - 278.
- Field RW, Smith BJ, Brus CP, et al., 1998. Retrospective temporal and spatial mobility of adult Iowa women, Risk Analysis, In Press.
- Fleiss JL, 1993. The statistical basis of meta-analysis. Statistical Methods in Medical Research, 2: 121 - 145.
- Gail MH, Wiend S, Piantadosi S, 1984. Biased estimates of treatment effect in randomized

experiments with nonlinear regressions and omitted covariates. *Biometrika* 71: 431-434.

Gilbert ES, Cross FT, Dagle GE, 1996. Analysis of lung tumor risks in rats exposed to radon. *Radiation Research*, 145: 350- 360.

Gorsek L, Humar M, Ilic R, et al., 1990. Indoor and outdoor radon measurements in the Saleska Valley. *Nuclear Tracks and Radon Measurements*, 17; 1: 73.

Grayson DA, 1987. Confounding confounding. *American Journal of Epidemiology*, 126: 3; 546- 553.

Greenland S, 1980. The effect of misclassification in the presence of covariates. *American Journal of Epidemiology*, 112: 4; 564- 569.

Greenland S, Robins JM, 1985. Confounding and Misclassification. *American Journal of Epidemiology*, 122: 3; 495- 506.

Greenland S, 1987. Quantitative methods in the review of the epidemiologic literature. *Epidemiology Reviews*, 9: 1 - 30.

Greenland S, 1989. Modeling and variable selection in epidemiologic analysis. *American Journal of Public Health*, 79: 3; 340- 349.

Greenland S, Longnecker MP, 1992. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *American Journal of Epidemiology*, 135: 1301 - 1309.

- Greenland S, Robins J, 1994. Invited commentary: Ecologic studies - biases, misconceptions, and counter-examples. *American Journal of Epidemiology*, 139: 8; 747- 760.
- Guimond RJ, Page SD, 1992. Indoor radon: A case study in risk communication. *Radiation Protection Dosimetry*, 42; 3: 159 - 176.
- Hammond EC, Horn D, 1958a. Smoking and death rates - report on 44 months of follow-up of 187,783 men. I. Total mortality. *Journal of the American Medical Association*, 166: 1159 - 1172.
- Hammond EC, Horn D, 1958b. Smoking and death rates - report on 44 months of follow-up of 187,783 men. II. Death rates by cause. *Journal of the American Medical Association*, 166: 1294 - 1308.
- Harley NH, Chittaporn P, Roman MH, et al., 1991. Personal and home 222-Rn and gamma-ray exposure measured in 52 dwellings. *Health Physics*, 61: 6; 737 - 744.
- Hegmann KT, Fraser AM, Keaney RP, et al. 1993. The effect of age at smoking initiation on lung cancer risk *Epidemiology* 4: 5; 444- 448.
- Hennekens CH, Buring JE, 1987. *Epidemiology in Medicine*. Little Brown and Co., Boston.
- Hobbs CH, McClellan RO 1986. Toxic effects of radiation and radioactive materials IN: (Klaassen CD, Ambdur MO, and Doull J, eds.) *Casarett and Doull's Toxicology: The Basic Science of Poisons*, Third Edition. McMillan Publishing Co., New York NY.

Holum JR 1979. **Elements of General and Biological Chemistry, Fifth Edition.** John Wiley and Sons Inc., New York .

Horgan J, 1994. Radon's risks. **Scientific American**, August: 15 - 16.

Hornung RW, Samuels S, 1981. Survivorship models for lung cancer mortality in uranium miners - is cumulative dose an appropriate measure of exposure ? IN: Gomez , M (ed.) **Radiation Hazards in Mining: Control, Measurement, and Medical Aspects.** New York: Society of Mining Engineers of the American Institute of Mining Metallurgical and Petroleum Engineers pages 363- 368.

Hornung RW, Meinhardt RW, 1987. Quantitative risk assessment of lung cancer in U.S. uranium miners. **Health Physics**, 52; 4: 417 - 430.

Hosmer DW, Lemeshow S, 1989. **Applied Logistic Regression.** John Wiley & Sons Inc., Toronto. 307 pp.

International Agency for Research on Cancer, 1989. Diesel and gasoline exhausts and some nitroamines IARC Monographs on the Evaluation of Carcinogenic Risks to Humans., Volume 46. Lyon FR.

International Agency for Research on Cancer (IARC), 1988. Man-made fibres and radon. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Lyon FR volume 43 pages 173- 241.

James AC, 1988. Lung dosimetry. IN: Nazaroff, WW and Nero, AV (eds.) **Radon and its Decay Products in Indoor Air.** John Wiley and Sons Inc., New York. Pages: 259-309.

- Jeng GT, Scott JR, Burmeister LE, 1995. A comparison of meta-analytic results using literature versus individual patient data. *Journal of the American Medical Association*, 274: 830 - 836.
- Jonassen N, Jenson B, 1989. Radon daughters in indoor air. Final Report to Vattenfall Technical University of Denmark. Lyng, Denmark.
- Jorgenson HS, 1973. A study of mortality from lung cancer among miners in Kiruna 1950 - 1970. *Work Environmental Health* 10: 126- 133.
- Kramer MF, 1988. *Clinical Epidemiology and Biostatistics: A Primer for Clinical Investigators and Decision Makers*. Springer Verlag, New York.
- Krewski D, Miller T, Meyerhof DP, et al., 1989. Managing environmental radon risks: A Canadian perspective. IN: L. Freij, (ed.), *Management of Risk from Genotoxic Substances in the Environment*. Swedish National Chemicals Inspectorate, Stockholm. 242 - 257.
- Kunz E, Sevc J, Placek V, et al., 1979. Lung cancer distribution in man in relation to different time distributions of radiation exposure. *Health Physics*, 36: 699 - 706.
- Kupper LL 1984. Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies. *American Journal of Epidemiology* 120: 4; 643- 648.
- L'Abbe KA, Detsky AS, O'Rourke K, 1987. Meta-analysis in clinical research. *Annals of Internal Medicine*, 107: 224 - 233.
- Lagarde F, 1992. Statistical aspects of the influence of indoor radon exposure and smoking

on the risk of lung cancer. Institute of Actuarial Mathematics and Mathematical Statistics, Report No. B:21. Stockholm: University of Stockholm.

Laird NM, Mosteller F, 1990. Some statistical methods for combining experimental results. *International Journal of Technology Assessment in Health Care*, 6; 9: 5 - 30.

Larsson S, 1971. Completeness and reliability of lung cancer registration in the Swedish Cancer Registry. *Acta Pathol. Microbiol. Scand. Section A* 79: 389.

Leech JA, Wilby K, McMullen E, et al., 1996. The Canadian Human Activity Pattern Survey: Report of Methods and Population Surveyed. Division of Air Quality Health Effects Research, Monitoring and Quality, and Biostatistics, Environmental Health Directorate, Health Canada.

Letourneau EG, Zielinski JM, Krewski D, et al. 1992. Levels of radon gas in Winnipeg homes. *Radiation Protection Dosimetry*, 45; 1/4: 531 - 534.

Letourneau EG, Krewski D, Choi NW, et. al. 1994. Case-control study of residential radon and lung cancer in Winnipeg Manitoba, Canada. *American Journal of Epidemiology*, 140: 4; 310- 322.

Letourneau EG, Krewski D, Goddard MJ, et al., 1995. The Authors Reply. *American Journal of Epidemiology*, 142; 10: 1121 - 1122.

Levi F, Franceschi S, La Vecchia C, et al. 1997. Lung cancer trends by histologic type in Vaud and Neuchatel, Switzerland, 1974 - 1994. *Cancer*, 79; 5: 906 - 914.

Leviton LC, Cook TD, 1981. What differentiates meta-analysis from other forms of review

? J. Pers. 49: 231 - 236.

Licciardone JC, Brownson RC, Chang JC, et al. 1990. Uterine cervical cancer risk in cigarette smokers. *American Journal of Preventative Medicine*, 6; 5: 274 - 281.

Litt BR, Waldman JM, Harley NH, 1991. Validation of a personal radon monitor for use in residential ²²²Rn exposure studies. *Health Physics* 61: 6; 727- 735.

Lo YMD, Darby S, Noakes L, et al. 1996. Screening for codon 249 p53 mutation in lung cancer associated with domestic radon exposure. *The Lancet*, 345: 59.

Louis TA, Fineberg HV, Mosteller F, 1985. Findings for public health from meta-analyses. *Annual Reviews of Public Health*, 6: 1 - 20.

Lubin JH, Blot WJ, Berrino F, et al., 1984. Modifying risk of developing lung cancer by changing habits of cigarette smoking. *British Medical Journal*, 288: 1953 - 1956.

Lubin JH, Samet JM, Weinberg C, 1990. Design issues in epidemiologic studies of indoor exposure to Rn and risk of lung cancer *Health Physics*, 59: 807- 817.

Lubin JH, 1994. Invited Commentary: Lung Cancer and Exposure to Residential Radon. *American Journal of Epidemiology* 140: 4; 323- 332.

Lubin JH, Boice JD, Edling C, et al. 1994b. Radon and Lung Cancer Risk: A Joint Analysis of 11 Underground Miners Studies. US Dept. of Health and Human Services, Public Health Service, National Institutes of Health, Publication No. 94-3644.

Lubin JH, Liang Z, Hrubec Z, et al. 1994a. Radon exposure in residences and lung cancer

among women: combined analysis of three studies *Cancer Causes Control*, 5: 114 - 128.

Lubin JH, Boice JD, Edling C, et al. 1995b. Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects. *Health Physics*, 69: 4: 1- 7.

Lubin JH, Steindorf K, 1995. Cigarette use and the estimation of lung cancer attributable to radon in the United States. *Radiation Research*, 141: 79- 85.

Lubin JH, Boice JD, Samet JM 1995a. Errors in exposure assessment, statistical power, and the interpretation of residential radon studies. *Radiation Research*, 144: 329- 341.

Lubin JH, Boice JD, Edling C, 1995c. Lung cancer in radon -exposed miners and estimation of risk from indoor exposure. *Journal of the National Cancer Institute*, 87: 11; 817- 827.

Lubin JH, Boice JD, 1997. Lung cancer risk from residential radon: Meta-analysis of eight epidemiological studies. *Journal of the National Cancer Institute*, 89; 1: 49 - 57.

Mann C, 1990. Meta-analysis in the breech. *Science*, 249: 476 - 480.

Marcinowski F, Lucas RM, Yeager WM, 1994. National and regional distributions of airborne radon concentrations in US homes. *Health Physics*, 66: 669- 706.

Martz DE, Rood AS, George JL, et al. 1991. Year-year variations in annual average indoor 222-Rn concentrations. *Health Physics*, 61: 3; 409 - 413.

Mason TJ, 1994. Descriptive epidemiology of lung cancer. IN: Samet JM, (ed.)

- Epidemiology of Lung Cancer. Marcel Dekker, New York NY. Pages: 51 - 69.
- McGregor RG, Vasudev P, Letourneau EG, et al. 1980. Background concentrations of radon and radon daughters in Canadian homes. Health Physics, 39: 285- 289.
- Miettinen OS, Cook EF, 1981. Confounding: essence and detection. American Journal of Epidemiology 114: 593 - 603.
- Monchaux G, Morlier P, Fritsch P, et al., In press. Co-carcinogenic effects in rats of various agents following exposure to radon and radon daughters.
- Moolgavkar SH, Venzon DJ, 1987. General relative risk regression models for epidemiologic studies. American Journal of Epidemiology, 126; 5: 949 - 961.
- Moolgavkar SH, Luebek EG, Krewski D, et al. 1993. Radon, cigarette smoke, and lung cancer: A re-analysis of the Colorado Plateau uranium miners' data. Epidemiology, 4: 3; 204- 217.
- Morrison HL, Wigle DT, Stocker H, et al. 1981. Lung cancer mortality and radiation exposure among the Newfoundland Flourospar miners. IN: Gomez , M (ed.) Radiation Hazards in Mining: Control, Measurement, and Medical Aspects. New York: Society of Mining Engineers of the American Institute of Mining Metallurgical and Petroleum Engineers, Pages 372- 376.
- Muller J, Wheeler WG, Gentleman JF, et al., 1983. Study of mortality of Ontario Uranium Miners 1955 - 1977. Part I. Toronto, Canada: Ontario Minstry of Labour.
- Muller J, Wheeler WC, Gentleman JF, et al., 1985. Study of the mortality of Ontario miners.

- IN: H. Stocker (ed.) Proceedings of the International Conference, Occupational Radiation Safety in Mining, October 14 - 18, 1984. Canadian Nuclear Association, Toronto. Pages 335- 343.
- Murray DM, Burmaster DE, 1995. Residential air exchange rates in the US: empirical and estimated parameteric distributions by season and climatic region. Risk Analysis, 15; 4; 459- 465.
- National Council on Radiation Protection and Measurements (NCRP), 1984. Evaluation of Occupational and Environmental Exposures to Radon and Radon Daughters in the United States. NCRP Report No. 78, Bethesda MD.
- Nero AV Jr, Schwehr MB, Nazaroff WW, et al., 1986, Distribution of airborne radon-222 concentrations in US homes Science, 234: 992- 997.
- Nero AV Jr, 1988, Radon and its decay products in indoor air: an overview. IN: Nazaroff, WW and Nero, AV (eds.) Radon and its Decay Products in Indoor Air. John Wiley and Sons Inc., New York. Pages 1- 56.
- Neuberger JS 1991. Residential radon exposure and lung cancer: an overview of published studies Cancer Detection and Prevention, 15: 6; 435- 443.
- Neuberger JS, 1992. Residential radon exposure and lung cancer: an overview of ongoing studies. Health Physics, 63: 503- 509.
- NRC (National Research Council), Committee on the Biological Effects of Ionizing Radiation (BEIR IV), Board on Radiation Effects Research, 1988. Health Risks of Radon and Other Internally Deposited Alpha-Emitters. National Academy Press,

Washington DC.

