LUNG CANCER RISK AMONGST URANIUM MINERS
THE RADIUM HILL STUDY

Arunthathi (Arul) Mylvaganam

Thesis submitted in fulfillment of the requirements for the
Degree of Doctor of Philosophy

Department of Community Medicine
University of Adelaide
August 1993

This work was supported by:
A National Health and Medical Research Council
Public Health Research and Development Committee Training Fellowship

CONTENTS

CHAPTER 2

2. Radon Epidemiology

CHAPTER 3

3. Study Design and Implementation

CHAPTER 4

4. Descriptive Epidemiological Findings

CHAPTER 5

5. Identification and Control of Confounders

CHAPTER 6

6. Temporal Effect Modifiers and Surrogate Measures of Exposure: Person-Years Based Analyses

Page

CHAPTER 7

7. Further Issues of Risk Evaluation Using Exposure-Time-Modifier-Response Models: Nested Case-Control Analyses

Page

CHAPTER 8

8. Global Perspectives: Summary, Interpretation and Implications of Findings from the Radium Hill Study

ABSTRACT

This research estimates the risk of lung cancer mortality associated with radon progeny exposure amongst a cohort of former workers at the Radium Hill uranium mine, and examines the nature of the exposure-response relationship in conjunction with relevant cofactors, covariates, temporal effects effect modifiers of effect. The Radium Hill study was conceived and initiated in the early 1980's by staff in the South Australian Health Commission. Work on this thesis began in 1988 and included extensions to study trace activities, design of data bases, compilation and validation of data, and all aspects of statistical analyses.

The Radium Hill uranium mine operated in a remote region of South Australia from 1948 to 1962. Uranium was produced under a seven-year contract between the Australian government and the governments of the United States and the United Kingdom, from 1955 to 1961; at the end of this contract it was found that mining at Radium Hill was no longer feasible. The mine stopped ore production in 1961 and was closed in 1962.

The Radium Hill study was designed as an historical cohort study. The study cohort - identified as all those on the nominal roll compiled from archived Radium Hill pay-roll records - included 2,574 workers (2,521 males and 53 females). Primary trace activities which commenced prior to my participation in the Radium Hill study yielded a trace rate of 66% over the follow-up period from 1948 through to the end of 1987.

During the course of my research, I initiated secondary trace activities which improved the study trace rate to 74% and also enabled more complete data to be obtained on those previously traced. Secondary trace activities mainly included matching of the Radium Hill study cohort with the Wittenoom Gorge asbestos study cohort in Western Australia, searches at the Perth Chest Clinic, personal contact with former Radium Hill workers, and contributing towards organizing a reunion of former Radium Hill residents. Where first-hand data were not available on individuals, proxy data were obtained through these secondary sources.

Individual exposures to radon progeny at Radium Hill were computed using individual work histories obtained from pay-roll records and radon progeny concentrations estimated from contemporary monitoring results by a team of health physicists for the purpose of this study. Only underground workers at Radium Hill were considered exposed to radon progeny. The 1,459 workers who were ever employed underground were all males. Cumulative exposures to radon progeny ranged from 0.07-112 Working Level Months (WLM) (mean = 7.7 WLM; median = 3 WLM); the average concentration of radon progeny was 0.88 Working Levels; the mean duration of employment underground was 12 months.

Risk estimation was based on 2,516 male cohort members for whom data on complete personal identification were available. Of this group, 1,849 were traced to 1987, including 606 deaths. Lung cancer was the underlying cause of death of 54 males. This comprised 9% of all deaths and 43% of the deaths from malignant neoplasms.

The average male cohort member entered the study cohort at the age of 31 when he commenced work at Radium Hill, worked for a total of 17 months, held 2.4 different jobs at Radium Hill and was aged 32 years at termination of employment at Radium Hill; he was followed up for 19.5 years and was aged 50 years at the end of follow-up.

The risk of radon progeny-related lung cancer was estimated using traditional epidemiological methods based on standardization and more recent methodology using statistical modelling techniques - Poisson regression methods based on the time-dependent allocation of person-years data, and nested case-control analyses using Cox proportional hazards models.

The potential confounding effects of age and calendar time at risk - used as a surrogate for other hidden confounders - were examined using stratification and standardization techniques. These factors were identified as confounders of the relation of radon progeny and lung cancer mortality and therefore controlled for in all further analyses.

Standardized mortality ratios showed a significantly elevated risk of lung cancer mortality amongst surface workers at Radium Hill compared to the Australian national population. The Australian national population was therefore not regarded as a suitable reference population capable of representing the lung cancer mortality experience of unexposed workers at Radium Hill. Hence, risk estimation was mainly based on internal comparisons between surface workers and categories of underground workers at Radium Hill. A comparative appraisal of risk evaluation based on internal and external references - including indirect and direct standardized methods - was made. These analyses were based on cumulative exposures lagged by five years. Results showed a significantly higher risk of lung cancer mortality amongst those exposed to over 40 WLM of radon progeny exposure at Radium Hill. Poisson regression analyses based on internal references and a linear additive excess relative risk model estimated an excess relative risk of 4.3% per WLM for all workers and 5.4% per WLM amongst underground workers at Radium Hill. These are the highest estimates of risk for radon-exposed miners reported to date.

The risk of lung cancer mortality declined significantly with time since last exposure, for a given level of cumulative exposure. This finding supports the view that radon progeny act as late-stage carcinogens under the multistage theory of carcinogenesis. Analysis based on *time since exposure windows* showed that recent radon progeny exposures (5-15 years prior to time of observation) were more hazardous than more distant exposures (15 years or more before).

The risk of radon progeny-related lung cancer mortality declined significantly with increasing radon concentration (intensity of exposure), or alternatively, increased with increasing duration of exposure for a given level of cumulative radon progeny exposure. These findings imply that the risk of lung cancer mortality per unit of exposure increases with protracted exposure, which has particular relevance to the setting of exposure limits. This significant protracted exposure effect is here identified in a cohort predominantly exposed to under 100 WLM. No other study has yet reported such an effect in this exposure range.

Age at first exposure was not a significant modifier of the exposure-response relationship in this cohort. Ages at first exposure varied little amongst the Radium Hill cohort; therefore, in its current state, this data set lacks the power to detect such a modifying effect.

The possible influence of error in measurements of radon progeny exposure was examined using a simple form of sensitivity analysis and analyses based on surrogate measures of exposure in nested case-control analyses. Sensitivity analyses were used to examine the assumptions on which exposures were extrapolated over the very early periods (1948-1952) and the last year of employment (1962) at Radium Hill, periods during which measurements of radon gas concentrations in the mine were not available. The results showed that extrapolation assumptions used in this study - direct extrapolation from exposure estimates closest in time - were reasonably appropriate for the present analyses. Analysis based on surrogate measures of exposure showed that duration of exposure was a slightly better predictor of radon progeny-related lung cancer mortality than cumulative exposure. This may imply greater measurement error in estimates of radon progeny concentrations and thereby, in cumulative exposures, compared to the duration of exposure. A review of the job exposure matrix is therefore suggested for future analyses.

Despite the limited data available, nested case-control analyses enabled examination of the roles of other relevant occupational exposures and smoking. Occupational exposures to radioactive material other than at Radium Hill and to asbestos, and smoking were found to be significant confounders of the risk of radon progeny-related lung cancer mortality.

Study participants who reported occupational exposure to asbestos were at over 4 times greater risk of lung cancer death than those who reported not having been exposed to asbestos. The interaction between asbestos and radon progeny exposure was found to be more additive than multiplicative. Due to the unavailability of quantitative estimates of asbestos exposure all analyses were based on a simple dichotomous classification, depending only on whether or not the individual was occupationally exposed to asbestos.

These findings on the role of asbestos exposure are of importance not only in studies of radon progeny-related lung cancer but also in the study of asbestos-related lung cancer. Further study of the 99 workers from the Radium Hill study cohort who are included also in the Wittenoom Gorge asbestos mine cohort is recommended. Long term follow-up and joint analyses between these studies should provide better quantitative data enabling the elucidation of the joint effects of radiation and asbestos exposures in the aetiology of lung cancer. It is noted that to date no other studies on occupational radon epidemiology have reported the effects of asbestos exposure.

After controlling for a potentially confounding effect of smoking, for a given level of cumulative radon progeny exposure, the excess relative risk per WLM was estimated to increase exponentially by 6% with each additional pack-year of smoking; this modifying effect was of marginal statistical significance. The interaction between smoking and radon progeny exposure was supra-additive, but sub-multiplicative. These findings are considered an indication of the importance of continued research based on improved smoking data.

In the final chapter of this thesis, findings from the Radium Hill study are examined in a global perspective. The Radium Hill study reports the highest risk estimates per unit exposure and the lowest cumulative radon progeny exposures of any underground miner study published to date. A statistically significant excess of lung cancer mortality has been found at relatively low levels of cumulative radon progeny exposure (<100 WLM). Furthermore, the protracted exposure effect that has been observed suggests another perspective on exposure limits. With further improvement in the follow-up and refinement in analyses, the Radium Hill study can continue to make an important contribution to the study of lung cancer mortality related to radon progeny exposure; these contributions could extend beyond the occupational arena, into the environmental setting.

DECLARATION

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge, this thesis contains no material previously published or written by any other person except where due reference is made in the text.

I agree to this thesis being made available for photocopying and loan if accepted for the award of the degree.

Arunthathi Mylvaganam

ACKNOWLEDGMENTS

The Radium Hill study commenced a decade before I knew of its existence. I am indebted to those whose vision made this study possible and all those who worked on this study from its inception. In particular: David Roder designed and coordinated the initial study trace activities and the questionnaire survey; Phil Crouch and Clair Corani studied radon progeny exposures at Radium Hill and derived the estimates of exposure used in this study; and Alistair Woodward performed the cause of death coding.

Kaye Robinson and Carol Holst have been mainly responsible for the collection and compilation of data and have patiently toiled over this seemingly endless task - I am especially, indebted to them for their meticulous contributions, dependability and the support they have provided me throughout this study.

Kevin and June Kackoschke provided valuable insight and enthusiasm which led to the success of many of the secondary trace activities. The Radium Hill community association and former residents of Radium Hill are thanked for their considerable support in the questionnaire survey. Jan Eccles and Nicholas DeKlerk contributed to the successful secondary trace activities through the Wittenoom Gorge study and the Perth Chest Clinic.

I wish to express my gratitude to colleagues at the Department of Community Medicine who have contributed towards this work in a multitude of ways. Tony McMichael and Alistair Woodward introduced me to radon epidemiology, helped me identify and structure my research hypotheses at the commencement of this work, provided valuable guidance throughout this work and commented extensively on the findings from my study and my thesis.

Apart from support as my supervisor, I am also grateful to Alistair for supporting some of my more unconventional methods, particularly my approach to the secondary trace activities.

Phil Ryan participated enthusiastically in the many brain-storming sessions on analytical techniques and cast a keen eye for detail in reviewing my work. Chris Reynolds shared some of his experiences as a former sufferer of doctoral research dilemmas. Neville Owen passed on many valuable tips on organizing my thoughts and on writing. Claudette Percy and Wendy Zweck were dependable anchors during my many visits away from Adelaide.

Vital contributions to my doctoral research were made during my overseas research fellowship which was made possible through the research fellowship awarded by the National Health and Medical Research Council, Public Health Research and Development Committee.

I am grateful to my hosts during my overseas visiting fellowships and to their support staff whose assistance was invaluable: Duncan Thomas at the University of Southern California; Jay Lubin and John Boice at the National Cancer Institute in Rockville, with whom I worked on the joint analyses of uranium miner cohorts; and Nicholas Day at the MRC Unit of Biostatistics. I am particularly grateful to Maria Cardenas for her help and her assistance in coordinating my overseas research fellowship. I am also grateful to Professors Lee Hin Ping and James Lee for the opportunities of working with them at the National University of Singapore, and to Nicholas DeKlerk for the many occasions on which I have drawn upon his resources and patient strength at the University of Western Australia. I also appreciate the award of a fellowship on radiation and health from the American Statistical Association which opened many doors, and the award of a travel grant from the Anti-Cancer Foundation of South Australia which enabled secondary trace activities in Western Australia.

The main emphasis of this thesis is quantitative and analytical. I am indebted to Duncan Thomas who provided me with a solid background on analytical techniques in cohort studies; he helped me to identify grey areas in my knowledge, clarified them, provided me with the analytical building blocks that are fundamental to this work and gave me the confidence to produce my own analytical algorithms and programs.

I owe a deep debt of gratitude to Duncan Thomas and Nicholas DeKlerk, working with whom was an exhilarating and deeply satisfying mental challenge - the most enjoyable and rewarding times I have known since I left Oxford.

I am also grateful to Duncan Thomas, Nancy Dryer and Nicholas Day who provided me with forums to present my algorithms and programs for analytical techniques in cohort studies. The peer review from Duncan Thomas, his post graduate students and Jonathan Buckley at the University of Southern California, Nancy Dryer, Kenneth Rothman, Alec Walker and others at Epidemiology Resource Incorporated, Nicholas Day, Brian Langholz, Stephen Duffy, Shiela Gore, David Clayton, David Spiegelhalter and others at the MRC unit of Biostatistics and Nick DeKlerk in Western Australia were especially valuable.

I am grateful to Professor Peter Armitage - my supervisor during my Masters Degree in Oxford who continues to guide me in my research activities - for hosting me at the Department of Statistics, Oxford University in 1989, enabling me to consult with him and others at his department, Sarah Darby, Richard Peto and Sir Richard Doll, all of whom provided valuable guidance on study design and analyses at the inception of my doctoral research. I am also grateful to Professors Nicholas Day, Norman Breslow, Geoffery Berry, Geoffery Howe and John Darroch, Ethel Gilbert, Sarah Darby, Neil Pearce and Gary Gloneck for their advice and helpful suggestions on analytical techniques, and to Phil Crouch, Andrew Johnston, Tammy Utteridge and Jill Fitch for their expert advice on radon and comments on aspects of my thesis.

This work would not have been possible if not for the strength I have derived from many personal friends: Lee and Mona have enthusiastically supported my academic pursuits and watched over for me for many years. Lee has been the sounding boards for many brain waves, mathematical models and algorithms. Carole has encouraged me to keep on at formulating 'my bright ideas'. Bob has been a tower of strength throughout my doctoral research.

I also wish to thank Mary, Bob, Cheng, Lawrence and Ken for their assistance during my overseas fellowship, and friends in America, England, Canada, Singapore, Perth and Adelaide whose support is most appreciated.

On reflection, it appears to me that the peripheral support vital to the completion of doctoral research must include good friends, practical advisors, a reliable banker (Karen Griffiths) and real estate agent (Ian Warhurst), and above all, an excellent G.P. (Chris Jenkins); with this combination, all I had to do over the past three years was concentrate on my thesis.

Last but not least, I wish to thank members of my family who have wished me well despite having often found it hard to figure what I do for a living and why I do it. I am grateful to Jano and Rama for their support and for bringing some colour into my thesis, to Chitta for his keeness to understand my work, and to Rocco for his valuable comments on aspects of my thesis. I am especially appreciative of my cousin Shantha's active support and encouragement of my academic interests since my childhood.

This work is dedicated to two people I have long admired: in memoriam to Lim Chooi Wan - Gor Ee - who was diagnosed with lung cancer at a time when I was exploring topics for my Doctoral Research and for whom I chose a study on lung cancer; and to my Mother - Tilakavati Mylvaganam nee Ambalavanar - whose experience and study on leukoderma inspired my interest in studying the determinants of diseases. Little did my mother know, that the holiday chore she assigned me when I was about eight would lead to a career and a doctoral thesis. I accidentally found the report she wrote to the Hope Ship on her illness and began pondering on the whys and wherefores of disease; coupled with a fascination for numbers and analytical techniques, analytical epidemiology must have been my undeniable destiny.

For:

Gor Ee (in memoriam)

and

Amma

LIST OF ABBREVIATIONS

AFE	Age at First Exposure - to Radon Progeny at Radium Hill
AFEG	Age at First Exposure - to Radon Progeny at Radium Hill - Group
AMFIT	Procedure for Person-Years Based Analyses in EPICURE
BEIR	Biological Effects of Ionizing Radiation (Committee)
BEIRIV	Committee on the Biological Effects of Ionizing Radiation - Fourth Report
BEIRV	Committee on the Biological Effects of Ionizing Radiation - Fifth Report
CB	Confidence Bounds
CDE	Cumulative (Relevant) Duration of Exposure to Radon Progeny at Radium Hill
CE	Cumulative (Relevant) Exposure - to Radon Progeny at Radium Hill
CI	Confidence Interval
DF	Degrees of Freedom
EPICURE	Risk Regression and Data Analysis Software
ERR	Excess Relative Risk
ERR/WLM	Excess Relative Risk per Working Level Month
IE	Intensity of Exposure - Radon Progeny Exposure Concentration at Radium Hill
JEM	Job Exposure Matrix - Matrix of Radon Progeny Exposure at Radium Hill
LR	Likelihood Ratio
LRB	Likelihood Ratio Bounds
LRT	Likelihood Ratio Test
NCC	Nested Case-Control

p, *p*-value	Statistical Significance Level
PCC	Perth Chest Clinic - Western Australia
PECAN	Procedure for Nested Case-Control Analyses in EPICURE
PYRS	Person Years at Risk
RH	Radium Hill
Rn	Radon or Radon Progeny
RR	Relative Risk
SAS	Statistical Analysis System - Data Management and Statistical Software
TSE	Time Since Exposure - to Radon Progeny at Radium Hill
TSLE	Time Since Last Exposure - to Radon Progeny at Radium Hill
TSLEG	Time Since Exposure - to Radon Progeny at Radium Hill - Group
WG	Wittenoom Gorge - Asbestos Study - Western Australia
WL	Working Level Unit of Radon Progeny Concentration Used in Occupational Epdemiology Defined as: any combination of short-lived Rn progeny in one liter of air, that will ultimately result in the emission of 1.3×10^5 MeV of potential α particle energy (see chapter 2, page 31).
WLG	Working Level Group Categorical Variable Representing Inensity of Exposure
WLL05	Working Levels Lagged by 5 years Continuous Variable Representing Intensity of Exposure