- NRC (National Research Council), Committee on the Biological Effects of Ionizing Radiation (BEIR V), Board on Radiation Effects Research, 1990. Health Effects of Exposure to Low Levels of Ionizing Radiation. National Academy Press, Washington DC.
- NRC (National Research Council), Committee on the Biological Effects of Ionizing Radiation (BEIR VI), Board on Radiation Effects Research, 1994. Health Effects of Exposure to Radon: Time for Reassessment ? National Academy Press, Washington DC.
- NRC (National Research Council), Committee on the Biological Effects of Ionizing Radiation (BEIR VI), Board on Radiation Effects Research, 1998. Health Effects of Exposure to Radon. Prepublication Copy: Uncorrected Proofs. National Academy Press, Washington DC.
- NRC (National Research Council), Panel on Dosimetric Assumptions Affecting the Application of Radon Risk Estimates, Board on Radiation Effects Research, 1991. Comparative Dosimetry of Radon in Mines and Homes. National Academy Press, Washington DC.
- NRC (National Research Council), Panel on Statistical Issues and Opportunities for Research in the Combination of Information, Committee on Applied and Theoretical Statistics, 1992. Combining Information: Statistical Issues and Opportunities for Research. National Academy Press, Washington DC.
- O'Brien C, 1996. Are pylons and radon a lethal cocktail ? New Scientist, February 17, Page

4.

Olkin I, Sampson A, 1998. Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics*, 54: 317 - 322.

Papastefanou C, Stoulos S, Manolpaulou M, 1994. Indoor radon in Greek apartment dwellings. *Health Physics*, 66: 270 - 273.

Pathak, B, 1989. Radon in Buildings, Canadian Centre for Occupational Health and Safety Publication No. P89-3E Hamilton ON .

Pershagen G, Liang ZH, Hrubec Z, et. al., 1992. Residential radon exposure and lung cancer in Swedish women. *Health Physics*, 63: 2; 179- 186.

Pershagen G, Akerblom G, Axelson O, et. al., 1994. Residential radon exposure and lung cancer in Sweden *New England Journal of Medicine*, 330; 3: 159 - 164.

Peto R, Doll R, 1984. Keynote address: the control of lung cancer. IN: Mizell M and Correa P (eds.) *Lung Cancer: Causes and Prevention*. Verlag Chemie International. pages: 1 - 19.

Piao CQ, Hei TK, 1993. Radon, cigarette smoke and oncogenic transformation. *Carcinogenesis*, 14; 3: 497- 501.

Piechowski JW, Gac LE, Brenot J, et al., 1981. Exposure to short-lived radon daughters: comparison of individual and ambient monitoring in a french uranium mine. IN: Gomez , M (ed.) *International Conference: Radiation Hazards in Mining: Control, Measurement, and Medical Aspects*, October 4 - 9, 1981. New York: Society of

Mining Engineers of the American Institute of Mining Metallurgical and Petroleum Engineers pages 538- 548.

Porstendorfer J, Robig G, Ahmed A, 1979. Experimental determination of the attachment coefficients of atoms and ions on monodisperse particles. *Journal of Aerosol Science* 10: 21- 28.

Puskin JS, 1992. An analysis of the uncertainties in estimates of radon-induced lung cancer. *Risk Analysis*, 12; 2: 277- 285.

Rickert WS, 1983. "Less hazardous" cigarettes: fact or fiction ? *NY State Journal of Medicine*, 83: 269 - 1272.

Robins JM, Greenland S, 1986. The role of model selection in causal inference from nonexperimental data. *American Journal of Epidemiology*, 123: 392 - 402.

Robkin MA, 1987. Terminology for describing radon concentrations and exposures IN: Bodansky D, Robkin MA, & Stadler DR, (eds.) *Indoor Radon and Its Hazards*. University of Washington Press. Seattle. Pages 17- 29.

Roggli VL, Vollmer RT, Greenberg ST, et al., 1985. Lung cancer heterogeneity: a blinded randomized study of 100 consecutive cases. *Human Pathology*, 16: 569- 578.

Rothman KJ, 1981. Induction and latent periods *American Journal of Epidemiology*, 114; 2: 253- 259.

Rothman KJ, 1986. *Modern Epidemiology*. Little, Brown & Co., Toronto.

- Rothman KJ, Poole C, 1988. Strengthening programme for weak associations. *International Journal of Epidemiology* 17 (suppl): 955- 959.
- Ruosteenoja E, 1991. Indoor Radon and Risk of Lung Cancer: An Epidemiological Study in Finland. Doctoral Dissertation. Department of Public Health, University of Tampere. Published by the Finnish Government Printing Centre, Helsinki Finland.
- Saba SR, Espinoza CG, Richman AV, et al., 1983. Carcinomas of the lung: An ultrastructural and immunocytochemical study. *American Journal of Clinical Pathology*, 80: 6 .
- Samet JM, 1989. Radon and lung cancer. *Journal of the National Cancer Institute*, 81: 745-757.
- Samet JM, Nero AV Jr, 1989. Indoor radon and lung cancer. *New England Journal of Medicine*, 320: 591- 593.
- Samet JM, Hornung RW, 1990. Review of radon and lung cancer. *Risk Analysis*, 10: 65-75.
- Samet JM, Lambert WE, 1991. Epidemiologic approaches for assessing health risks from complex mixtures in indoor air. *Environmental Health Perspectives*, 95: 71- 74.
- Samet JM, 1992. Indoor radon and lung cancer: estimating the risks. *The Western Journal of Medicine* 156 25 29.
- Samet JM, 1994. Indoor radon and lung cancer: risky or not ? *Journal of the National Cancer Institute*, 86: 1813- 1814.

- Samet JM, 1997. Indoor radon exposure and lung cancer: Risky or not ? - All over again. *Journal of the National Cancer Institute* 89; 1: 4 - 7.
- SAS Institute Inc., 1988. SAS /STAT User's Guide, Version 6, Fourth Edition, Volume 2, Chapter 27: The logistic procedure. SAS Institute, Cary NC. 1848 pp. .
- Schoenberg JB, Wilcox HB, Mason TJ, et. al., 1989. Variation in smoking-related lung cancer risk among New Jersey Women. *American Journal of Epidemiology*, 130: 688- 695.
- Schoenberg JB, Klotz JB, Wilcox HB, et. al., 1990. Case-Control Study of Residential Radon and Lung Cancer among New Jersey Women. *Cancer Research*, 150: 6250-6254.
- Schoenberg JB, Klotz JB, Wilcox HB, et. al. 1992. A case-control study of residential radon and lung cancer among New Jersey women. IN: Cross, FT (ed.), *Indoor Radon and Lung Cancer: Reality or Myth ? Twenty-Ninth Hanford Symposium on Health and the Environment*, Richland WA, October 1990. Battelle Press, Columbus OH Pages: 907 - 915.
- Schuman LM, 1977. Patterns of smoking behavior. IN: Jarvik, ME, Cullen JW, Gritz, FR, et al., (eds.), *Research on smoking behavior*. NIDA Research Monograph 17, US Dept. of HEW, Public Health Service, National Institute on Drug Abuse. DHEW Publication No. (ADM) 78-581. Pages: 36 - 65.
- Schwartz AG, Swanson GM, 1997. Lung carcinoma in african americans and whites: A population-based study in Metropolitan Detroit, Michigan. *Cancer*, 79; 1: 45 - 52.

- Scott AG, Robertson A, 1991. Precision of alpha track radon exposure estimates determined from field measurements. *Health Physics*, 61; 2: 267- 269.
- Scram RJ, Bokova B, Dobias L, et al. 1993. Monitoring genotoxic exposure in uranium miners. *Environmental Health Perspectives*, 99: 303 - 305.
- Sevc J, Placek V, Jerabek J, 1971. Lung cancer risk in relation to long-term radiation exposure in uranium miners. IN: *Proceedings of the 4th Conference on Radiation Hygiene*. Parzska, Lhota, CSSR. Pages: 315 - 326.
- Sexton K, Gong H, Bailar JC, et al., 1993. Air pollution health risks: do class and race matter ? *Toxicology and Industrial Health*, 9; 5: 843- 878.
- Shopland DR, Eyre HJ, Pechacek TF, 1991. Smoking-attributable lung cancer mortality in 1991: Is lung cancer now the leading cause of death among smokers in the United States ? *Journal of the National Cancer Institute*, 83: 1142 - 1148.
- Siemiatycki J, 1991. *Risk Factors for Cancer in the Workplace*. CRC Press. Boca Raton, FL.
- Siemiatycki J, Dewar R, Krewski D, et al., 1994. Are the apparent effects of cigarette smoking on lung and bladder cancers due to uncontrolled confounding by occupational exposures ? *Epidemiology*, 5: 57 - 65.
- Siemiatycki J, Krewski D, Franco E, et al. 1995. Associations between cigarette smoking and each of 21 types of cancer: a multi-site case-control study. *International Journal of Epidemiology*, 24; 3: 504 - 514.

- Smigel KA, 1988. Radon revisited; Smokers beware. *Journal of the National Cancer Institute*, 80; 16: 175 - 1276.
- Snihs JO, 1974. The approach to radon problems in non-radium uranium miners in Sweden. *Proceedings of the 3rd International Congress of the International Radiation Protection Association*. CONF-730997. Oak Ridge Tennessee: US Atomic Energy Commission.
- Spitzer WO, 1995. Editor's keynote address: The challenge of meta-analysis. *Journal of Clinical Epidemiology*, 48; 1: 1 - 4.
- Steck WJ, Field RW, 1998. Outdoor radon in central North America, *Proceedings of SPERA 1998: Environmental Radioactivity and its Application in Environmental Studies*, Christchurch, New Zealand, 1998.
- Steinhausler F, 1988. Epidemiological evidence of radon-induced health risks. IN: Nazaroff, WW and Nero, AV (eds.) *Radon and its Decay Products in Indoor Air*. John Wiley and Sons Inc., New York. Pages 311- 371.
- Stidley CA, Samet JM, 1993. A review of ecological studies of lung cancer and indoor radon. *Health Physics*, 65: 234- 251.
- Stuart LA, Parmar MK, 1993. Meta-analysis of the literature or of individual patient data: is there a difference ? *The Lancet*, 341: 418 - 422.
- Syme SL, Berkman LF, 1976. Social class, susceptibility and sickness. *American Journal of Epidemiology*, 104: 1 - 8.

- Taylor PR, Qiao Y, Schatzkin A, et al., 1989. The relation of arsenic exposure to lung cancer among tin miners in Yunnan province, China. *British Journal of Industrial Medicine*, 46: 881 - 886.
- Thacker SB, 1988. Meta-analysis: A quantitative approach to research integration. *Journal of the American Medical Association*, 259: 1685 - 1689.
- Thomas DC, 1983. Statistical methods for analyzing effects of temporal exposure on lung cancer risk. *Scandinavian Journal of Work and Environmental Health*, 9: 353 - 366.
- Thomas DC, McNeill KG, Dougherty C, 1985. Estimates of lifetime lung cancer risks resulting from Rn progeny exposures. *Health Physics*, 49: 825- 846.
- Thomas DC, 1987. Pitfalls in the analysis of exposure-time-response relationships. *Journal of Chronic Disease*, 40; suppl 2: 71s- 78s.
- Thomas DC, 1988. Models for exposure-time-response relationships with applications to cancer epidemiology. *Annual Review of Public Health*, 9: 451 - 482.
- Thomas J, Lamb D, Ashcroft T, et al., 1993. How reliable is the diagnosis of lung cancer using small biopsy specimens ? Report of a UKCCCR lung cancer working party. *Thorax*, 48: 1135- 1139.
- Thompson SG, Pocock SJ, 1991. Can meta-analyses be trusted ? *The Lancet*, 338: 1127 - 1130.
- Upfal M, Divine G, Siemiatycki J, 1995. Design issues in studies of radon and lung cancer: implications of the joint effect of smoking and radon. *Environmental Health*

Perspectives, 103; 1: 58- 63.

US Department of Energy (DOE) Office Energy Research, 1989. International Workshop on Residential Radon Epidemiology: Workshop Proceedings. Commission of European Communities, Radiation Protection Program. CONF-8907178.

US Department of Energy (DOE) Office Energy Research, Commission of European Communities Radiation Protection Programme, 1995. Proceedings of the Third International DOE/CEC Residential Radon Epidemiology Workshop February 1995, and Planning Meeting Combined Analysis North American Residential Radon Studies. Commission of European Communities, Radiation Protection Program. CONF-8907178.

US Department of Health and Human Services (DHHS), 1982. The Health Consequences of Smoking. Cancer: A report of the surgeon general. Rockville, Md.

US Department of Health & Human Services (DHHS), 1989. Reducing the Health Consequences of Smoking: 25 Years of Progress. A Report of the Surgeon General. US DHHS Publ. No. (CDC) 89-8411. Public Health Service, Washington DC.

US Environmental Protection Agency (EPA), Office of Air and Radiation, 1992. Respiratory Effects of Passive Smoking: Lung Cancer and Other Disorders. EPA/600/6-90/006F. Washington DC.

US Environmental Protection Agency (EPA) & US Dept of Health & Human Services (DHHS), 1986. A Citizen's Guide to Radon, What it is and What to do About it. Document OPA-86-001 (August), Washington DC.

- US Environmental Protection Agency (EPA) & US Dept of Health & Human Services (DHHS), 1994. A Citizen's Guide to Radon, Second Edition. The Guide to Protecting Yourself and Your Family From Radon. Indoor Air and Radiation (6604J) 402-K92-001 September 1994. Washington DC.
- Vandenbrouke JP, 1987. Should we abandon statistical modeling altogether ? American Journal of Epidemiology, 126; 1: 10 - 13.
- Venitt S, Biggs PJ, 1995. Radon, mycotoxins, p53 and uranium mining. The Lancet, 345: 59.
- Victor N, 1995. The challenge of meta-analysis': discussion. Indications and contra-indications for meta-analysis. Journal of Clinical Epidemiology, 48; 1.
- Wacholder S, McLaughlin JK, Silverman D, et al. 1992. Selection of controls in case-control studies I. Principles. American Journal of Epidemiology, 135; 9: 1019- 1028.
- Wallman KK, 1998. Data on race and ethnicity: revising the federal standard. The American Statistician, 52; 1: 31 - 33.
- Wang Y, Krewski D, Lubin JH, 1996. Meta-analysis of multiple cohorts of underground miners exposed to radon. Proceedings of the Statistics Canada Symposium 95. From Data to Information -Methods and System, Ottawa ON, Pages: 21 - 28.
- Weinberg CM, Moledor ES, Umbach DM, et al. 1996. Imputation for exposure histories with gaps under an excess relative risk model. Epidemiology , 7; 5: 490 - 497.
- White H, 1989. Estimation, Inference, and Specification Analysis. Cambridge University

Press, New York, NY.