WLM	Working Level Month Unit of Radon Progeny Exposure Used in Occupational Epdemiology Defined as: the exposure accumulated by a miner exposed to 1WL over one working month (WM) of 170 hours. Cumulative exposure in WLM is therefore the sum of the products of Rn progeny concentrations (in WL) and duration of exposure (in months) (see chapter 2, page 31).
WLMG	Working Level Month Group Categorical Variable Representing Cumulative (Relevant) Exposure
WLML05	Working Level Months Lagged by 5 years Continuous Variable Representing Cumulative Relevant Exposure
WM	Working Month Defined as: 170 Hours
WMG	Working Month Group Categorical Variable Representing Cumulative Duration of Exposure
WML05	Working Months Lagged by 5 years Continuous Variable Representing Cumulative Relevant Duration of Exposure

COMMONLY USED NOTATION

w	Exposure
z, y	Modifier (Potential)
x	Confounder (Potential) Also used to denote a Joint Confounder and Effect Modifier (Potential)
s	Surrogate Measure
5_	Time Since Exposure Window - Exposures Prior to 5 year Lag
15_	Time Since Exposure Window - Exposures Prior to 15 year Lag
5_15	Time Since Exposure Window - Exposures 5 to 15 years Prior
15_25	Time Since Exposure Window - Exposures 15 to 25 years Prior
25_	Time Since Exposure Window - Exposures Prior to 25 year Lag

CHAPTER 1

1. Radiation and Health

1. Radiation and Health

1.1 Background and Importance of Study Topic

Though the association between illness and mineral ore mining was recognized in the 16th century (Agricola 1597), it was not until late in the 19th century that lung cancer was identified as a principal cause of death amongst some mining groups (Harting and Hesse 1879). In the early nineteen hundreds, radon exposure was identified as the causative agent for lung cancer amongst mine workers (Ludewig and Lorenser 1924). In recent decades, radon progeny have been identified as the largest single component of radiation exposure in the natural environment. Concern has spread from the occupational setting to the general environment. This change in emphasis highlighted the need for a more detailed understanding of radon exposure in generalized and specific settings. Environmental monitoring indicated that the levels of radon progeny concentrations in some indoor settings substantially exceeded outdoor levels and occasionally approached the levels experienced in underground mines. This has focused attention on the domestic environment as a major non-occupational source of exposure to radon progeny.

Numerous studies of underground miners exposed to radon progeny in the air of mines have shown an increased risk of lung cancer in comparison with non-exposed populations. These studies - some of which were cohort studies extending over several decades - provide the richest source of epidemiological data on radiation related lung cancer. Now that mines are designed to avoid workers' exposure to high levels of radiation, concern lies mainly with the effects of relatively small doses of ionising radiation (Anthony 1988). In the absence of long term observations of non-occupational cohorts exposed to low-levels of ionising radiation, miner populations exposed to low-levels of radon provide the most suitable longitudinal data available to study the long term effects of low levels of exposure. Occupational radon epidemiology has thus taken on an additional role in providing a basis for extrapolations to other settings.

1.2 Overview of Radiation

Radioactivity is the transformation process through which unstable nuclides - *radionuclides* - reach stability. During the process of radioactive decay, radionuclides tend towards stability by emitting excess energy, which is known as *radiation*. The time taken for a radionuclide to halve its activity through decay is known as its *half-life*; the half-life of a radionuclide is unique and unalterable.

Artificial, or man-made radiation is used for medical, military, or industrial purposes. The majority of human exposure to radiation, however, can be attributed to natural sources. These can be terrestrial or cosmic. Recent estimates by the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) suggest that half the annual effective dose equivalents received by individuals from all natural sources and three-quarters of the dose from terrestrial sources, can be attributed to a specific radionuclide - radon, and to its progeny (UNEP 1985; UNSCEAR 1988).

Although there are several isotopic forms of radon, only radon-222 and radon-220 (thoron) are found in significant concentrations in the human environment. Radon-222, the most common isotope of radon, is a member of the uranium-238 decay chain whilst radon-220 is a member of the thorium-232 decay chain. Of these two isotopes of radon, radon-220 together with its decay products makes only a small contribution (less than 20%) to occupational and environmental exposure (UNSCEAR 1982). This thesis will concentrate on radon-222 and its progeny; henceforth, *'radon'* (Rn) will refer to radon-222 and, where relevant, to its progeny.

Radon-222 and its progeny are short-lived decay products of uranium. When uranium decays, its atoms go through a series of radioactive changes, each resulting in the formation of a different element, until finally an isotope of lead remains. Each radioactive element in the series is unstable, with half-lives ranging from a fraction of a second to millions of years.

The forms of radiation most commonly emitted in the process of uranium decay are, *alpha (α) particles*, *beta (β) particles* and *gamma (γ) rays* (NRPB 1986). Alpha particles comprise two protons and two neutrons; they are therefore positively charged, massive, have low penetrating power, and can, for example, be stopped by a piece of paper. Beta particles comprise single electrons which are created when a neutron in an unstable radionuclide changes into a proton. The negatively charged β-particles are much less heavy and are more penetrating than α-particles. Gamma rays consist of a discrete quantity of energy which has no mass or charge. Propagated as waves, γ-rays travel at the speed of light and have extremely high power. They require a thick slab of lead or concrete to prevent penetration.

The intensity with which energy is dispersed as radiation transverses matter is known as *linear energy transfer* (LET). There are two types of LET; *low-LET* radiation which loses energy slowly and is thus able to penetrate deeply (e.g., X-rays and γ-rays), and *high-LET* radiation which deposits large amounts of energy over a relatively short distance (e.g., α-particles).

Evidence from epidemiological studies suggests that the major component of health risk due to Rn exposure amongst uranium miners, stems from the inhalation of short-lived α-emitting Rn progeny. Inhalation of Rn progeny results in the emission of α-particles in the lung which can cause cell damage and can ultimately lead to lung cancer (BEIR IV 1988). This work will therefore concentrate on the risk of exposure to low levels of internally deposited α emitting Rn progeny, with specific emphasis on the outcome of lung cancer.

Prerequisites essential to the task of evaluating the risk of exposure to radiation include an understanding of radiation dose quantification, the biological effects of ionising radiation, the temporal stages of radiation carcinogenesis and methods for studying radiation related health effects. These topics will now be addressed in turn.

1.3 Measurement of Radiation Dose

The number of transformations that take place each second in an amount of radioactive material is called its *activity*. The basic unit for measuring radio-activity in the international system of units - *Systeme International d'Unites (S.I units)*, is *Becquerels (Bq)*; a becquerel equals one transformation or disintegration per second.

The amount of radiation energy that is absorbed per gram of any medium of absorption (e.g., tissue) is called *absorbed dose* and is measured in units called *Grays* (*Gy*), where 1 *Gy* equals 1 *J/kg* medium of absorption. The biological effect of radiation however, also varies with the type of radiation (α, β, γ, and type of *LET*). Therefore, the measure for absorbed dose needs to be weighted for its potential to produce damage. The weighted dose is known as *equivalent dose* (ICRP 60 1991) and is measured in *Sieverts* (*Sv*), which are a function of dose in *Gy*.

The amount of radiation absorbed depends on the medium of absorption. The biological effects of radiation depend on the part of the body that is exposed, some parts being more vulnerable than others e.g., a given dose equivalent of radiation is more likely to cause fatal cancer in the lung than in the thyroid. Therefore, dose equivalents that are appropriately weighted for the different parts of the body are termed *effective dose* (ICRP 60 1991), also expressed in Sieverts. In referring to the radiation exposure of a group of people, the cumulative effective dose equivalent received by the whole group is called the *collective effective dose* (ICRP 60 1991) and is expressed in *man-sieverts* (*man-Sv*). Since many radioactive nuclides decay slowly, their radioactive effect extends far into the future and to future generations. The collective effective dose delivered to generations of people over time is called the *collective effective dose commitment* (ICRP 60 1991).