- White SB, Bergsten JW, Alexander BV, et al., 1992. Indoor 222 Rn concentrations in a probability sample of 43,000 houses across 30 states. *Health Physics* 62: 41 - 50.
- Whitehead A, Whitehead J, 1991. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine*, 10: 1665 - 1677.
- Whittemore AS, McMillan A, 1983. Lung cancer mortality among US uranium miners: a reappraisal. *Journal of the National Cancer Institute*, 71: 489- 499.
- World Health Organization (WHO), Regional Office for Europe 1985. *Environmental Health Series Number 13: Indoor Air Quality: Radon and Formaldehyde*. Copenhagen.
- Wu-Williams AH, Samet JM, 1994. Lung cancer and cigarette smoking. IN: Samet JM, (ed.) *Epidemiology of Lung Cancer*. Marcel Dekker, New York NY. Pages: 71 - 108.
- Wynder EL, Graham EA, 1950. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. (Reprinted as "Landmark Research" in *JAMA* 1985; 253: 2986-94, and in *JNIH Res* 1992;4:63-72). *Journal of the American Medical Association*, 143: 329 - 336.
- Wynder EL, 1997. Tobacco as a cause of lung cancer: some reflections. *American Journal of Epidemiology*, 146; 9: 687 - 694.
- Xiang-Zhen X, Lubin JH, Li JY, et al., 1993. A cohort study in southern China of tin miners

- exposed to radon and radon decay products. *Health Physics*, 64: 120 - 131.
- Yao SX, Lubin JH, Qiao YL, et al., 1994. Exposure to radon progeny, tobacco use, and lung cancer in a case-control study in southern China. *Radiation Research*, 138: 326 - 336.
- Yeager WM, Lucas RM, Daum KA, et al., 1991. A performance evaluation study of three types of alpha track detector radon monitors. *Health Physics*, 60; 4: 507- 515.
- Yeh HC, Phalen RF, Raabe OG, 1976. Factors influencing the deposition of inhaled particles *Environmental Health Perspectives*, 15: 147 .
- Yusuf S, 1987. Obtaining medically meaningful answers from an overview of randomised clinical trials. *Statistics in Medicine*, 6: 281 - 286.
- Zhai P, Feng Y, 1993. Passive integrating radon level measurement of the building for nuclear and radiochemistry. *Atomic Energy Science and Technology*, 27; 1: 74 - 77.

Appendix A:

Detailed Tables Comparing Design Features of The Eight Published Case-Control Studies

Table A.1	Subject Selection
Table A.2	Proxy Data
Table A.3	Radon Exposure Information
Table A.4	Levels of Exposure
Table A.5	Prevention and Control of Confounding
Table A.6	Effect Modification
Table A.7	Smoking-Specific Analyses
Table A.8	Treatment of Age
Table A.9	Histological Analyses

Table A.1

Subject Selection

Location / Author	Subject Source/ Dates of Ascertainment		Matching	Method and Dates of Lung Cancer Case Confirmation	Attrition: Number and Proportion of Eligible and Interviewed Subjects Retained in the Analysis(♀ + ♂)	
	Cases	Controls			Cases	Controls
Shenyang China: Blot et al. 1990	Dx: 1985 -1987 Shenyang Cancer Registry	DNG Three-stage pop'n sampling from city administrative units	None	Review of pathology, cytology slides and xrays by an expert panel ; DNG: 'special rapid reporting system' between registry and expert panel	308 ♀ = 79% of those "eligible" (390 ?) but 397 = 95% of those interviewed (418?) *	356 ♀ = 91% of those eligible (391 ?) and 391= 99% of those interviewed (391); eligible = interviewed
New Jersey USA: Schoenberg et al. 1990	Dx: 1982 - 1984 1. Rapid reporting system with hospital pathology depts & 2. hospital pathology records, state cancer registry, & death certificate files	DNG 1. Live Cases: Driver licence (<65 yrs); medicare files (65 +) 2. Deceased cases: Death Certificates	Respondent Type (P); 1. Live (Direct): age & race (F) 2. Deceased (Proxy): age, race, closest date of death (P)	DNG Pathology reports were reviewed by Health Dept. physicians (no slide review)	433 ♀ = 44% of 994 interviewed, 31% of 1306 eligible ΔCUM: 346	402 ♀ = 40% of 995 interviewed, 27% of 1449 eligible ΔCUM: 314

Location / Author	Subject Source/ Dates of Ascertainment		Matching	Method and Dates of Lung Cancer Case Confirmation	Attrition: Number and Proportion of Eligible and Interviewed Subjects Retained in the Analysis(ϕ + Δ)	
	Cases	Controls			Cases	Controls
South Finland: Ruosteenoja 1991	Registry 1980 - 1982; Hospital 1983 - 1985	Population register 1980	Smoking and age ('balanced')	Hospital records	238 σ = 75% of the 318 eligible Δ + ϕ + smoking data: 222 Δ + smoking data : 164 (=69% of those selected)	495 σ = stratified sample of 1320 interviewed (88% of 1500 eligible) Δ + ϕ + smoking data: 379 Δ + smoking data : 334 (=68% of those selected)
Stockholm Sweden: Pershagen et al. 1992	1983 - 1985 ^b Surgical/ Pulmonary Wards at all city Hospitals	1. Population register DNG 2. Hospital 1983 - 1985	1. Pop'n Controls: date of birth (P); 2. Hospital Controls: none	Pathology reports	201 η = 96% of the 210 recruited EPNG	378 η = 95% of the 400 recruited EPNG
Winnipeg Canada: Létourneau et al. 1994	1983 - 1990 Manitoba Registry	DNG phone directory	Age (P) Gender (P)	Histologically Confirmed	738 488 σ 250 η = 53% of 1400 eligible	738 488 σ 250 η EPND
Missouri USA: Alavanja et al. 1994	1986 - 1991 State Registry	DNG Driver Licence (30 - 64 yrs) Medicare Files (65 - 84 yrs)	Age (F)	Histological confirmation of 76% of cases	538 η = 83% of 650 eligible	1183 η = 78% of the 1587 eligible

Location / Author	Subject Source/ Dates of Ascertainment		Matching	Method and Dates of Lung Cancer Case Confirmation	Attrition: Number and Proportion of Eligible and Interviewed Subjects Retained in the Analysis(σ + Δ)	
	Cases	Controls			Cases	Controls
Sweden: Pershagen et al. 1994	1980 - 1984 Registry	DNG 1.Live Cases: Population Register; 2. Deceased Cases: Death Register	<u>Control Gp 1:</u> Age (F); Cal. yr. of residence (F) <u>Control Gp 2:</u> Age (F); Cal. yr. of residence (F); Respondent Type (F)	84.2 % of cases histologically; 14.6 % cytologically; 1.1 % on 'other criteria'	1281 σ ♀ : Δ but only 1118 = 82.2% of 1360 eligible completed a questionnaire	2576 σ ♀ : Δ = 90% of 2847 eligible; only 2327 completed a questionnaire ^c
Finland: Auvinen et al. 1996	Dx: 1986 - 1992	Population Register DNG (Alive when cased diagnosed)	Age; Gender (P)	Histologic verification in 73% of cases, Cytologic in an additional 19%	1973 σ ♀ Identified: Δ + σ : 1055 = 53% of those identified ^d	2885 σ ♀ Identified: Δ + σ : 1544 = 54% of those identified

Table A.1 Notes:

DNG: Dates not given.

Dx: Diagnosis.

EPNG: Eligible population not given.

EPND: Eligible population not determined.

σ Male-restricted study.

♀ Female-restricted study.

σ ♀ Males and females included.

F: Frequency matching.

ID: Identification.

P: Pair-wise matching.

σ Analysis includes some interpolation of missing data on these subjects.

Δ Indicates that home radon measures were available for at least one home of each subject.

^a By deduction, more cases were interviewed than "eligible", and the number of "eligible" controls equaled the number of interviewed controls

^b Year of ascertainment (end of observation) for each matched pair was the date of interview of the case

^c Fewer cases and controls than included in the analysis actually completed a questionnaire.

^d Numbers of cases and controls shown were those included in the non-gender-matched analysis; 517 matched pairs (29% of those identified) were included in a matched analysis.

Table A.2

Proxy Data

Location / Author	Subjects Included		Matching	Adjustment	Proxy-Specific Analysis ^a	Comments
	Cases	Controls				
Shenyang China: Blot et al. 1990	Direct	Direct	-	-	-	If survival "selective", bias minimized through the use of special rapid reporting system
New Jersey USA: Schoenberg et al. 1990	Direct 246 57% Proxy 187 43%	Direct 212 53% Proxy 190 47%	✓ pairwise	✓	×	Both the main effect of respondent type and its interaction with smoking status were retained in the final model
South Finland: Ruosteenoja 1991	Direct 29 12% Proxy 209 88%	Direct 342 79% ^b Proxy 92 21%	×	×	×	Non-response greatest among proxy controls and least amongst live cases ^b
Stockholm Sweden: Pershagen et al. 1992	Direct	Direct	-	-	-	If survival "selective", bias minimized through the use of incident cases
Winnipeg Canada: Lévesque et al. 1994	Direct 481 65% Proxy 257 35%	Direct 660 89% Proxy 78 11%	×	×	✓	Analysis restricted to the 471 case-control pairs who were directly interviewed "failed to provide evidence of increased risk" with radon
Missouri USA: Alavanja et al. 1994	Direct 197 37% Proxy 341 63%	Direct 1183 100%	-	-	✓	A marginally significant (p=0.06) dose-response trend was limited to direct interview subjects
Sweden: Pershagen et al. 1994	Direct 8% Proxy 92%	Gp1 Direct 82 % Proxy 18% Gp2 Direct 8 % Proxy 92 %	✓ Control Gp2 only	×	×	Failure to adjust for the proxy matching variable may have introduced confounding into the data
Finland: Auvinen et al. 1996	Pre-Interview: Direct 15% Proxy 85%	Pre-Interview: Direct 90% Proxy 10%	×	×	×	

Table A.2 Notes:

Direct: subject alive and well enough and to be interviewed.

Proxy: next of kin provided information on housing history and covariates such as smoking.

Gp Group.

a Investigation of effect modification (EM) through either stratified analysis or testing for the importance of an interaction term in the model.

b Among 1450 controls originally selected, 74% (1066) were direct, and 26% (384) proxy. Among cases: of those originally selected 10% (29) were direct, and 90% (262) proxy: Percent non-response therefore was 0% for live cases, 20% for dead cases, 68% for live controls and 76% for deceased controls.

Table A.3

Radon Exposure Information

Location / Author	Duration and Method	Residence Inclusion Criteria	Location of Dosimeter Placement	Exposure Time Window	Exposure Time Window (ETW) Coverage	Analysis	Method of Interpolating Missing Data
Shenyang China: Blot et al. 1990	1 year α Track T	Last residence of ≥ 5 years in Shenyang; 1 home monitored	Bedroom and living Room; maximum reported	None specified	Only 1 residence monitored; Median residence time 24 years	CONC	None
New Jersey USA: Schoenberg et al. 1990	1 year α Track T	Last NJ residence of ≥ 10 years during the period 10 - 30 years prior to dx or selection; 1 home monitored	Living Area (76%); basement (5%); 4 day charcoal canister (8%)	5 - 30 years prior to dx or selection	Only 1 residence monitored; Median ETW residence time in years: 20 (cases) & 21 years (controls); 82% cases & 79% controls resident >15 years	CONC CUM ϕ	Median value of controls assigned for periods not residing in index home; apartments assigned 0.4 pCi/liter
South Finland: Ruosteenoja 1991	2 month α Track G	All residences of ≥ 1 year between 1950 and 1975	Bedroom or living Room	5 - 35 years but monitored 1950 - 1975	50% of eligible homes monitored: ≥ 1 home for 72% of cases, 79% of controls; Mean ETW coverage: Δ : 14 (cases) & 16 (controls) (approx. 50% person-time covered); $\phi + \Delta$: 20 years (both groups)	TWAC ϕ	HouseType/location or if no such information available, Finnish average of 100 Bq/m ³ ; Dwellings above ground assigned 50 Bq/m ³
Stockholm Sweden: Pershagen et al. 1992	1 year α Track T (87%)*	All Stockholm county residences of ≥ 2 years during index period	Bedroom & living room (averaged)	1945 to 5 years prior to interview	74% (1573) of eligible (2118), 45% of lifetime residences monitored; Mean of 26 years of ETW covered	TWAC CUM ϕ	House Type/Age or if no such information, previous home concentration
Winnipeg Canada: Lévesque et al. 1994	1 year α Track G	All Winnipeg residences of ≥ 1 year during index period	Bedroom and basement (reported separately)	5 - 30 and 5 - 15 years prior to interview	33% of eligible residences monitored; Mean years covered: 17 in 5 - 30 years ETW (68% of person-time); 8 in the 5 - 15 year ETW (80% of person-time)	CUM ϕ	Calibration to bedroom or basement monitored; if no measurement, average study value for all subjects
Missouri USA: Alavanja et al. 1994	1 year α Track T	All in-state index period residences	Bedroom and kitchen (differences slight, but not stated which used)	5 - 30 years prior to interview	Average coverage of 20 years; ETW coverage: living cases: 78.5%; deceased cases: 76%; controls: 78.8%	TWAC CUM ϕ	Status-specific mean (cases and controls assigned the respective group mean)

Location / Author	Duration and Method	Residence Inclusion Criteria	Location of Dosimeter Placement	Exposure Time Window	Exposure Time Window (ETW) Coverage	Analysis	Method of Interpolating Missing Data
Sweden: Pershagen et al. 1994	3 month α Track G	Must have lived in Sweden at some time between 1980 and 1984, and on Jan 1, 1947; each residence of ≥ 2 yrs monitored	Bedroom and living room (not stated which used)	1947 to 3 years prior to dx (cases) or selection (controls)	Mean years covered: 23.5 for cases; 23 for controls; ETW coverage: cases 72.4%; controls 71.1%	TWAC CUM ϕ	Residence characteristics or median value for all subjects
Finland: Auvinen et al. 1996	1 year α Track G	Cohort inclusion criteria was residence in the same single family house from Jan 1 1967 or earlier to end of 1985 (19 years); 1 home monitored	Bedroom or living room.	Non-specified, but inclusion criteria guaranteed minimum 19 years coverage.	Median number of years at index address: cases 38, controls 36; For all subjects, average 54% of life years, and 73% of adult years were covered by measurement.	CONC	None

Table A.3 Notes:

- * The remaining 13% of monitored homes had 2 week thermoluminescence measures, from which regression was used to estimate annual concentration.
- τ Terradex Corporation manufactured and read dosimeters.
- CONC only radon concentration in the one monitored home considered.
- CUM Exposures were cumulated over the ETW.
- dx Diagnosis.
- ETW Exposure time window.
- G Government office responsible for dosimeter provision.
- PT Person-time.
- TWAC Analysis was by time weighted (by residence time) averaging of measured concentrations.
- ϕ Results were analysed with imputation of missing data as described.
- Δ Monitored homes.