1.4 Biological Effects of Ionising Radiation

The effects of radiation on living things vary with the penetrating power of the radiation and the amount of energy it loses as it traverses matter. The biological effects of radiation which include cellular and tissue effects, depend on the nature of the target of exposure or the absorbing medium and the dose, form and *LET* of the radiation emitted.

Alpha emitters are scarcely able to penetrate the dead outer layer of the skin and are not hazardous unless they are absorbed into the body through inhalation, ingestion, or contaminated open wounds. Beta particles may penetrate about a centimeter of tissue, proving hazardous to superficial tissues; beta emitters can only harm internal organs if they are incorporated in them. The greater penetrating power of γ-rays allows them to pass through the body; γ-rays are therefore hazardous to the body, regardless of whether they are deposited internally or externally.

Once radionuclides enter the body through inhalation, ingestion, contaminated open wounds or injection (for diagnostic or therapeutic use), they may be absorbed and upon decay will irradiate adjacent cells. Radioactive emission is a random process and the effects of radiation are stochastic. Hence, exposure to radiation need not necessarily result in deleterious chemical changes.

Radiation related chemical changes are thought to occur in various ways, including effect on DNA molecules. Radiation energy, if absorbed by DNA molecules, results in a series of excitations and ionisations. Ionisation of DNA molecules may result *directly* through interactions with charged radioactive particles or *indirectly* through free radicals which are generated in the cytoplasm (water) of the cell, diffuse to the DNA and produce biological lesions.

Exposure to charged α and β particles results in direct ionisation through electrical interactions. In the case of sparsely ionising radiations such as γ and *X*-rays, about two-thirds of the biological lesions result from indirect ionisation. Ionising action is therefore determined by the quality of radiation, with the process shifting from indirect to direct as the type of radiation changes from low to high LET (BEIR V 1990). Both direct and indirect ionisation may result in highly localised damage or gross chromosomal damage, and are implicated in inherited genetic defects and the development of cancer.

Due to their densely ionising properties, particulate radiations from *high LET* α and β emitters are substantially more effective with regards to cell killing, mutagenicity, cell transformation and carcinogenic potential, than the weakly ionising *low LET* radiations such as *X* and γ rays. The *quality factor (Q)* is a measure that takes this increased potential for biological damage by certain types of radiation into consideration and provides a means of relative assessment. The quality factor *Q* is defined as the ratio of occupational exposure dose limits (BEIR V 1990). It is therefore, arbitrarily defined as a constant for a particular radiation (Cameron and Skofronick 1978).

Since the biological effectiveness of radiation also depends on the target irradiated, a comparative measure of biological effectiveness also becomes necessary. The relative ability of a radiation as compared with a standard (usually *X* or γ-rays), to create a specific disorder is given by the *relative biological effectiveness (RBE)*. The *RBE* therefore varies with the *LET* of the radiation, the *dose*, the *dose rate*, the type of cell or tissue under consideration and the endpoint studied (Broerse and Barendson 1973; Barendson 1968; Coggle 1983; Mettler and Moseley 1985; BEIR V 1990).

The *RBE* of one radiation relative to that of another is given by the inverse ratio of the doses of each radiation that will be required to produce the same biological effect (BEIR V 1990). Consequently, unlike *Q*, *RBE* is not constant for a specific type of radiation. Furthermore, though conceptually, *Q* has a similar meaning to *RBE*, by virtue of their definitions, they are not necessarily identical. Estimates of *RBE* are determined experimentally from radio-biological data and play an important role in modelling radiation exposure-response relationships.

The health effects of exposure to radiation are governed not only by the biological effects of radiation exposure; they are heavily dependent on the physiology and biochemistry of the individuals exposed. Individuals vary in their susceptibility to the damaging effects of radiation (Lewis 1987). Radiation related health effects can be somatic or genetic with perhaps the most feared somatic effect being *carcinogenesis,* the induction or cancer. Radio biological aspects of carcinogenesis are comprehensively described in an article by Adams *et al.,* (Adams *et al.* 1987), who also provide some interpretations of dose and dose-rate relationships in the context of the biological effects of ionising radiation.

1.5 Temporal Stages of Radiation Carcinogenesis and the Multistage Theory of Carcinogenesis

1.5.1 Temporal Stages of Radiation Carcinogenesis

Radiation carcinogenesis is a complex process that extends from the very early physical, chemical and cellular changes initiated by the absorption of radiation, to delayed effects that only appear many years later. The primary action of irradiation is ionisation. The temporal stages of radiation action that follow ionisation may be described as the *physical*, *chemical* and *biological* processes. The *physical stage* of irradiation includes *ionisation*, *electronic excitation, molecular vibrations and dissociation, and rotational relaxation*, all of which take a very small fraction of a second, ranging from 10^{-16} seconds to 10^{-12} seconds. This is followed by the *physico-chemical* and *chemical stage* which include *free radical* reactions that take less than a second and *biochemical changes* such as *enzyme reactions* that could take seconds, minutes or hours. The ensuing *cellular* and *tissue stage* comprises of cell division and reproductive death which may occur within hours, gastrointestinal tract damage and central nervous system damage at high doses that may take days, haemopoietic death, acute damage to skin and other organs and late normal tissue morbidity which may take months and lastly, carcinogenesis and the expression of genetic damage to offspring which may take years to manifest themselves (Franks and Teich 1986).

The process of radiation carcinogenesis is now widely accepted as a *multistage process* and, its essential features have been summarized by the *multistage theory of carcinogenesis*. The multistage theory of carcinogenesis provides a unified framework in which to view experimental and epidemiological findings on carcinogenesis (Kaldor and Day 1989). An outline of the multistage theory of carcinogenesis follows next, and this section is concluded with a presentation of aspects of interpreting the temporal factors in radiation carcinogenesis related to this thesis, in the context of the multistage theory.

1.5.2 The Multistage Theory of Carcinogenesis

Multistage models of carcinogenesis which predict an increase in cancer incidence as a function of time since exposure to some carcinogen, provide some of the most popular theories used in explaining temporal patterns in studies of cancer mortality. Several multistage models of carcinogenesis have been proposed since the introduction of the multistage theory in mathematical terms. All these models share a few common principles: first, they assume that a normal cell passes through two or more stages before becoming fully malignant; second, that the transitions between these stages are a random process and the rate of transitions may depend on the level of carcinogens to which the cell is exposed; and finally, that all cells are at risk of transitions independently.

Based on these principles, the multistage theory proposes that cancer arises from a single cell (originally, normal) which has undergone a series of heritable changes after the last of which it is capable of uncontrolled malignant replication; each of these changes may be considered a distinct stage of the process of carcinogenesis; in the absence of carcinogenic exposure, each of these changes has a low probability of occurrence and a slow progression time to the next stage. Carcinogens may act at any of these stages. Carcinogens that act at the first stage are referred to as *initiators* and those that act at later stages are generally referred to as *promoters*. The multistage nature of the process of carcinogenesis may therefore, primarily be divided into two phases - the *initiation* phase followed by the *promotion* phase; an alternate formulation proposes that the carcinogenic process is one of at least three successive stages - *initiation, promotion* and *progression* (Upton 1987). By the nature of their definition, initiators characteristically have a longer latent period to death than promoters which have fewer stages of the carcinogenesis process to transgress. Specific definitions of initiators and promoters are avoided in this work due to recent experimental evidence that has thrown the functional definitions heretofore accepted into question (Hennings *et al.* 1983; Kaldor and Day 1989).

The simple but broad framework of the multistage theory, being based on multiple and variable numbers of stages, lends itself to several multistage theories of carcinogenesis.

Multistage theories of carcinogenesis have their foundations in both experimental and epidemiological investigation. In many organ systems, the multistage nature of carcinogenesis can be demonstrated experimentally; epidemiologically, multistage models are capable of drawing a wide array of observations into a single coherent framework (Day 1984; Kaldor and Day 1989).

Although it is impossible to prove whether or not the mathematical form of the multistage model actually holds in a given situation, a number of its predictions have been verified experimentally (Borzsonyi *et al.* 1984; Peto *et al.* 1975). Experimentally, the multistage nature of carcinogenesis can be demonstrated as follows: Of two events A and B, event A followed by event B induces tumours while A or B alone does not induce tumours. In the laboratory setting A and B may comprise sequences of events with events broadly described as initiation or promotion phases. Epidemiologically however, such clarity of distinction is rarely possible due to the difficulty in identifying the sequence of causative actions; therefore, it is not always possible to establish a direct one-to-one link between stages of carcinogenesis that are epidemiologically inferred with those experimentally demonstrated (Day 1984). Because of this lack of correspondence between the experimental and epidemiological stages of carcinogenisis, the terms initiators and promoters are avoided in favour of the less specific terms *early-stage* and *late-stage* carcinogens, in epidemiological contexts.

The mathematical formulation of the multistage model thus provides a link between qualitative description of events at the cellular level and quantitative description of cancer risk in human populations. One of the simplest multistage models of carcinogenesis proposed is the Armitage-Doll model (Armitage and Doll 1961), which may be used as an example in illustrating the mechanics of the multistage model.

Under certain assumptions, in the absence of specific carcinogenic exposures, the Armitage-Doll model proposes that the incidence rate of cancer - I - at a given age - t - is proportional to a power of the age, the power representing one less than the number of stages - k - underlying the specific carcinogenic process, which can be represented by the following expression:

$$I(t) \propto t^{k-1}$$

where, the parameter k can be estimated from the slope coefficient - as the slope + 1 - in a linear fit of the logarithm of age against age-specific incidence rates. The assumptions on which the model are founded are that transitions between stages of the carcinogenic process are rare, vary little with time and that transition rates are constant (Moolgavkar 1978; Breslow and Day 1987). Under these assumptions, this model had been found to fit background age-specific incidence rates for many tumours in a variety of different populations, with k being equal to 5 or 6 (Breslow and Day 1987; Cook *et al.* 1969); thus, implying that these tumours may have risen from an underlying five or six stage carcinogenic process.

In extending the simple multistage model to situations concerning carcinogenic exposures over and above those in the background, the effect of changing exposures changes the transition rates on which the model is based. The transition rates being modified, can no longer be assumed as constant and the multistage model results in a polynomial expression of degree *k-1* (Breslow and Day 1987). These polynomial expressions on transition stages to have been fitted to experimental data and epidemiological data on exposures to known carcinogens (Whittemore and Keller 1978; Thomas 1982, 1983 and 1990). In radiation carcinogenesis, Thomas (Thomas 1990) has also fitted an alternative form of the multistage model - the *two-stage model* suggested by Moolgavkar and Knudson (Moolgavkar and Knudson 1981; Moolgavkar and Venzon 1979) - which proposes two transition stages between cell normalcy and malignancy whilst explicitly allowing for cell division and cell killing in the process of transition, and has also proposed and demonstrated another form of the multistage model - the *three-stage model* (Thomas 1990).

As discussed by Breslow and Day (Breslow and Day 1987), epidemiologically, data are seldom extensive or detailed enough to fit precise multistage models that enable estimation of specific transition rates and their interpretation. In epidemiology therefore, such models are generally used in a heuristic way to examine the phenomena predicted by the Armitage-Doll model and to draw biologically plausible interpretations from the epidemiological data (Day and Brown 1980). They also help provide further insight into which variables to use in modelling temporal effects on cancer risk.

1.5.3 Temporal Factors in Radiation Carcinogenesis in the Context of the Multistage Theory of Carcinogenesis

The relationship between cancer risk and different temporal variables corresponds closely with the behaviours predicted by theories of the multistage process; the multistage theory provides a conceptual framework within which the behaviour of various temporal variables and their relationship with cancer risk can be explained; thus, enabling classification of the specific actions of carcinogenic agents, as early or late stage carcinogens (Peto 1984; Day 1984; Breslow and Day 1987).

If one subscribes to the multistage model, it is possible to predict whether exposure acts at an early or late stage in the carcinogenic process by examining the temporal patterns in the data (Hornung and Meinhardt 1987; Breslow and Day 1987).

Two of the most important temporal effects associated with exposure-response relationships in radiation carcinogenesis are age at first exposure and time since last exposure. The role of these factors have been interpreted in the context of the multistage theory in several recent epidemiological studies (Whittemore 1977; Day and Brown 1980; Brown and Chu 1983; Hornung and Meinhardt 1987). The relevance of these temporal variables to the multistage theory may be explained through the following rationale: if a carcinogen acts at a late stage, the multistage theory predicts a rising relative risk with increasing age at first exposure and a falling relative risk with increasing time since last exposure, with the opposite being true for early stage carcinogens (Hornung and Meinhardt 1987; Breslow and Day 1987). Such interpretations may also be extended to address the modifying effect of attained age. If such relationships do hold, there may be important consequences for occupational and public health.

Further uses of multistage models of carcinogenesis in the interpretation of epidemiological data include implications of the forms of dose-response relationships, the role of latency, effects of measurement error and joint effects of several exposures. These issues have been discussed in detail by Breslow and Day (Breslow and Day 1987; Day 1984).

The Radium Hill study dataset lacks the sufficient follow-up and detail necessary to fit multistage models of carcinogenesis. Therefore, the analyses undertaken in this thesis will not include any attempts at fitting multistage models. However, attempts will be made to explain the implications of the study findings in the general context of the multistage theory.

1.6 Radiation Related Lung Cancer Risk

1.6.1 Methods of Study

Methods for characterizing the lung cancer risks associated with Rn progeny exposure in humans consist of dosimetry and epidemiology; both approaches being important and neither sufficient.

The dosimetric approach provides an estimate of lung cancer risk related to Rn progeny exposure that is based specifically on modelling the dose to target cells. Dosimetric models are all based on assumptions on radionuclide dose modifying factors, some of which are not subject to direct verification e.g., assumptions as to breathing rates, the disposition of Rn progeny in the respiratory tract, the type, nature and location of the target cells for cancer induction and the RBE. Controlled animal experiments have been the main source of knowledge on radiation dosimetry with extensions being made to human populations.

Although animal studies have provided much information on the biology of radiation carcinogenesis, estimates of radiation risk in humans still rest heavily on the data from epidemiological studies. Studies on humans however, are often fraught with problems relating to the lack of accurate data on radiation dosage and on other response modifying factors. Furthermore, since cancers associated induced by ionising radiation are generally indistinguishable from cancers induced by other causes, their occurrence can only be identified by a statistical analysis of excess incidence over the 'natural' incidence. Much of our information on human carcinogenesis is derived therefore from epidemiological sources that focus on population perspectives rather than individual outcomes (Franks and Teich 1988).

Epidemiological studies consist mainly of longitudinal cohort studies and case-control studies of exposed populations. Cohort studies include historic or retrospective cohort studies, prospective cohort studies and more recently, nested case-control studies (within the cohort).

Historic cohort studies and case-control studies have yielded most of the evidence currently available and also helped generate new hypotheses which may be tested through prospective studies. Basic epidemiological methods applied to cohort studies enable the evaluation of excess risk due to exposures of interest. More detailed study of exposure-response relationships is aided by the use of statistical and mathematical modelling, discussed in the sub-sections 1.6.2 and 1.6.3.

Though the term *'dose-response relationship'* has been commonly used in most epidemiological studies, it is not appropriate in some cases where specific dose is not measured. In epidemiological studies of uranium miner cohorts, exposure is measured in terms of Rn progeny concentration levels in the workplace and there is, currently, no information available on tissue or body dose of radiation. In such situations the use of the term *dose* is inappropriate, and it is recommended that "*exposure-response relationships*" be used instead (Thomas 1987).

1.6.2 The Role of Modelling

Statistical models for studying radiation related carcinogenesis can be considered to have a *deterministic component* that accounts for identifiable systematic effects or response-modifying effects, and a *stochastic component* which would encompass the random nature of radiation absorption and any unidentified systematic effects that are not explained by the deterministic component in the model. Factors that may influence the biological response to radiation identified to date include dose, dose rate, non-uniformity of dose distribution and other temporal factors related to dose, age, species, health status and exposure to other carcinogenic agents in combination with radiation.

Statistical modelling enables the separation of the effects due to radiation exposure and exposure-related variables from those due to the other identifiable effect modifiers. Factors other than radiation exposure may explain a sizable component of the disease patterns observed. Yet, radiation exposure remains the exposure of greatest interest in evaluating the excess health risk associated with underground mining. Studies of the effect of radiation must control for confounders of the radiation exposure-response relationship and elucidate the role of interactions and effect modifiers. Statistical modelling techniques enable the simultaneous adjustment for multiple confounders, and demonstration of the role of effect modifiers. Statistical modelling is based on postulating the functional form of the exposure-response and exposure-modifier relationships, and the stochastic variation. Due to the time-dependent nature of most exposures and effect modifiers in epidemiological studies, the more comprehensive terminology of *exposure-time-response* relationships and *exposure-time-modifier-response relationships* which provide fuller descriptions of temporal associations, has been widely advocated in recent years (Thomas *et al.* 1987).