Table A.4
Mean Radon Exposure

Location / Author	Radon Concentration (Time Weighted Average) Bq/m ³				Cumulative Radon Exposure	
	Median	Geometric Mean	Arithmetic Mean	Extreme Values ^a	Reported	Central Estimate
Shenyang China: Blot et al. 1990	85	89	–	20 % > 148	×	–
New Jersey USA: Schoenberg et al. 1990	22	–	–	1 % > 148 (1.4% cases; 0.5% controls)	✓	Not Specified
South Finland: Ruostenoja 1991	–	–	220	40 % > 174	×	–
Stockholm Sweden: Pershagen et al. 1992	–	96	127	28 % > 148 (29% cases; 25% controls)	✓	Not Specified
Winnipeg Canada: Lévesque et al. 1994	–	–	120 ^b	24 % > 144	✓	Mean 32,250 Bq/m ³ -years for 5 - 30 year time window (b)
Missouri USA: Alavanja et al. 1994	–	44	67	7 % > 148 (6.5% cases; 6.8% controls) 20 % > 91 (23% cases; 18% controls)	(c)	–
Sweden: Pershagen et al. 1994	–	60	107	25 % > 117	×	–
Finland: Auvinen et al. 1996	67	–	–	–	×	–

Table A.4 Notes:

- a Revised based on Lubin et al. 1995b.
- b Bedroom measures given, to be consistent with living area measurements in other studies; average basement: concentration 200 Bq/m³; cumulative estimate for 5-15 year time window 1440 Bq/m³-years.
- c Unit of cumulative exposure defined as a combination of the average period for which (time-weighted average) exposures were estimated (32.5 years) and a radon level of 2.7 pCi per liter, with missing measurements replaced by median levels, and after exclusion of subjects who slept with an open window.

Table A.5

Prevention and Control of Confounding

Location / Author	Restriction ^a	Potential Confounders Investigated	Matching with Adjustment [✓]	Variables Adjusted for in Results Presented ^b
Shenyang China: Blot et al. 1990	♀ Age: 30 - 69 years	Age; smoking status; spouse smoking; air pollution; education.	No Matching	Age; smoking status (2); air pollution (2); education.
New Jersey USA: Schoenberg et al. 1990	♀	Age; active smoking (several variables); 'race'; education; county of residence; occupation; vegetable consumption; respondent type.	Respondent type [✓] Age [✓] Race [✗] Closest date of death (Proxy) [✗]	Age; lifetime nonsmoking (2), years since quit smoking; cigarettes per day (3); occupation; respondent type; respondent type X smoking.
South Finland: Ruosteenoja 1991	♂	Age; education; active smoking; farming; air pollution; asbestos.	Smoking ('balancing') [✓]	Smoking (pack-years)
Stockholm Sweden: Pershagen et al. 1992	♀	Age; active smoking; municipality of residence.	Date of birth (age, for population controls) [✓]	Age; smoking (4); municipality of residence.
Winnipeg Canada: Lévesque et al. 1994	Age: 35 - 80 years (cases)	Age; gender; smoking; education; country of birth (NEF); occupation (NEF).	Age [✓] Gender [✗]	Age; smoking status (2); education
Missouri USA: Alavanja et al. 1994	♀ Non-smokers White 'race'	Age; marital status; previous smoking; passive smoking, pre-existing nonmalignant lung disease; education; saturated fat; respondent type.	Age [✓]	Age (6) + one (or 2 for smoking) of several other variables
Sweden: Pershagen et al. 1994	Age: 35 - 74 years	Age; active smoking status; gender; occupation; urbanization.	Age [✓] Cal. year of residence [✗] Respondent type (gp 2) [✗]	Age (4); smoking status (4); gender (2); occupation (3); urbanization(2)
Finland: Auvinen et al. 1996	Residency criterion only	Active smoking (several variables); environmental tobacco smoke; occupational hazard exposure.	Age [✗] Gender [✗]	Smoking status; smoking intensity; smoking duration; age at start of smoking.

Table A.5 Notes:

^a Aside from residency restrictions^b Adjustment refers to inclusion in a multiple regression model; parenthesis contains number of categories in adjustment variable or C for continuous

NEF; No effect on model fit

(#) Bracketed number indicates number of levels of a categorical variable(when given)
c A matched analysis was conducted but was limited to 27% of originally recruited case/control pairs

Table A.6

Effect Modification

Location / Author	Model form	Variables Reportedly Investigated ^a	Stratum-Specific Results Provided For			
			Smoking	Age	SWOW ^b	Histological Type
Shenyang China: Blot et al. 1990	Logistic	(I & S) Smoking	√ 3 levels	×	×	√
New Jersey USA: Schoenberg et al. 1990	Logistic	(I) Smoking X respondent type	√ 4 levels	×	×	√
South Finland: Ruosteenoja 1991	Logistic	(S) Smoking (S) Asbestos (S) Farming (S) Birth cohort (S) Education (S) Pollution	√ 2 levels	×	×	√
Stockholm Sweden: Pershagen et al. 1992	Logistic ^d	(S) Smoking (S) Age	√ 4 levels	√ 3 levels	×	√
Winnipeg Canada: Lévesque et al. 1994	Conditional logistic	(S) Respondent type	×	×	×	√
Missouri USA: Alavanja et al. 1994	Logistic	(S) Smoking (S) Age	√ 2 levels	√ 6 levels	×	√
Sweden: Pershagen et al. 1994	TWAC: Conditional logistic; CUM: ERR	(I & S) Smoking (S) SWOW	√ TWAC only	×	√	√
Finland: Auvinen et al. 1996	Conditional and Un-Conditional Logistic Regression	(I & S) Age (I & S) Gender (I & S) Smoking status	√ status and 3 dose levels	√ 4 levels	×	√

a Estimation of an interaction term (I) or stratum-specific analyses (S).

b SWOW = sleeping with an open window.

c Because the effect of smoking seemed to account for most of the effect of age.

d Logarithmic transformation of the radon exposure data improved estimation.

Table A.7

Smoking-Specific Analyses

Location / Author	Proportion of Smokers	Treatment of Passive smoking	Model Adjustment for Confounding	Stratum-specific Results for Active Smoking			Overall Results
				Current	Ex-smokers	Never-smokers	
Shenyang China: Blot et al. 1990	Not Given	Spousal smoking Dichotomous but did not appear to adjust for	2 Levels of "smoking status": smoked ≥ 6 months or < 6 months	✓ CONC 2 levels (light and heavy)	✗ current and ex smokers not differentiated	+/- CONC R nonsmokers used cigs up to 6 months	P = 0.15 for interaction between smoking status (dichotomous) and radon; RR's higher for heavy smokers vs light or nonsmokers
New Jersey USA: Schoenberg et al. 1990	Ever- smokers: cases: 86% controls: 47%	Taking spousal smoking into account did not change nonsmokers pattern of risk	Lifetime nonsmokers vs. ever-smokers (3 levels) adjusted for years since quit smoking	✓ TWA R Strongest ER trend (p=0.01) occurred among light (<15cigs/day) smokers	✗ R consideration implicit by lifetime average cigs/day and years since quit but no results provided	✓ TWA R OR's 0.9, 1.2 NS ER trend (P=0.39 one-sided)	For continuous, cumulative radon exposure, OR greater for smokers than nonsmokers. Among smokers of less than 15 cigs/day, Sig. ER trend (p=0.01) Trend NS for 15-24 and negative for 25+.
South Finland: Ruosteenoja 1991 (a)	Ever-smokers: cases: 98% controls: 90%	✗	Pack-years of smoking	✓ TWA Only heavy smokers in 174-274 Bq/m ³ Rn exposure class had significant OR	✗	✗ R Non & light smokers mixed(< 20 pack-years)	In non&light smokers there was no systematic variation of Rn conc with risk, but for heavy smokers, OR's increased incrementally for 2nd, 3rd, and 4th quintiles (NS)
Stockholm Sweden: Pershagen et al. 1992	Ever-smokers: cases: 81% controls: 49%	✗	4 levels: never-smokers; ex-smokers; 1 - 9 cigs/day; ≥ 10 cigs/day	✓ CUM NS ER trend among current smokers (p = 0.24 for light, 0.29 for heavy smokers)	✓ CUM OR's consistently significant but ER trend NS (P=0.91)	✓ CUM R Both OR's for upper tertiles (2.6, 3.2) were significant ER trend (P = 0.04)	"consistent with additive interaction" but little power to distinguish between the 2 models; Highest OR among the heaviest smokers

Location / Author	Proportion of Smokers	Treatment of Passive smoking	Model Adjustment for Confounding	Stratum-specific Results for Active Smoking			Overall Results
				Current	Ex-smokers	Never-smokers	
Winnipeg Canada: Létourneau et al. 1994	Ever-smokers: cases: 97% controls: 70%	×	Ever/Never Smoking Status	×	×	×	- Not tested
Missouri USA: Alavanja et al. 1994	All currently non-smokers	✓ ^a OR 1.24 for highest radon exposure conc. Adjusted for age	Previous smoking (pack-years, previous smoker, years since cessation)	-	✓ TWA ^a OR 1.32 for highest radon exposure conc. No ER trend within group (b)	✓ TWA ^a OR 1.20 for highest radon exposure conc. No ER trend	Could not test as only current nonsmokers
Sweden: Pershagen et al. 1994	Ever-smokers: cases: 68% controls: 46% Unknown: cases: 18% controls: 18%	×	Four levels: never, ex-smoker, and current smoker of 1 - 9 or 10 or more cigarettes/day	✓ TWA All 5 radon exposure levels demonstrated significant OR's Largest (NS) trend was for current smokers	✓ All 5 radon exposure levels demonstrated significant OR's NS ER trend	✓ ^a OR's of 1.5 & 1.2 NS elevated in two highest radon conc. NS ER trend	When radon and smoking analysed as continuous variables, interaction exceeded an additive effect (p=0.02)
Finland: Auvinen et al. 1996	Ever-smokers: cases: 91% controls: 56%	Intensity and duration (not included in final model)	Smoking status, intensity, duration, and age at start of smoking	✓ CONC	✓ CONC	✓ CONC	All OR's centered around null value

Table A.7 Notes:

a Restriction of sample by "attempt to balance" cases and controls on smoking status see page 87.

b No confidence limits given.

cigs/day cigarettes/day.

conc Radon concentration category.

CONC Radon concentration in the one measured home.

a Restriction of sample by "attempt to balance" cases and controls on smoking status see page 87.
 b No confidence limits given.
 cigs/day cigarettes/day.
 conc Radon concentration category.
 CONC Radon concentration in the one measured home.
 CUM Cumulative radon exposure analysis augmented a TWA analysis.
 NS Not statistically significant assuming an alpha level of 0.5.
 TWA Time weighted average concentration analysis only.
 X Data not provided.
 ✓ Data explicitly provided.
 LR Logistic regression.
 ER Exposure-response.
 OR Odds ratio.
 R This strata contains its own reference group, that of lowest radon level, which is the basis for its odds ratios.

Table A.8

Treatment of Attained Age

Location / Author	Design		Matching	Analysis		Comments
	Restriction			Adjustment for Confounding	Stratum Specific Results (EM)	
	Case Inclusion Criteria	Proportion in highest age category				
Shenyang China: Blot et al. 1990	Cases 30 - 69 years	Not Given	× Controls of a "similar age"	√	×	Type of variable/ categories for adjustment not √
New Jersey USA: Schoenberg et al. 1990	None stated	25% cases 25% controls > 72 years	In 5 year age groups	√	×	Type of variable/ categories for adjustment not specified
South Finland: Ruosteenoja 1991	None	Not Given	Informal "balancing" of controls to case age (pg 15)	×	×	Cases on average 1.8 years older than controls;
Stockholm Sweden: Pershagen et al. 1992	None stated	43% cases 33% controls ≥65 years	Population (not hospital) controls matched on DOB	√ ≤54; 55-64; ≥65	√	Significant trend within the lowest age group ≤54 years (p=0.01) and all ages (p=0.02)
Winnipeg Canada: Létourneau et al. 1994	Cases 35 - 80 years	1.6% cases 4.7% controls > 75 years	Controls +/- 5 years from case age	√ CLR	×	
Missouri USA: Alavanja et al. 1994	Controls 30 - 84 years	46% cases 24% controls ≥80 years	In 5 year age strata	√ in 6 categories	√ in 3 age strata	Strongest trend (n.s.) with increasing radon in <65 years (youngest age group)
Sweden: Pershagen et al. 1994	All subjects 35 - 74 years	44 - 52% 65-74 variable among gender and outcome groups	In 5 year intervals	√ in 5 year intervals	×	Among cases, men older than women
Finland: Auvinen et al. 1996	None	20% cases 20% controls ≥75 years	Year of birth	×	√ in 4 year age strata	Trend P across age strata 0.62

Table A.8 Notes:

a Schoenberg et al. 1989
DOB Date of birth

Table A.9

Histological Analyses

Location / Author	Overall Results	Small Cell	Large Cell	Squamous Cell	Adenocarcinoma	Other/Unknown Type
Shenyang China: Blot et al. 1990	All OR estimates <1 (compared to reference category)	All OR estimates > 1 (CI's not given); NS trend	—	—	All OR estimates <1	All OR estimates ≤ 1
New Jersey USA: Schoenberg et al. 1990	All showed an ER trend except squamous	most significant CUM result (p=0.079)	most significant CONC result (p=0.027)	No ER trend	NS ER trend	—
South Finland: Ruosteenoja 1991	Increasing OR's except for highest exposure category	All OR's negative	—	Similar pattern to overall	—	—
Stockholm Sweden: Pershagen et al. 1992	Significant ER trend (p=0.05)	Significant ER trend (p = 0.05)	—	NS ER trend (p = 0.09)	No ER trend; highest exp OR 1.1	NS trend (p = 0.27)
Winnipeg Canada: Létourneau et al. 1994*	All point estimates > 1 but none significant	OR = 1.35 (0.44 - 4.11)	—	OR = 1.19 (0.9 - 1.56)	OR = 1.33 (0.83 - 2.13)	OR = 1.35 (0.16 - 1.53)
Missouri USA: Alavanja et al. 1994	Significant trend ^b (p<0.05)	—	—	—	Significant ER trend ^c 49% of tumours	other =28% of all tumours unknown = the 24% not typed
Sweden: Pershagen et al. 1994	Significant trends for both TWA and CUM analysis	4th Rn quartile significant (OR pt est: 2.8) NS trend (p = 0.15)	—	3rd & 4th Rn quartiles sig (pt est: 1.5; 1.7) NS trend (p = 0.09)	3rd & 4th Rn quartiles sig (pt est: 1.4; 2.3) No trend	NS trend (p=0.27)
Finland: Auvinen et al. 1996	No significant OR's in any histologic group	OR = 1.02 {0.81 - 1.27}	—	OR = 1.0 {0.90 - 1.12}	OR = 1.01 {0.87 - 1.18}	'Other' OR = 1.17 {0.98 - 1.39}; 'Unknown' OR = 0.99 {0.86 - 1.14}

Table A.9 Notes:

- a Analysis of cumulative radon (bedroom measures) analysed as a continuous variable (OR per 3750 Bq/m³-years) , and restricted to subjects with minimum 75% coverage of the 5 - 30 year ETW.
- b Significant trend limited to the log-transformed cumulative radon exposure data, analysed as a continuous variable; P value ranged from 0.01 to 0.04 depending on the model covariates.
- c Significant trend was observed for adenocarcinoma when continuous results were adjusted for saturated fat and age (p=0.04), and when categorical results adjusted for age and one of several other variables: previous smoking(p=0.04); smoking pack-years + years since cessation (p = 0.01); previous lung disease (p=0.05); education (p=0.4); saturated fat intake (p = 0.01).
- CI Confidence interval.
- CUM Analysis of cumulative radon exposure.
- ER Exposure-response.
- NS Not statistically significant.
- OR Odds Ratio.
- pt est Point estimate.
- TWA Analysis of time-weighted (by residence duration) in monitored homes only.
- { - } 95% confidence interval.
- * Every type not classified in a particular study (-) was included in "other/unknown".

Appendix B:

Common Data Format

The data received from each site was organized into two separate records: 1) covariates and general information, and, 2) year by year variables, including those related to radon exposure. Tables B.1 and B.2 outline each record has been coded and formatted. For all variables, missing values have been set to '.' Note however, that missing-ness may represent one of several distinct conditions specific to the respective variable definition.

Record 1: Covariates

Table B.1
Data Format For Record 1: Covariates

SAS VARIABLE NAME	DESCRIPTION	COLUMN START IN ASCII FILE	COLUMN WIDTH IN ASCII FILE ¹	CODED (SAS FORMATTED) VALUES
ID	Study-defined unique subject identifier	1	8.	Integer
SET_NO	Pair-Matched set number or Frequency-Matched stratum number	10	4.	NA (sequential)
STATUS	Subject status (type)	15	1.	Case 1 Control 0
HISTO	Histological type	17	1.	Control 0 Squamous 1 Small/oat cell 2 Other 3 Adenocarcinoma 4 Uncertain 5
AGE_ASC	Age at diagnosis for cases; Age at interview for controls	19	2.	Integer value between 30 and 84

SAS VARIABLE NAME	DESCRIPTION	COLUMN START IN ASCII FILE	COLUMN WIDTH IN ASCII FILE ¹	CODED (SAS FORMATTED) VALUES	
YR_ASC	Last two digits of calendar year of ascertainment defined as year of diagnosis for cases and interview for controls	22	2.	Numeric 19'__' where '__' values can range from 82 to 91 for cases	
INTERV	Interview type (source of information) self (subject) or proxy (spouse or other surrogate) respondent	25	1.	Subject	1
				Proxy	2
GENDER	Subject gender	27	1.	Females	1
				Males	2
EDU_CAT	Highest level of education undertaken (but not necessarily completed) by subject.	29	1.	Some Elementary school (Grades 0-7)	1
				Some Secondary school (Grades 8-13)	2
				Some Post-Secondary Education	3
				Unknown	.
INCOME	Average annual family income in Canadian or US dollars (respectively). (Note that income was not ascertained in the New Jersey study).	31	1.	<\$10,000	1
				\$10,000-\$20,000	2
				\$20,000-\$30,000	3
				\$30,000-\$50,000	4
				>=\$50,000	5
				Unknown	.
SM_TYPE	Smoking category of subject. 'Never'-smokers include those who smoked < 1yr or < 100 cigs. (original category 5 for Missouri); mixed refers to those smoking both cigarettes and pipes or cigars. Those subjects with missing values are eliminated from the data set.	33	1.	Never-smoked	1
				Cigarettes only	2
				Pipe/cigar only	3
				Mixed	4
SM_RATE	Smoking rate as average number of cigarettes smoked/day during the active smoking period. Defined as 0 for non or pipe/cigar smokers.	35	5.2	Numeric with 2 digits before the 2 decimal places	
SM_START	Age started smoking. Defined as missing (.) for non or pipe/cigar smokers.	41	2.	Positive integer value	
SM_STOP	Age stopped smoking. Defined as missing (.) for non or pipe/cigar smokers.	44	2.	Positive integer value	

SAS VARIABLE NAME	DESCRIPTION	COLUMN START IN ASCII FILE	COLUMN WIDTH IN ASCII FILE ¹	CODED (SAS FORMATTED) VALUES	
SM_DUR	Net years of cigarette smoking adjusted for periods of cessation intermediate between start and stop age. Defined as 0 for non or pipe/cigar smokers.	47	4.1	Numeric with 2 digits before the single decimal place	
SM_LAG	Number of years prior to ascertainment since the subject last smoked cigarettes. Defined as missing (.) for non or pipe/cigar smokers.	52	2.0	Positive integer value	
SMKHM	Lived with cigarette smoker at any time prior to ascertainment	55	1.0	Yes No Missing	1 2 .
ETHNICITY	Race (New Jersey and Missouri) or country of origin (Winnipeg). Missouri study was restricted to subjects of 'white' race (all coded as 1).	57	1.0	<u>New Jersey:</u> White/Hispanic Black/Native/Asian or Pacific islander/ Other <u>Winnipeg</u> Canadian birth Non-Canadian birth	 1 2 1 2

Record 2: Year By Year Variables

This record was output in two different formats:

- 1) The *person-years format* for ease of reporting, with 50 records (lines of observations) per person, one for each calendar year prior to ascertainment, indexed by an additional variable TIME, and
- 2) The *analysis format*, as a flat file with 1 observation per person and blocks of 50 time-indexed arrays for the variables (CONC, ES_TYP, CONCB, ES_TYB, OTH_SM, O_CIGS, S_CIGS) specific to each of the 50 years prior to ascertainment.

It is the former reporting format which is delineated in Table B.2 below.

Table B.1
Data Format For Record 2: Year By Year Variables

VARIABLE NAME	DESCRIPTION	START COLUMN	COLUMN WIDTH	CODED (SAS FORMATTED) VALUES
ID	Study-defined subject identification number	1	8.0	Positive integer value
STATUS	Subject status	10	1.0	Case 1 Control 0
TIME	Index variable for the following variables which take on unique values for each of 50 years prior to ascertainment	12	2.0	Sequential Positive Integer from 1 - 50
HOME	Sequential number starting at 1 for the most recent, increasing as time prior to ascertainment increases. ¹	15	2.0	Positive integer value
CONC	Average annual radon concentration in living area in Bequerels/cubic meter	18	5.1	Numeric value with 3 digits prior to a single decimal place
ES_TYP	Estimation type for living area radon concentration during indexed year.	24	1.0	Regression from basement 1 OCMI ² 2 1 Year Alpha Track Measure 3 Canister Measure 4 Glass Measure 5 Not an Applicable Time Period for Subject 6 Missing (No estimation) '.'

¹ In Missouri and New Jersey data sets, home identification numbers were available only for homes with complete measurement. In these data sets, therefore, a missing ('.') value can denote either that the relevant home was not identified, or measurement was incomplete (refusal, loss of dosimeter, less than a year of measurement). The Winnipeg file contains home numbers every home identified in the housing history whether measured or not. In Winnipeg, missing values indicate lack of identification. The availability of a complete house history with identification numbers is important, as errors in estimated duration of residence in a home will be correlated for adjacent homes. Such information will be considered in an uncertainty analysis during the final (phase II) analysis.

² OCMI refers to the Observed Control Mean Imputation method (Weinberg et al. 1996)

VARIABLE NAME	DESCRIPTION	START COLUMN	COLUMN WIDTH	CODED (SAS FORMATTED) VALUES
CONCB	Average annual radon concentration in basement in Becquerels/cubic meter (Only routinely available in Winnipeg data set) during indexed year.	26	5.1	Numeric value with 3 digits prior to a single decimal place
ES_TYB	Estimation type for basement radon concentration during indexed year.	32	1.0	Regression from basement Imputation by OCMI 1 year alpha track meas. Canister Glass Measure Not an Applicable Time Period for Subject Missing (No estimation)
OCCUP	Estimated fraction of the average 24 hour period that the subject occupied the home (direct information available only in Missouri data set) during indexed year.	34	4.2	Numeric value (a proportion between 0.0-1.0)
OTH_SM	Number of smokers (other than subject) in home during indexed year.	39	2.0	Positive integer value
O_CIGS	Number of cigarettes smoked/day by smokers other than subject in indexed year. (Note that Winnipeg lacks direct data on this variable).	42	3.0	Positive integer value
S_CIGS	Number of cigarettes smoked/day by the subject in indexed year.	46	3.0	Positive integer value

Appendix C:

Models Considered During Development of a Common Logistic Regression Model

The two following tables provide details pertaining to covariate specification in the more than 70 distinct models considered during development of a common logistic regression model. As detailed in section 3.6.4.1.1, the models considered were of four basic types with respect to the treatment of active smoking:

- A Combined Intensity and Duration In Pack-years
- B $\text{Log}(\text{intensity}^1)$ and $\text{Log}(\text{duration}^*)$
- C Intensity and Duration
- D $\text{Log}(\text{intensity})$ and Age

Models of each type were fit to each data set (site) separately. Each “site-specific” model was coded with the respective letter (A through D) as a prefix, and a sequential number. The D models followed A through C in time. D models analysed using the pooled data from all three sites together, (with common coefficients for covariates or in some cases, site-specific covariates), were designated by the prefix E and a sequential number for the “combined model”.

Table C.1 provides the coding key for models of type A, B & C with respect to smoking all of which incorporated duration of smoking either as a separate term, or as an element in the

¹ The term for smoking intensity was denoted in the following tables as “dose”.

pack-years term which was formed as the product of duration and average intensity (cigarettes smoked per day).

Table C.2 specifies models of type D: those including age at ascertainment and not necessarily duration of smoking. Note that models D30 and D31 were in fact B type models since they included the duration term (for the purpose of a final re-test of this term); to avoid confusion in the results archives, these analyses were not re-labelled.

Table C.1

Models of Type A, B & C

Model Code	Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent ¹	Active Smoking*							Radon	Calendar Year of Ascertainment
								Pack-years	Log dose	Log duration	Dose	Duration	Recency	Start Age		
ABC	D	2-1		4-1**	D	D _{EN}	D (+/-)	5-1	C	C	C	C	D	C		
AB	D	2-1		4-1**	D	D _{EN}	D (+/-)	5-1	C	C			D	C		
BC	D	2-1		4-1**	D	D _{EN}	D (+/-)				C	C	D	C		
AC	D	2-1		4-1**	D	D _{EN}	D (+/-)	5-1			C	C	D	C		
A1	D	2-1		4-1**	D	D _{EN}	D (+/-)	5-1					D			
A2	D	2-1		4-1**	D	D _{EN}	D (+/-)	C								
A3	D	2-1		4-1**	D	D _{EN}	D (+/-)	5-1								
A4	D	2-1		4-1**	D	D _{EN}	D (+/-)	5-1						C		
A5	D	2-1		4-1**	D		R(2)	5-1								
A6	R	2-1		4-1**	R		D (+/-)	5-1					R			
A7		2-1		4-1**	D		D (+/-)						4-1			
A8		2-1		4-1**	D		D (+/-)	5-1					R			
B1	D	2-1			D	D _{EN}	D (+/-)		C	C			D			
B2	D	2-1			D	D _{EN}	D (+/-)		C	C						
B3 ^a	D	2-1			D	D _{EN}	D (+/-)		C	C			D			
B4 ^a	D	2-1			D	D _{EN}	D (+/-)		C	C						