Early difficulties that arose in handling time-dependent exposures and covariates included differences of opinion on the meaning and definition of such terms as latency (Peto 1985), dose and exposure (Thomas 1987), and difficulties in analysis including, incorrect handling of time-dependent exposures (Enterline and Henderson 1973), failure to account for multiple time dimensions (Chu 1987), latency analyses without denominators (Guess and Hoel 1977; Breslow and Langholz 1987), the failure to account for latency in constructing exposure variables, the fallacy of analyzing cumulative exposure (Peto 1985; Doll and Peto 1978), failure to consider the sequence of exposure in studies of interactions (Thomas 1982), fallacies arising from the use of incidence rates versus cumulative incidence (Peto 1985) and the lack of available software for analysing data with time-dependent exposures and covariates (Thomas 1987). These issues were reviewed at a Symposium on Time Related Factors in Cancer Epidemiology, held at the National Cancer Institute in 1985. This symposium aimed at bringing together a group of quantitatively oriented cancer epidemiologists and epidemiologically oriented statisticians in an attempt to develop clarify issues that lacked clear definition and to address difficulties in analytical issues. The proceedings of this symposium - published in a special issue of the journal of chronic diseases (Thomas *et al.* 1987) - provided a rich basis for understanding issues on time-related factors and future work. Models for exposure-time-response relationships have been extensively reviewed by Thomas (Thomas 1988), and Thomas and Whittemore (Thomas and Whittemore 1988).

1.6.3 Models of Exposure-Time-Response Relationships

The nature of the exposure-response relationships can be complex. Overall response curves are known to be influenced by two factors viz., the *cell multiplying effect* and the *cell killing effect*. As radiation dose increases it raises the probability of malignant transformation at the cellular level - the cell multiplying effect. However, when dose levels are sufficiently high to sterilize some cells - the cell killing effect - the number of cells that survive and therefore, are capable of transformation, falls with increasing dose. Gray suggested in 1965 that the overall dose response relationship for cancer resulted from the balance of these two effects (Gray 1965).

The functional form of the exposure-response relationship in radiation carcinogenesis is postulated to vary with the type of radiation, amount of exposure and temporal factors such as age at first exposure, attained age, time since last exposure and duration of exposure. The general belief was once that the excess rate of cancer was a nearly linear or linear-quadratic function of dose at low doses - nearly linear for high-LET radiation and linear-quadratic for low-LET radiation - and increased with age at exposure and time since last exposure; for low-LET radiation, longer exposures at low dose rates were thought to produce a lower risk than shorter, more intense exposure of the same total dose, whereas the reverse may be the case for high-LET radiation (Thomas 1990).

The linear-quadratic formulation of the dose-response relationship was widely interpreted in terms of the *Kellerer-Rossi theory of dual radiation action* (Kellerer and Rossi 1972). The Kellerer-Rossi theory of dual radiation action postulates that exposure to radiation can cause lesions either in a single strand (SS), or simultaneously in both strands (DS) of the DNA. The linear component of the dose-response relationship may be attributed to a single quantum of radiation inducing a DS lesion or simultaneous homologous SS lesions. The quadratic component may result from two separate quanta of radiation acting simultaneously on a pair of homologous SS lesions.

Microdosimetry theory suggested that simultaneous DS lesions were rare in low-LET radiation and common in high-LET radiation. The reduced effect resulting from prolonged low-LET radiation was therefore, interpreted under the Kellerer-Rossi theory of dual radiation action, as resulting from a repair of SS lesions; i.e., the longer the duration of exposure, the greater the probability that the first lesion has been repaired before the next mutation occurs at the same locus. Though there is a growing body of evidence that suggests protracted exposures to high-LET radiation are more, not less, hazardous than short exposures (Thomas 1987), dual theory of radiation does not attempt to explain this phenomenon (Thomas 1990). If this phenomenon is confirmed, then risk estimates derived from occupational studies of mining populations exposed to Rn progeny for only a portion of their lives may underestimate the risk from lifetime exposure in domestic environments (Thomas 1987). The *three-stage* model of carcinogenesis proposed by Thomas appears capable of explaining both the decreasing effects of protracted low-LET radiation and the increasing effect for high-LET radiation; this model also appears consistent with the observed exposure-response relationships and the temporal modifying effects of age at exposure and time since last exposure (Thomas 1990). In proposing the three-stage model, Thomas also showed some limitations of the Moolgavkar-Knudson two-stage model, that rendered it unsuitable in radiation carcinogenesis (Thomas 1990).

The linear linear-quadratic formulations were based on the biophysical model that predominantly characterized single-track (i.e., resulting from entry of one quantum of radiation) events with a linear dose-response function for low doses at low dose rates, and dual-track events resulting in a quadratic dose-response relationship for high doses received at high-dose rates. This model has been challenged in recent years by the view that only single-track events contribute to biological damage, thus, advocating the linear dose-response function in preference to the previously postulated linear-quadratic function (BEIR V 1990; Thomas 1990). However, as stated in the BEIR V (BEIR V 1990) report, "an unequivocal choice is yet to be made", and research into the functional form of the exposure-response relationship, continues.

Other important questions which still remain unanswered regarding exposure-response relationships in radiation carcinogenesis include issues of threshold doses and the specific action of repair mechanisms (IAEA 1986). Though some tolerance level is considered likely, it is yet not known what threshold exists below which radiation effects could be discounted (Anthony 1988; Southwood 1987). It is also considered most unlikely that there is no repair mechanism at work. These issues were summarized in the opening address of the British Nuclear Energy Society conference on health effects of low dose ionising radiation by Anthony in 1988 (Anthony 1988), who further suggested that: answers on threshold and non-linearity of the cause effect relationship are most likely to emerge from micro or cellular radio-biology of individuals in conjunction with well designed epidemiological studies; individual risk estimates could then be extended to provide societal estimates. Risk estimates and relationships thus obtained would have to be validated by epidemiological studies. Estimates of radiation related risk can then be refined through iterative application of the process of obtaining estimates for individuals, extending them to populations and validating the findings.

Since Rn progeny emit high-LET alpha radiation, the exposure-response relationship may be expected to be linear. In this thesis, linear dose-response models will be fitted to the data from the Radium Hill uranium miner study and departures from linearity will be examined in turn by fitting an additional quadratic term to the linear model and also by fitting separate exponential and power models.

1.6.4 Risk Assessment

Risk assessment involves *risk estimation* and *risk evaluation*. Risk estimation is the process whereby risk is expressed in numerical terms, and is the focus of this work. Risk evaluation deals with broader issues such as the acceptability of risks and appraisal of risks in terms of their social costs. The stages of quantitative risk assessment may be described as *modelling the nature of the exposure, the metabolism of absorption, dose to sensitive tissue, the exposure-time-response* and *dose-response relationships* and finally, *performing sensitivity analyses to investigate the significance of uncertainties in the models used and the parameters estimated.* These aspects are described in the Fifth report of the Committee on the Biological Effects of Ionizing Radiation (BEIR V 1990). Similar issues with particular reference to social and scientific problems of radiation risk assessment in the context of predictive modelling are summarized by Crouch (Crouch 1987).

Cancer risk may be estimated in various ways. *Additive* risk expresses the number of *excess* cases per unit of time per unit of measured exposure, in a given number of exposed individuals. The *multiplicative* or *relative* risk model expresses the ratio of the risk in the irradiated population to that in a non-irradiated control group. Under the additive risk model, it is postulated that the risk due to exposure adds to the background risk, whereas, the relative risk model postulates that the risk due to exposure is a multiple of the background risk.

Recent findings (BEIR IV 1988; ICRP 60 1991; Xuan *et al.* 1993; Hornung and Meinhardt 1987; Lundin *et al.* 1979; Whittemore and McMillan 1983), indicate that relative risk models are more suitable for evaluating the excess risk of lung cancer mortality in relation to Rn exposure; in the words of the BEIR V committee "dose-dependent excess of cancers is now more compatible with *relative* risk estimates than with previous *absolute* risk estimates; the Committee believes that the constant absolute or additive risk model is no longer tenable" (BEIR V 1990); at present therefore, preference is given to relative or multiplicative risk models rather to purely additive models.

It is however, emphasized that the constant relative risk model should only be considered as an approximation, because recent epidemiological studies (summarized in chapter 2) show that the relative risk is likely to be time-dependent and varies with time since exposure and attained age. Alternatives to the constant relative risk model - modified relative risk models - which allow for the variation of relative risk with time since exposure and other temporal factors, have therefore been proposed; these include the BEIR IV model which allows for the variations in time since exposure and age at risk (BEIR IV 1988), the three component model proposed by Hornung and Meinhardt (Hornung and Meinhardt 1987), and the model recently proposed by Jacobi *et al.* (Jacobi *et al.* 1992) - the GSF model - which considers the age-specific excess rate of lung cancer as a function of age at exposure and time since exposure.