Model Code	Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent ¹	Active Smoking*							Radon	Calendar Year of Ascertainment
								Pack-years	Log dose	Log duration	Dose	Duration	Recency	Start Age		
B5	D	2-1			D	D _{BN}	D (+/-)		C	C						C
B6	D	2-1	O (5)		D	D _{BN}	D (+/-)		C	C						
B7	D	2-1	O (5)		D	D _{BN}	D (+/-)		C	C						C
B8	D	2-1	O (5)		D	C	D (+/-)		C	C						C
B9		2-1	2-1		D	4-1	D (+/-)		3-1	3-1						C
B10		O (3)	O (5)		D	D _c	D (+/-)		C	C						C
B11		O (3)	O (5)		D	D _{BN}	D (+/-)		C	C						C
B12		2-1	2-1		D	D _c	D (+/-)		C	C						C
B13		2-1	O (5)		D	D _c	D (+/-)		C	C						C
B14		2-1	O (5)		D	D _{BN}	D (+/-)		C	C					3-C	
B15		2-1	O (5)		D	D _c	D (+/-)		C	C						
B16	D	O (3)	O (5)		D	D _{BN}	D (+/-)		C	C					3-C	C
B17		O (3)	O (5)		D	D _c	D (+/-)		C	C					1-C (5-15)	
B18		O (3)	O (5)		D	D _c	D (+/-)		C	C					1-C (5-15)	
B19	D	2-1	O (5)		D	D _c	D (+/-)		C	C					1-C (16-25)	
B20	D	O (3)	O (5)		D	D _c	D (+/-)		C	C					1-C (5-15)	
B21	D	2-1			D	D _c	D (+/-)		C	C					3-C	

Model Code	Gender	Education	Income †	Age	Ethnicity ‡	ETS	Proxy Respondent §	Active Smoking*							Radon	Calendar Year of Ascertainment
								Pack-years	Log dose	Log duration	Dose	Duration	Recency	Start Age		
B22	D	2-I			D	D _c	D (+/-)		C	C					2-C	
B23	D	2-I			D	D _c			C			C			3-C	
C1	D	2-I		4 -I **	D	D _{EN}	D (+/-)				C	C				
C2	D	2-I		4 -I **	D	D _{EN}	D (+/-)				C	C	D			
C3	D	2-I		4 -I **	D	D _{EN}	D (+/-)				C	C		C		
C4	D	2-I		4 -I **	D	D _{EN}					C	C	D	C		

- § For the site-specific analyses variables denoted by this character were included in models for Winnipeg and New Jersey only. Missouri subjects were restricted to white race and live controls. For the models in this table, recency was coded as a dichotomous variable (smoking in the 15 years prior to ascertainment or not), and since the Missouri study excluded subjects who had smoked in the 15 years prior to ascertainment ("current" smokers), the dichotomous recency variable is not relevant to this study, as it is defined as 0 for all Missouri subjects)
- * Active smoking was considered in the 5 - 50 years prior to ascertainment.
- ** Age has only 3 indicator variables in Winnipeg due to small numbers (7 control subjects) in ≥ 80 years age category.
- † Family income was available and adjusted for in Winnipeg and Missouri only.
- ‡ Relevant to and adjusted for in Winnipeg only.
- Dose Average number of cigarettes smoked per day during the period that the subject smoked (average is over smoking duration).
- Duration Age at ascertainment - age at smoking initiation. Duration was zero for never-smokers.
- Log dose Log(Cigarettes smoked per day + 6).
- Log duration Log (Age at ascertainment - age at smoking initiation), where the average value of start year amongst smokers was employed for age at smoking initiation for never-smokers.
- C Continuous variable.
- D Dichotomous variable.
- D_{EN} A dichotomous variable indicating ever or never exposure to Environmental Tobacco Smoke (ETS) exposure.
- D_c Dichotomous ETS exposure created from a continuous measure of ETS with a cut-off point at 10 (cohabitant-smokers x years) in Winnipeg, and 200 (cohabitant-cigarettes/day x years) in New Jersey and Missouri.
- I Indicator variable with a number following a dash to indicate the number of "dummy" variables representing the covariate.
- O Ordinal variable (bracket indicates number of levels).
- (+/-) Indicates that models were analysed with and without the dichotomous variable for proxy respondent versus self-respondent.
- R 'R' as a superscript on the model code number denotes restriction of subjects to those who were under 80 years of age at ascertainment, smokers of 40 or fewer cigarettes per day on average for the duration of smoking, and were 'regular smokers' (ie. They did not have periods of intermediate cessation between the time of onset and either the completion of the study or final quit date; Model B3 and B4 were analogous to models B1 and B2 respectively, but were carried out on restricted samples of subjects.
- 'R' as a variable description indicates restriction to subjects within certain limited strata defined by the respective variable.

Table C.2

Models of Type D and E

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
D1		D	O (3)	O (5)	C	D	O (5)		C		4-1												
D2		D	O (3)	O (5)	O (5)	D	O (5)		C		O (5)												
D3		D	O (3)	O (5)	C	D	O (5)		C		O (5)												
D4		D	O (3)	O (5)	O (5)	D	O (5)		C			O (5)											
D5		D	O (3)	O (5)	O (5)	D	O (5)		C		O (5)	O (5)											
D6		D	O (3)	O (5)	O (5)	D	O (5)		C		O (5)		C	C	C								
D7		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C	C	C								
D8		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C	C	C			C cum	C		C	C	C
D9		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C	C	C			C cum					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
D10		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C	C	C			C conc	C		C	C	C
D11		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C	C	C			C conc					
D12		D			C	D																	
D13 a		D	A O (3)	A O (5)	C	D	A O (5)		A C	A C	A O (5)	A O (5)						A C cum					
D13 b		D	A O (3)	A O (5)	C	D	A O (5)		A C	A C	A O (5)	A O (5)	A C	A C	A C			A C cum	A C		A C	A C	A C
D14 a		D	A O (3)	A O (5)	C	D	A O (5)		A C	A C	A O (5)	A O (5)						A C conc					
D14 b		D	A O (3)	A O (5)	C	D	A O (5)		A C	A C	A O (5)	A O (5)	A C	A C	A C			A C conc	A C		A C	A C	A C
D15		D	O (3)	O (5)	C	D	O (5)																
D16 a		D	O (3)	O (5)	C	D	O (5)		A C	A C	A O (5)	A O (5)						A C cum					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
D16 _b		D	O (3)	O (5)	C	D	O (5)		A C	A C	A O (5)	A O (5)	A C	A C	A C			A C cum	A C		A C	A C	A C
D17 _a		D	O (3)	O (5)	C	D	O (5)		C	A C	A O (5)	A O (5)						A C conc					
D17 _b		D	O (3)	O (5)	C	D	O (5)		A C	A C	A O (5)	A O (5)	A C	A C	A C			A C conc	A C		A C	A C	A C
D18		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C	C	C			C cum	C				
D19		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C	C	C			C conc	C				
D20		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C					C cum					
D21		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C					C conc					
D22		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C					C cum	C				
D23		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C					C conc	C				
D24		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C		C			C cum					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
D25		D	O (3)	O (5)	C	D	O (5)		C		O (5)		C		C			C conc					
D26		D	O (3)	O (5)	C	D	O (5)		C				C					C cum					
D27		D	O (3)	O (5)	C	D	O (5)		C				C					C conc					
D28	E1	D	O (3)	O (5)	C	D	O (5)		C			O (5)	C					C cum					
D29	E2	D	O (3)	O (5)	C	D	O (5)		C			O (5)	C					C conc					
	E3	D	O (3)	O (5) SSMI	C	D SSMI	O (5)		C		O (5)		C					C cum					
	E4	D	O (3)	O (5) OMI	C	D OMI	O (5)		C		O (5)		C					C cum					
	E5	D	O (3)	O (5)	C	D	O (5)		C		O (5)		C					C cum* overall					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ² x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
	E6	D	O (3)	O (5)	C	D	O (5)		C		O (5)		C					C conc overall ¹¹					
D30	E7	D	O (3)	O (5)		D	O (5)		C	C			C					C cum ⁴					
D31	E8	D	O (3)	O (5)		D	O (5)		C	C			C					C conc					
	E9	D	O (3)	O (5)		D	O (5)		C	C			C					C cum ⁴ overall ¹¹					
	E10	D	O (3)	O (5)		D	O (5)		C	C			C					C conc overall ¹¹					
D32	E11																	C cum ⁴					
D33	E12																	C conc					
D34	E13	D	O (3)	O (5)	C	D	O (5)	D	C		O (5) ¹		C					C cum ⁴					
D35	E14	D	O (3)	O (5)	C	D	O (5)	D	C		O (5) ¹		C					C conc					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
	E15	D	O (3)	O (5)	C	D	O (5)	D	C		O (5)		C					C cum overall ¹					
	E16	D	O (3)	O (5)	C	D	O (5)	D	C		O (5)		C					C conc overall					
	E17	D	O (3)		C		O (5)	D	C		O (5)		C					C cum F = 0.5					
	E18	D	O (3)		C		O (5)	D	C		O (5)		C					C conc					
	E19	D	O (3)		C		O (5)	D	C		O (5)		C					C cum overall F = 0.5					
	E20	D	O (3)		C		O (5)	D	C		O (5)		C					C conc overall					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
	E21	D	O (3)		C		O (5)	D	C		O (5)		C					C cum overall + site ⁶ F = 0.5					
	E22	D	O (3)	O (5)		D	O (5)		C	C	O (5)		C					C cum ⁶					
	E23	D	O (3)	O (5)		D	O (5)		C	C	O (5)		C					C conc					
	E24	D	O (3)	O (5)	C	D	O (5)	D	C		O (5)		C					C cum -site					
	E25	D	O (3)	O (5)	C	D	O (5)	D	C		O (5)		C					C conc -site					
	E26																	C cum -site					
	E27																	C conc -site					
D36	E28	D	2 - 1	O (5)	4-1**	D	4-1	D	3-1		4-1		C					5-1 cum ⁶					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
D37	E29	D	O (3)	O (5) & *site A	D	D	O (5)	D & *site	C & *site A	C A	O (5) & *site A	O (5) & *site A	C & *site A	C A	C & *site A	C & *site A		C cum ⁰	C & *site A	C & *site A	C A	C A	C & *site A
D38	E30	D	O (3)	O (5) & *site A	D	D	O (5)	D & *site	C & *site A	C A	O (5) & *site A	O (5) & *site A	C & *site A	C A	C & *site A	C & *site A		C conc	C & *site A	C & *site A	C A	C A	C & *site A
D39																		C cum					
D40																		C conc					
D41																		I-4 cum					
D42																		I-4 conc					
D43	E31	D	O (3)	O (5) & site	S (2)	D	O (5)	D & *site	C		S (5)	O (5)	C & *site					C cum		(S)			(S)

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
D44	E32	D	O (3)	O (5) & site	S (2)	D	O (5)	D & *site	C		S (5)	O (5)	C & *site					C conc		(S)			(S)
D45		D	O (3)	O (5) & site	S (2)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site					C cum					(S)
D46		D	O (3)	O (5) & site	S (2)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site					C conc					(S)
D47		D	O (3)	O (5) & site	D	D	O (5)	D & *site	C		S	O (5)	C & *site		C	(S)		C cum		(S)			
D48		D	O (3)	O (5) & site	D	D	O (5)	D & *site	C		S	O (5)	C & *site		C	(S)		C conc		(S)			
D49	E33	D	O (3)	O (5) & site	D	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C		C cum					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
D50	E34	D	O (3)	O (5) & site	D	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C		C conc					
D51	E35	D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C		C cum					
D52	E36	D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C		C conc					
D53		D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C		I-4 cum					
D54		D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C		I-4 conc					
D55	E39	D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C		C cum					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
D56	E40	D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C		C conc					
D57	E41	D	O (3)	O (5) & *site A	D	D	O (5)	D & *site	C & *site A	C A	O (5) & *site A	O (5) & *site A	C A	C A	C & *site A	C & *site A	O (5)	C cum ^Δ	C & *site A	C & *site A	C A	C A	C & *site A
D58	E42	D	O (3)	O (5) & *site A	D	D	O (5)	D & *site	C & *site A	C A	O (5) & *site A	O (5) & *site A	C A	C A	C & *site A	C & *site A	O (5)	C conc	C & *site A	C & *site A	C A	C A	C & *site A
D59	E43	D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C	O (5)	C cum					
D60	E44	D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C & *site		C & *site	C	O (5) & *site	C conc					

Model Code		Gender	Education	Income ¹	Age	Ethnicity ¹	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose ¹ x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy ¹		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender ¹	Rn x Age
D61	E45	D	O (3)	O (5) & site	O (5)	D & *site	O (5)	D & *site	C		O (5) & *site	O (5)	C			C		C cum					
D62	E46	D	O (3)	O (5) & site	O (5)	D & *site	O (5)	D & *site	C		O (5) & *site	O (5)	C			C		C conc					
D63		D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C			C		I-4					
D64		D	O (3)	O (5) & site	O (5)	D	O (5)	D & *site	C		O (5) & *site	O (5)	C			C		I-4					
D65		D	O (3)	O (5)	O (5)	D	O (5)	D	C		(S)	O (5)	C					C cum					
D66		D	O (3)	O (5)	O (5)	D	O (5)	D	C		(S)	O (5)	C					C conc					
D67 Wp g only		D	O (3)	O (5)	O (5)	D	O (5)	D	C		O (5)	O (5)	C			C		C cum					

Model Code		Gender [†]	Education	Income [‡]	Age	Ethnicity [§]	ETS	Proxy Respondent (NJ)	Active Smoking				Smoking Interactions					Radon	Radon Interactions				
Site-Specific	Combined								Log dose	Log duration	Recency	Start Age	Log dose x Proxy	Log dose [¶] x Gender	Log dose x Age	Recency x Log dose	Recency x Proxy [¶]		Rn x Log dose	Rn x Recency	Rn x ETS	Rn x Gender [¶]	Rn x Age
D68 Wp g only		D	O (3)	O (5)	O (5)	D	O (5)	D	C		O (5)	O (5)	C			C		C conc					

- SSMI Status-specific mean imputation; the respective case and control mean from Winnipeg and Missouri (combined) were employed as a fixed value for the missing income variable in New Jersey.
- OMI Overall mean imputation; the overall mean (cases and controls combined) from the other two sites was employed as a fixed value for the missing income variable in New Jersey.
- NJ Proxy was included as a model covariate in New Jersey, where it was a matching variable, and in Winnipeg in models D1 to D38 and E1 - E36; all subsequent models excluded the main effect of proxy in Winnipeg.
- § These variables were available and adjusted for only in Winnipeg and New Jersey.
- ‡ Family income was available and adjusted for only in Winnipeg and Missouri. In combined analyses (E models), unless otherwise specified, the observed control mean family income from Winnipeg and Missouri subjects (combined) was employed as a fixed value for family income in New Jersey.
- ¶ Gender varied only in Winnipeg; New Jersey and Missouri subjects were restricted to females.
- * Active smoking was considered in the 5 - 50 years prior to ascertainment; since the Missouri study excluded subjects who had smoked in the 15 years prior to ascertainment ("current" smokers), recency of smoking as a dichotomous variable was not included for this study, as it is defined as 0 for all Missouri subjects. When considered as an ordinal variable (0 through 4, with 0 designating never-smokers, and 4 designating those who smoked within the 13 years prior to ascertainment), Missouri subjects were included with coding from 0 through 3.
- ** Age has only 3 indicator variables in Winnipeg due to small numbers (7 control subjects) in highest age category.
- Log(dose) $\text{Log}(\text{Cigarettes smoked per day} + 6)$, where dose is the average number of cigarettes smoked per day during the period that the subject smoked (average is over smoking duration).
- Log(dur) Age at ascertainment minus age at smoking initiation (the average value of age at initiation was applied to non-smokers).
- C Continuous variable.
- D Dichotomous variable.
- I Indicator variables (the number following the dash indicates the number of "dummy" variables used to represent the covariate)
- O Ordinal Variable (bracket indicates number of levels)
- A Available for Automated Selection (alpha level entry = 0.2, alpha level stay = 0.2)
- Cum For A-D models up to and including model D33, the equation which generated the cumulative exposure measure did not include an adjustment for equilibrium factor (F); it was therefore by default equivalent to 1.0; Models D34 & D35 included a fixed F factor of 0.5
- ★ Except for models in which radon was calculated as "overall" (meaning one common coefficient for all sites, i.e. no between-site variability in radon slope), the combined (E) models contained terms for site and site*radon which enabled calculation of site-specific radon OR estimates, adjusted for covariates. The covariates except for those denoted '& *site' had estimated coefficients which were common across sites (assumed not to vary between sites).
- Overall radon indicates that for the combined model, only one term was fit for the radon slope, which was assumed to be constant across site (i.e. no between-site variation).
- € Combined analysis re-analysed on June 17 with an equilibrium fraction assumed F= 0.5; previous analyses did not incorporate an explicit F factor in cumulative radon exposure calculation,

- implicitly therefore assuming an F value of 1.
- ◆ Radon terms included: cum_m (overall) and two indicator terms for site=New Jersey and site=Missouri; no interaction terms with site were included as was the case in all other combined models not designated as "overall"
 - site Indicates that these combined models were analysed with a term for radon and interaction terms between radon and site (2 indicators) , but without the main effects for site (2 indicators).
 - & *site Indicates that interactions of the respective covariate with site were considered in the combined (E type) model.
 - △ Indicates that in all analyses including and following this one, the occupancy factor used in calculating WLMs, formerly assumed to be 100% among both males and females, was modified to be 80% among females and 50% among males.
 - ✓ All E models including and following models (E13,E14, and D34, D35) were analysed with a recency variable coded 0-3 in Missouri; all previous analyses containing an ordinal recency variable were incorrectly coded for Missouri.
 - ** Age has only 3 indicator variables in Winnipeg due to small numbers (7 control subjects) in highest age category (≥80 years).
 - S Indicates a stratification variable; the number which follows indicates the number of strata.

Interaction terms were centred to prevent collinearity amongst the model coefficients

Appendix D:

Covariates Unrelated to Cumulative Radon Exposure in the 5 - 50 Years Prior to Ascertainment

- Figure D.1 Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Environmental Tobacco Smoke Exposure (in Pack-Years) and Outcome
- Figure D.2 Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Education and Outcome
- Figure D.3 Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Income and Outcome
- Figure D.4 Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Age at Ascertainment and Outcome
- Figure D.5 Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Ethnicity and Outcome
- Figure D.6 Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Source of Information and Outcome
- Figure D.7 Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Year of Ascertainment and Outcome

Figure D.1

Median and Inter-quartile Range of Radon Exposure 5 - 50 Years Prior to Ascertainment in Weighted Working Level Months by Environmental Tobacco Smoke Exposure (in Pack-Years) and Outcome

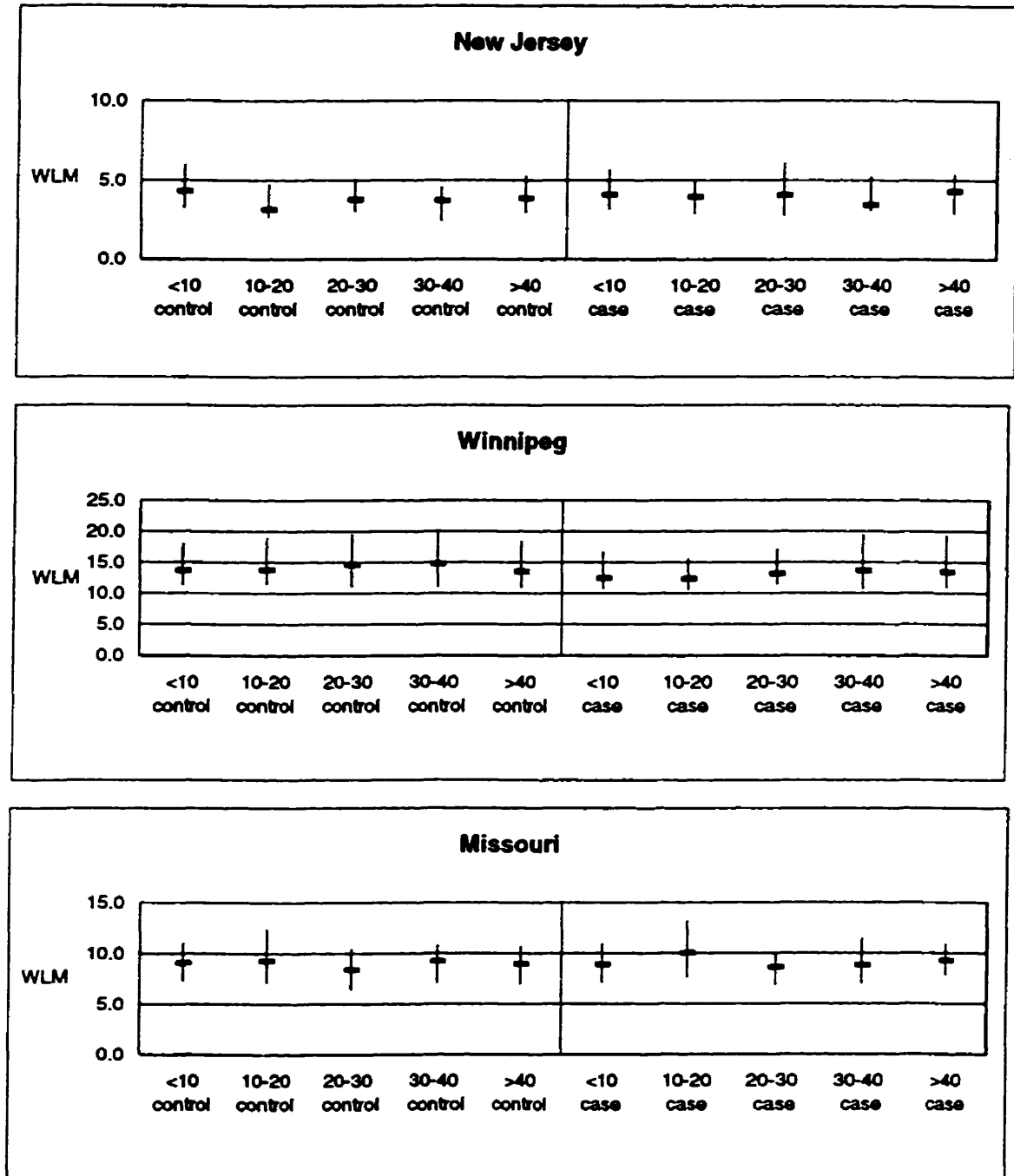


Figure D.2

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Education and Outcome**

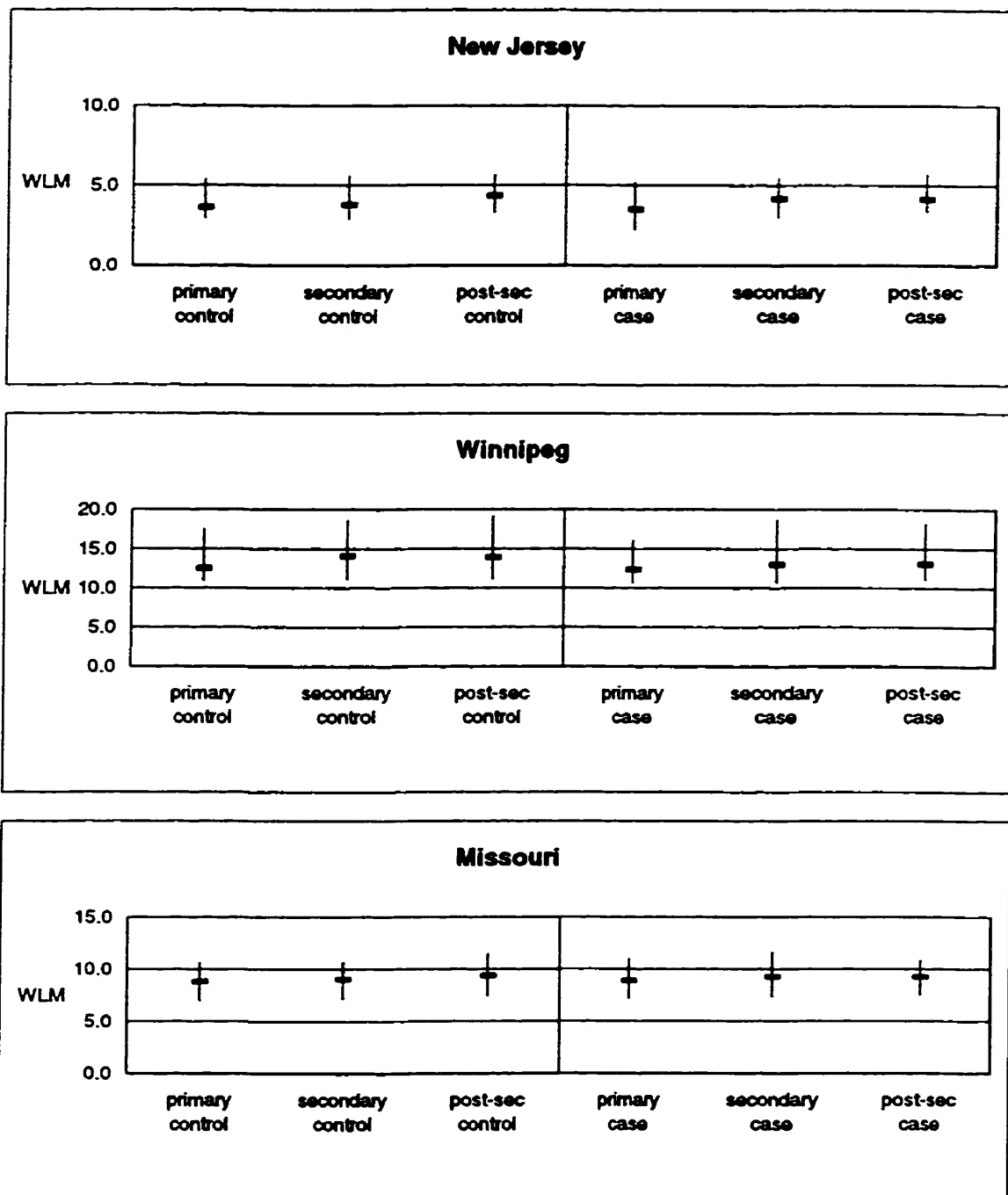


Figure D.3

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Income and Outcome**

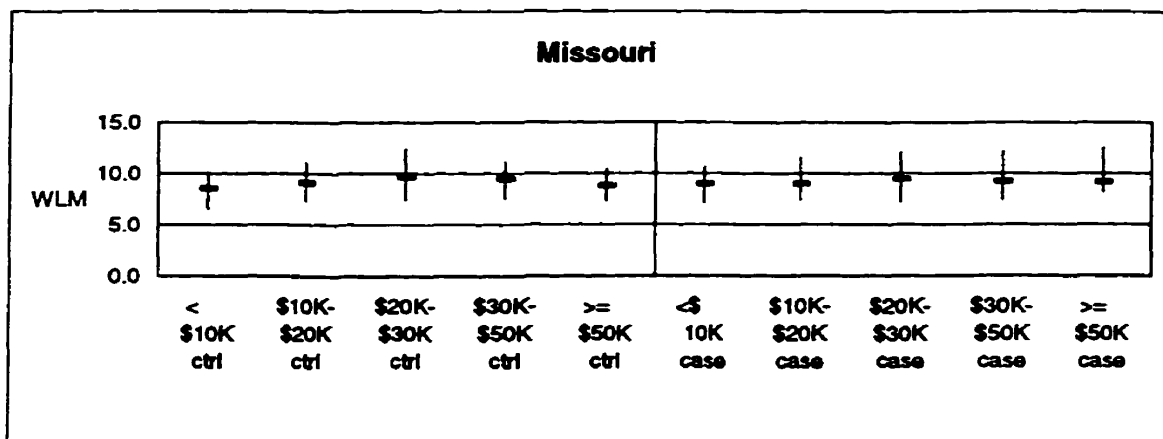
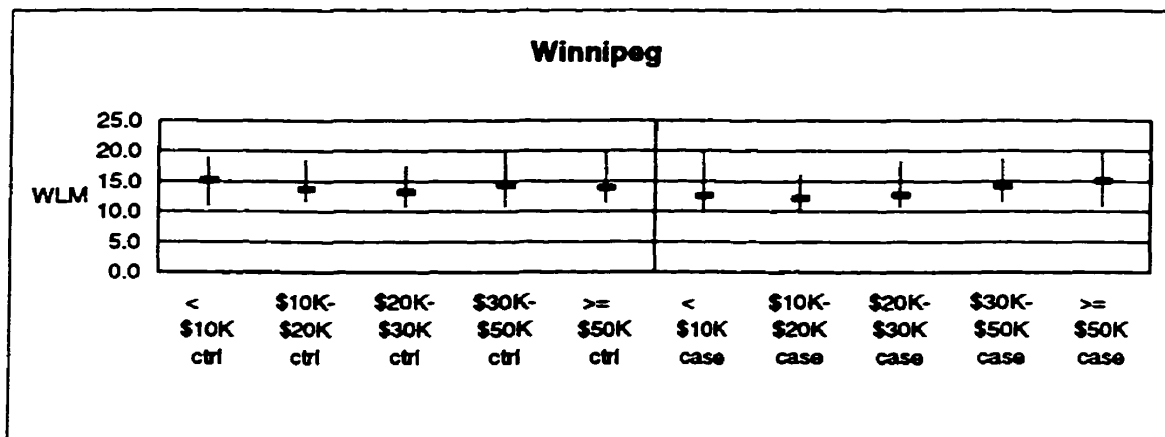
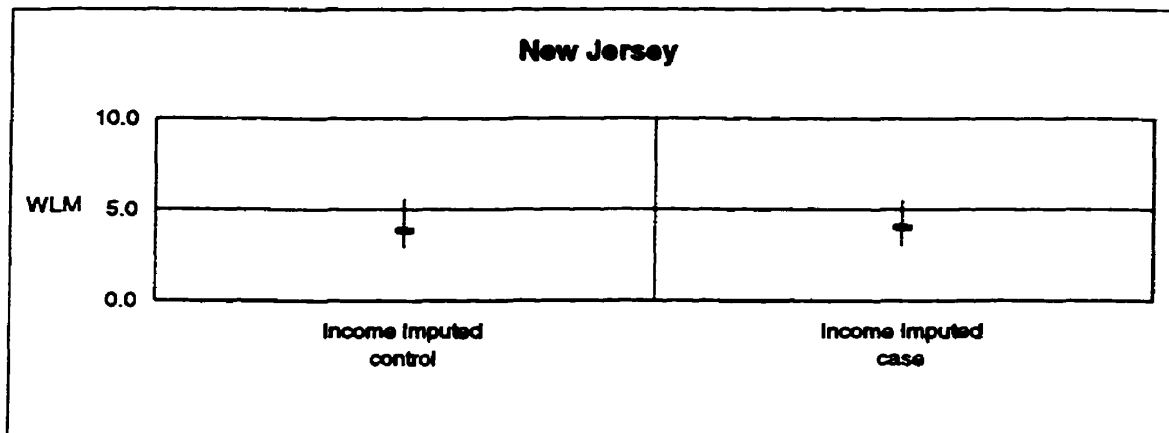