Samet and Hornung (Samet and Hornung 1990) provide a comparative appraisal of five models of risk assessment viz., the NCRP model absolute risk model (NCRP 78 1984), the ICRP constant relative risk model (ICRP 50 1987), the BEIR IV time-dependent relative risk model (BEIR IV 1988), the EPA constant relative risk model (US EPA 1987), and the NIOSH time-dependent relative risk model (NIOSH 1987). Samet and Hornung conclude that constant relative risk models may not be biologically appropriate in view of the declining risk with time since exposure (BEIR IV 1988; Hornung and Meinhardt 1987) and that the epidemiological data from studies of miners are not consistent with an additive model (Hornung and Meinhardt 1987).

These findings steered the course in the choice of risk assessment models used in this thesis. The evaluation of radon progeny related lung cancer mortality risk undertaken in this work will therefore be based only on relative risk models including both, constant and time-dependent relative risk models. The specific forms of models used in this thesis will be discussed in conjunction with their application in the analytical chapters of this work.

1.7 Importance and Scope of This Work

1.7.1 The Importance of This Work

The Radium Hill uranium mine operated in South Australia from 1948-1962 and levels of cumulative individual exposures to Rn progeny were relatively low, compared with other underground uranium mines of that period. It is therefore a particularly important setting for an epidemiological study of the health effects of exposure to Rn progeny.

The value of the Radium Hill cohort is enhanced by the data that are available, including reconstructed estimates of Rn progeny levels, comprehensive work histories and estimates of individual exposure to Rn progeny, and some additional data on smoking histories and other occupational exposures in radiation and asbestos related underground mining.

The comprehensiveness of the data available and the many years of follow-up makes the Radium Hill cohort a rich independent source of epidemiological data. However, it is recognized that the analytical strength of this study, may be better realized in joint analysis with the many overseas studies conducted. As noted by the BEIR IV committee, the quality of data on exposure, possible confounders, and disease outcome from each individual study, are variable and fraught with uncertainties; therefore it is important that no study, however comprehensive, should be regarded in isolation; findings from individual studies should be compared and combined to enable cross-validation and support cross-references (BEIR IV 1988). Efforts are being made to perform such comparative and combined analyses of the Radium Hill cohort through collaborative work in conjunction with the National Cancer Institute of the United States.

The opening of the Roxby Downs Uranium Mine in Olympic Dam, South Australia which commenced ore production in 1988, makes the analyses and findings from the Radium Hill cohort study particularly topical and of interest in the Australian context.

1.7.2 The Scope of This Work

The aim of this work is to provide a substantive analysis of the Radium Hill cohort study with a view to addressing several of the questions that are topical in occupational radon epidemiology.

This is essentially an epidemiological investigation. The nature of this investigation is fundamentally analytical involving the selection of analytical and inferential methods in epidemiology and statistics and the extensive application of quantitative techniques.

The derivation of analytical methodology is not the aim of this thesis; the primary aim is to identify and apply appropriate analytical techniques of study design and analyses. The choice of analytical techniques used in this work will be based on the application of available methods of study design and analyses which, when necessary, will be adapted for the particular use in this study.

The suitability of each analytical technique will be assessed through comparative appraisal of analytical methods. Many of the research questions will therefore be examined through alternate analytical methods; analyses will also be repeated through various alternative methods of investigations. However, this work is not intended at finding the *best* analytical methodology, but at examining the research questions without being constrained by the shortfalls of any particular analytical method. Each method will be examined with reference to this particular investigation.

The ultimate aim of this work is to estimate the magnitude of the risk of lung cancer mortality associated with radon progeny exposure amongst Radium Hill uranium mine workers, and to examine the nature of the exposure-response relationship in conjunction with relevant cofactors, covariates, temporal effects and modifiers of effect.

This thesis is structured as follows. Chapters 1 and 2 summarize the relevant background on radiation and health and radon epidemiology.

Chapter 3 identifies the specific aims of the Radium Hill study, provides a comprehensive description of the Radium Hill study design and describes the trace status of the study cohort. Appendices A and B provide supplementary information on the Radium Hill study cohort and data collection procedures. Chapter 4 - the descriptive epidemiology chapter - also summarizes the demographic characteristics of the study cohort.

Chapters 5-7 - the analytical epidemiology chapters - provide comprehensive analyses of the Radium Hill cohort study with a view to evaluating the risk of lung cancer mortality associated with Rn progeny exposure. Analyses presented in chapters 5 and 6 use analytical methods for cohort studies based on person-years-at-risk; chapter 5 addresses the issue of confounding whilst chapter 6 focuses mainly on temporal issues related to effect modifiers and surrogate measures of exposure. The final analytical chapter of this work - chapter 7 - is based on analytical methods for nested case control study designs which are used to study exposure-time-modifier-response relationships and the effect of cigarette smoking and other occupational exposures to radioactive materials and asbestos. Each analytical chapter is self contained in its description of analytical methodology. The basic epidemiological background on the processes of epidemiological enquiry and inference are presented in appendix C. A brief overview of the main analytical tool used for mathematical and statistical modelling in chapters 5-7 of this work - EPICURE (EPICURE 1992) - is provided in appendix D.

Chapter 8 - the final chapter of this thesis - provides an overall summary of the findings from this study of the cohort of former workers at the Radium Hill uranium mine, interprets these findings, and discusses the implications of the conclusions drawn from this study in a global perspective with suggestions for future directions in the Radium Hill study.

CHAPTER 2

2. Radon Epidemiology

Tables and Figures

2. Radon Epidemiology

2.1 Properties of Radon

Radon is a very dense, chemically inert gas that is tasteless and odourless. Radon is denser than air and highly soluble in water (Weast 1985); it is also known to be a highly mobile radionuclide. Since the source of radon - radium-226 - is found in the earth's crust, radon (Rn) is ubiquitous throughout the geosphere, biosphere and atmosphere (IARC 1988).

When the gaseous Rn is released into the air, (e.g., as happens in uranium mining), it decays into non-gaseous products that can attach themselves to particles in the air; the probability of attachment is high and increases with the concentration of particles in the air. *Radioactive equilibrium* is said to occur when the *activity concentration* of every short-lived Rn progeny present in air is at the same level as that of Rn. Individual activities of Rn progeny cannot exceed Rn activity.

The concentration of Rn progeny levels in air varies with ventilation rates. If the air is not circulated, the Rn progeny concentrations increase and build up to equilibrium level. The circulation of air reduces *plateout* - Rn progeny particles anchoring to walls and surfaces - by flushing out decay products before they can increase their concentration in air. Therefore, if air is circulated the Rn progeny are given less chance to increase in concentration to equilibrium levels; they taper off at lower concentration levels. A major contribution of human exposure results from the inhalation of the short-lived decay products which attach themselves to dust particles in the air. Radon gas which is released into dustier or less ventilated surroundings therefore poses a greater health hazard.

2.2 Measurement of Radon Progeny Concentrations

2.2.1 Measurement Criteria

Measurement of Rn progeny *concentration* is based on the total *potential* α energy concentration of the Rn decay product mixture present in air. The *potential* α energy of a decay product atom is defined as the total α energy emitted during the decay of the atom to the long level decay product, lead-210. Radon decay product concentration in air is therefore the sum of the α energies of all short-lived Rn progeny present per unit volume of air.

The individual activity levels of Rn progeny cannot exceed levels of Rn activity. Radon progeny concentrations are often expressed in terms of *equilibrium equivalent concentration* (EEC_{Rn}). Non-equilibrium mixtures of Rn and Rn progeny can be defined in terms of an *equilibrium factor* (F), which is defined as the *ratio of potential* α *energy concentration* of the decay product mixture to the corresponding concentration if they were in radioactive equilibrium with Rn. EEC_{Rn} is therefore equal to the product of F and the *activity concentration* (αR_n) of Rn. With the *activity concentration* (αR_n) of Rn generally being expressed in terms of Bq/m^3 of air, EEC_{Rn} is also measured in Bq/m^3.

The *cumulative potential energy exposure* (E) to short-lived Rn progeny is defined as the product of a conversion factor between potential α energy and the equilibrium equivalent activity in air (k), the time spent in the area (T), the equilibrium factor for the decay product mixture (F) and the activity concentration of Rn in air (αR_n);

$$i.e., \qquad E = k*T*\alpha R_n = k*T*EEC_{Rn}$$

Cumulative exposure to inhaled Rn progeny is measured by the potential α concentration of the short-lived Rn decay product mixture in inhaled air, integrated over the exposure time in that air space; expressed in terms of an additional time component of hours, units of measurement for *cumulative exposure* may be expressed as Jh/m^3 in SI units.

In occupational radon epidemiology amongst uranium miners where the α emitting progeny of Rn pose the major component of health risk, potential exposure to Rn progeny concentrations in air is conventionally measured in *working-level*s (WL*)*. Consequently, cumulative exposures over time are measured in *working-level months* (WLM*)*.