Figure D.4

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Age at Ascertainment and Outcome Figure 4.19**

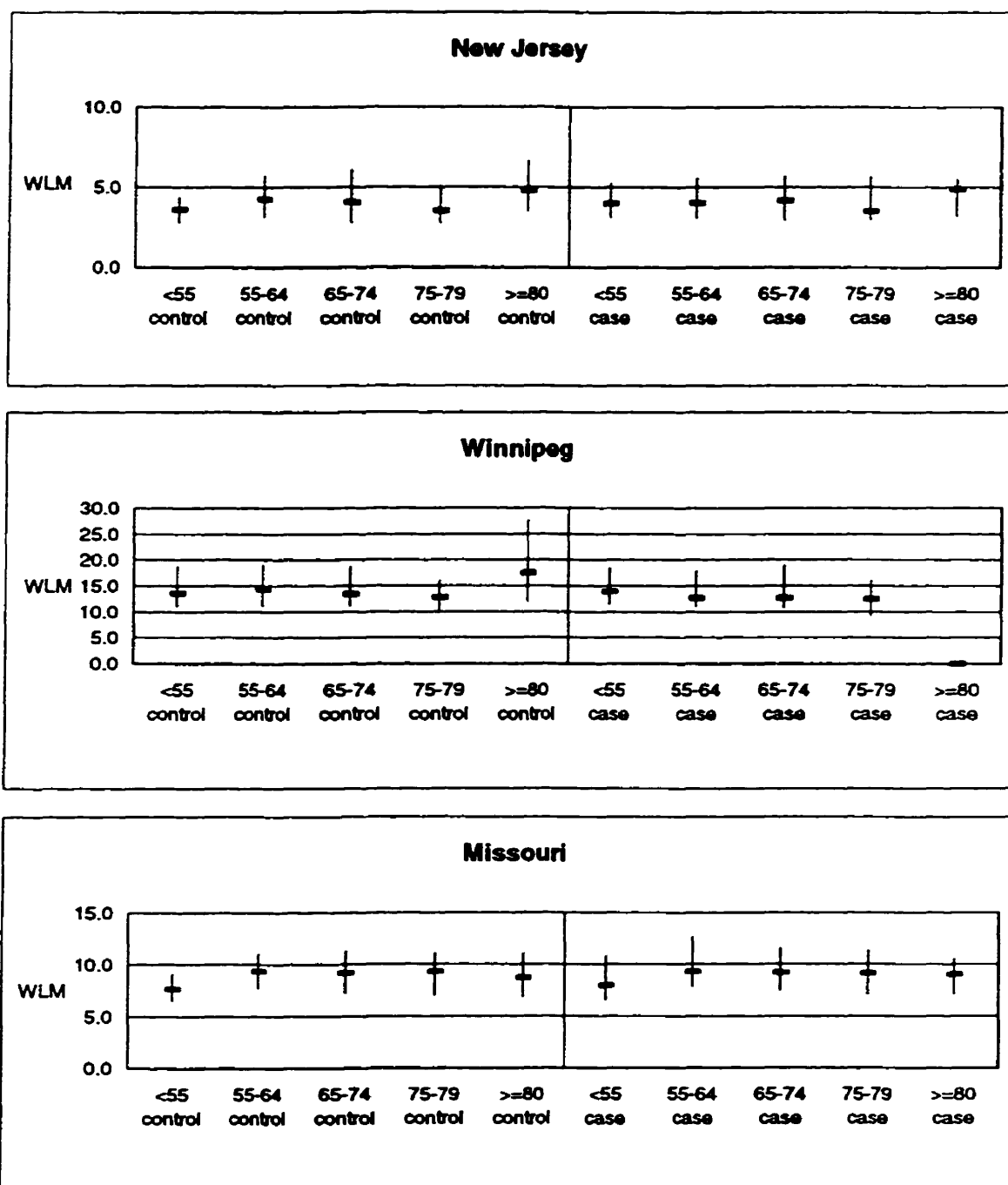


Figure D.5
Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Ethnicity and Outcome

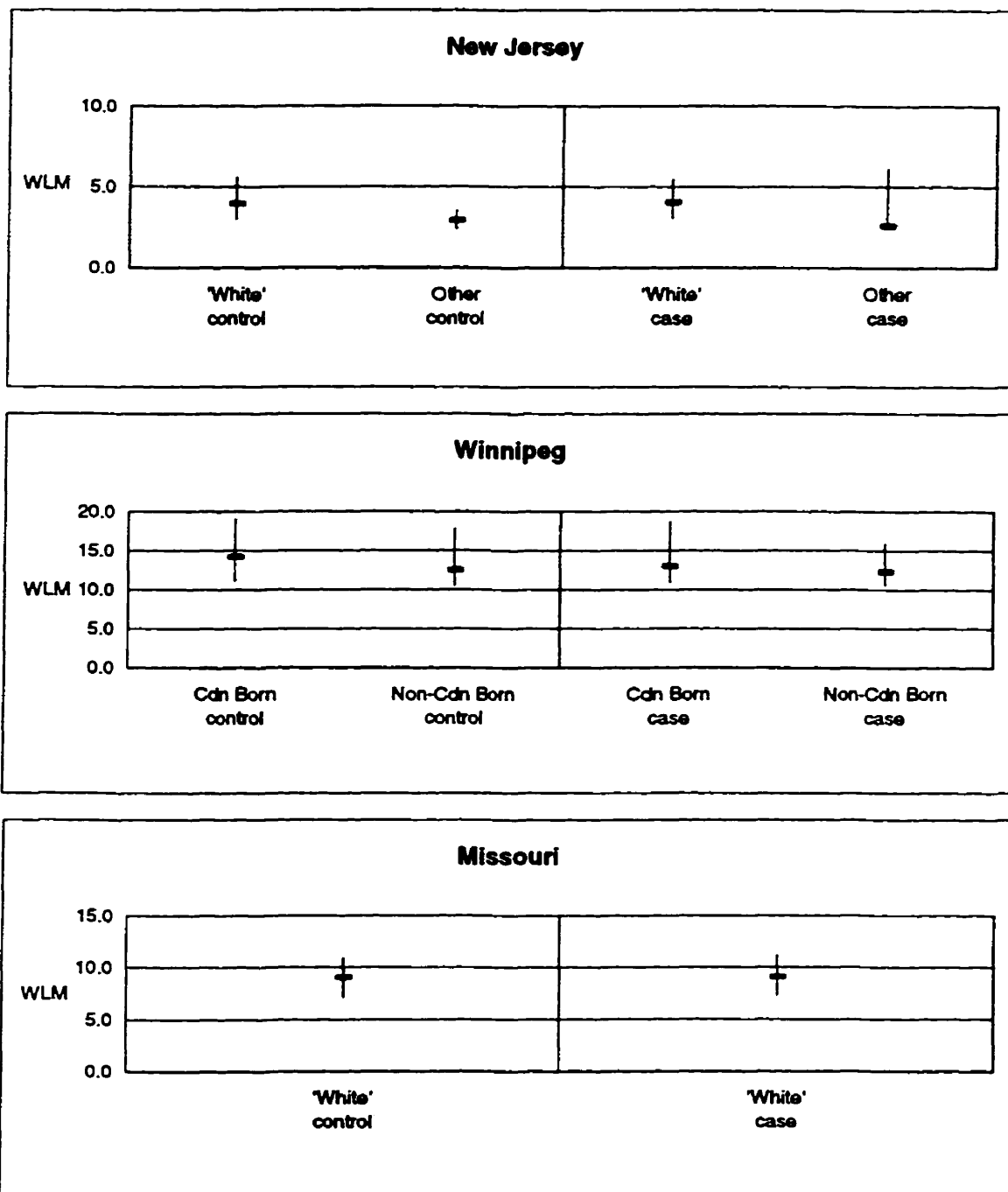


Figure D.6

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Source of Information and Outcome**

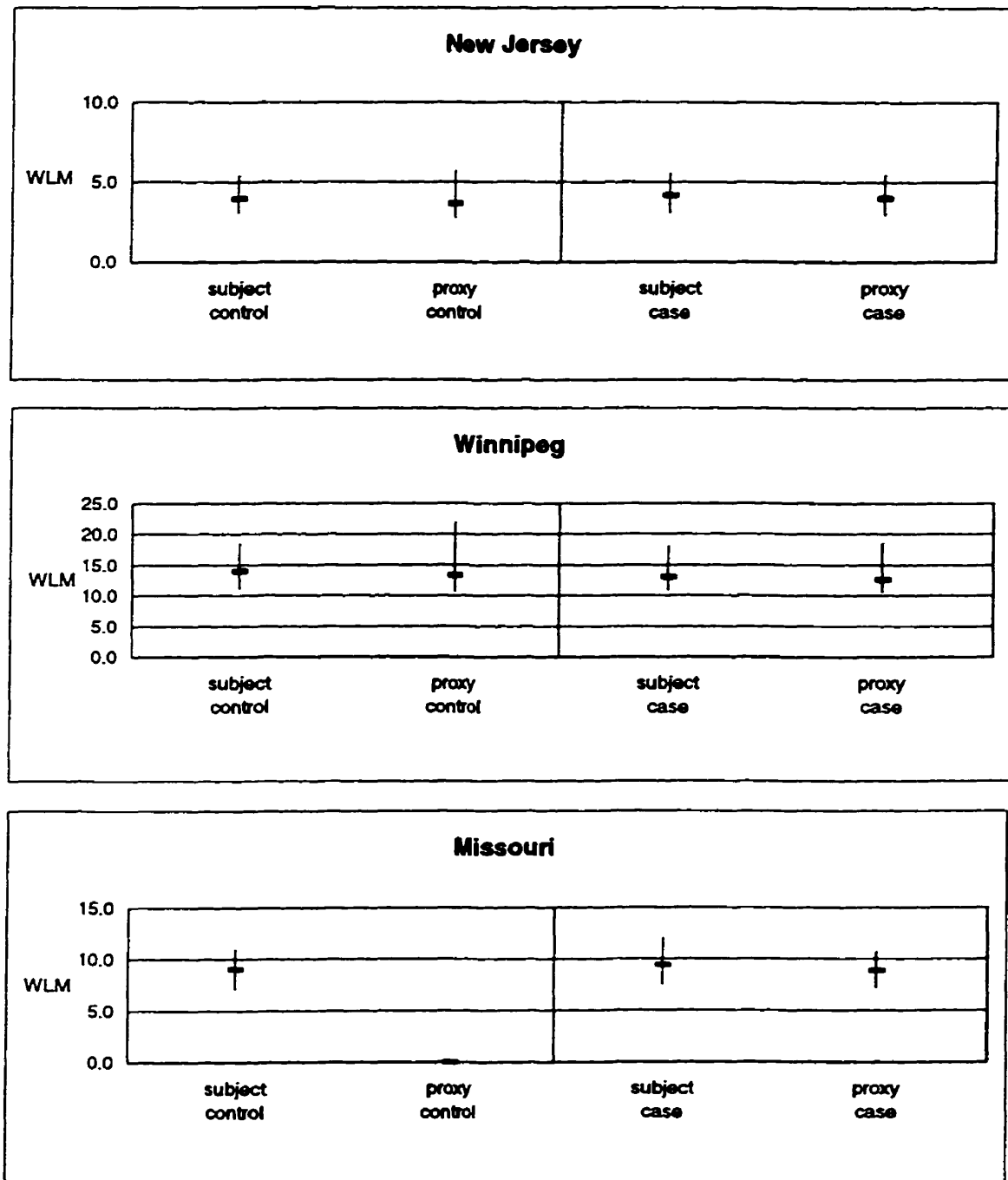


Figure D.7

**Median and Inter-quartile Range of Radon Exposure
5 - 50 Years Prior to Ascertainment in Weighted Working Level Months
by Year of Ascertainment and Outcome**