A WL is defined as any combination of short-lived Rn progeny in one liter of air, that will ultimately result in the emission of 1.3×10^5 MeV of potential α particle energy. A WL is equivalent to 2.08×10^{-5} J/m^3 of air in SI units and in terms of *pCi,* the original unit of measurement for activity, a WL *equals 100pCi/l (activity concentration of* Rn *at equilibrium).* A WL corresponds approximately to the potential α energy concentration of short-lived Rn progeny that are in radioactive equilibrium with a concentration of 3.7×10^3 Bq radon/m^3 of air; i.e., 1WL = $3.7 \times 10^3 Bq/m^3$ (EEC_{Rn}).

A WLM is defined as the exposure accumulated by a miner exposed to 1WL over one working month of 170 hours. Cumulative exposure in WLM is therefore the sum of the products of Rn progeny concentrations (in WL) and duration of exposure (in months).

2.2.2 Evaluation of Concentrations in Underground Mines

Individual exposure estimates used in most retrospective cohort studies of underground miners have been based on reconstructed estimates of Rn progeny concentrations. The reconstruction of early exposures was generally based on measurements of Rn that were made on air samples taken periodically during the operation of the mine for purposes such as ventilation control and routine monitoring. In the absence of Rn progeny concentration measurements, the radioactive equilibrium factor is used as a conversion factor in deriving estimates of Rn progeny concentrations from measurements of Rn concentrations.

Projections of individual exposures were then made on the basis of mine or workplace averaged measurements and the duration of exposure. Estimation techniques and averaging techniques varied considerably between mines. The obvious impact of this variation on estimates of individual exposure, emphasizes the need for careful assessment of measurement and estimation techniques in any comparative appraisal between studies. Therefore, for each of the studies reviewed in this chapter, a brief outline and reference are also provided as to the basis of exposure evaluation and the techniques that governed the estimation of individual exposures.

2.3 Occupational Radon Epidemiology: Overview

Since the identification of elevated incidence of lung cancer amongst underground miners, epidemiological studies have shown that the cause of the increase was occupational exposure to Rn progeny (IARC 1988). The high exposure levels experienced by miners before the seriousness of the hazard was fully recognized provided sufficient numbers of cancer cases for research into quantitative relationships between Rn progeny exposure and lung cancer incidence. These early studies were however, hampered by the lack of data on individual exposure levels and employment records. It was therefore difficult to obtain accurate risk estimates from these studies. Furthermore, since these findings resulted in greater controls being placed on the recommended standards of individual exposure limits it was now necessary that risk estimates be available for these lower levels of exposure. Though the extrapolation of findings from high-dose epidemiology did provide some estimates of low-dose risk, these estimates had to be refined through epidemiological studies on populations exposed to lower levels of exposure. Consequently, several epidemiological studies were initiated amongst underground mining populations which have resulted in a steadily improving basis for calculating risk. These improved data coupled with the more sophisticated analyses techniques ought to provide more accurate measures of risk at lower levels of exposure than those obtained through extrapolations from high-dose studies.

The principal sources of epidemiological data on miners exposed to Rn progeny have been the Canadian, Colorado Plateau, Czechoslovakian and Swedish mining populations. More recently, studies have also been reported from China, France and the United Kingdom. In addition to providing a basis for obtaining estimates of radon-induced lung cancer risk, findings from these studies have also helped identify potential risk and response modifiers; most importantly, the role of ventilation in mines as a risk modifier and the age of individuals exposed, other temporal factors regarding exposure and the smoking habits of individuals as possible response modifiers.

Epidemiological studies on the risk of lung cancer associated with Rn progeny have been extensively reviewed by several task forces and individuals with the aim of elucidating the role of smoking and temporal characteristics of exposure (IARC 1988). These efforts include work by the United Nations Scientific Committee on the Effects of Atomic Radiation, (UNSCEAR 1977); the Committee on the Biological Effects of Ionizing Radiations (BEIR III 1980; BEIR IV 1988; BEIR V 1990); the International Commission on Radiological Protection, (ICRP 32 1981; ICRP 50 1987); the National Council of Radiation Protection and Measurements, (NCRP 78 1984); Thomas *et al.*, (Thomas *et al.* 1985); the United States Department of Energy (US DOE 1988 and 1992), SENES Consultants (SENES 1984; SENES 1990), Lundin *et al.* (Lundin *et al.* 1971) and most recently, the National Cancer Institute of the United States National Institutes of Health (NCI 1992).

Risk estimates from these studies have been progressively revised with reanalyses over time. The detailed description of the evolution of analyses and revisions on individual studies have been extensively reported and summarized elsewhere; therefore, repetition of these details is minimized in this work. For each of the studies reviewed in this work, study characteristics are briefly outlined in section 2.4 which concludes with a presentation of analytical findings that are relevant to this thesis, summarized in tabular form and a discussion of these findings. This chapter is concluded in section 2.5 with a summary of the current needs in Rn epidemiology and the relevance of the Radium Hill study in addressing these issues.

2.4 Occupational Exposure to Radon Progeny: Review of Individual Studies

2.4.1 Canadian Studies

I. Preliminary Investigations

Canadian studies on occupational radon epidemiology have centered around three distinct populations viz., the Eldorado Nuclear Ltd., Newfoundland flourspar and Ontario miners.

The principal uranium mines under Eldorado Nuclear Ltd., were located at Port Radium in the Northwest Territories and Beaverlodge in nothern Saskatchewan. The Port Radium mine operated from 1930 to 1960 which included a brief closure between 1940 and 1942; the Beaverlodge mine started operations in 1949 and closed down in 1982. The Beaverlodge mine commenced full production of ore only in 1953, by which time ventilation and other safety procedures had been substantially increased in uranium mines. Therefore, Beaverlodge workers experienced lower levels of exposure than those at Port Radium. Since employment records were not available for workers at Port Radium prior to 1940, estimates of exposure could not be made for this period. A pilot study of miners at Port Radium reported lung cancer rate ratios for workers by three exposure groups based on the location and duration of work, viz., *surface, underground for < 5 years* and *underground for > 5 years* (Grace *et al.* 1980). A study of all male workers (14,022 with complete employment records) employed at the two uranium mines, and the uranium and radium refinery at Port Hope in Ontario reported standardized mortality ratios and directly standardized mortality rates for exposure groups based on location of work with reference to the Canadian population for 64 disease categories (Nair *et al.* 1985).

These studies were followed by several detailed studies which are summarized below.

II. The Beaverlodge Study

The Beaverlodge cohort was further studied using reconstructed estimates of individual exposures (Howe *et al.* 1986). In each occupied location of the Beaverlodge mine, routine measurements of Rn progeny concentrations were made several times a month since 1967. Individual exposures for this period were estimated from monthly average Rn progeny concentration measurements and duration worked in each location of the mine. Radon progeny levels prior to 1967 were estimated from Rn gas measurements from air samples that were taken for ventilation control purposes and annual equilibrium factors; concentrations were estimated on a workplace basis from November 1 1966, and on a mine average basis prior to that. Annual median Rn progeny estimates were then used to evaluate individual exposures, on an annual basis for those exposed after 1966 and as a single lifetime total for those exposed before then (Frost 1983). After the exclusion of workers with incomplete employment histories and those who worked at other uranium mines the analysis included 8,487 workers (77.5% of the total Beaverlodge cohort) and 65 lung cancer deaths.

Lung cancer risk was estimated using person-years-at-risk and a five-year lagged exposures. Expected deaths were calculated through indirect standardization with reference to age and calendar-year specific Canadian male mortality rates from 1950 to 1980. Weighted linear regression techniques were used to estimate relative and attributable risks for various ranges of exposure, age at first exposure and age at observation. Salient findings included a highly significant linear dose-response relationship and an increased risk of lung cancer with a relative risk estimate of 3.28% per WLM and an attributable risk coefficient of 20.8 per WLM per 10^6 PYR. Age at first exposure was a significant effect modifier and the interaction of exposure with age at observation was found to fit a relative risk model well. Subsequent reevaluation of exposure estimates have resulted in individual exposures being revised to considerably higher levels (SENES 1991; Chambers *et al.* 1992). Consequently, the revised risk estimates - a 1.3% increase in excess relative risk of lung cancer mortality per WLM (95% CI: 0.59%-2.98%)- are lower than those previously published.

III. Port Radium Study

The Port Radium cohort studied using similar methods of selection and analysis as the Beaverlodge study was reported in 1987 (Howe *et al.* 1987). The study included 2,103 workers (78% of the initial cohort) and 57 lung cancer deaths.

Exposure measurements were reconstructed from 261 Rn gas samples collected between 1945 and 1958. The distribution of annual Rn concentrations was found to vary, from being normal or log normal in some years and showing no clear distribution in other years. Measures of annual average concentration were obtained by using estimators appropriate for each type of distribution; viz., arithmetic means for normal distributions, geometric means for log normal distributions and medians in instances where no clear distribution could be discerned. Equilibrium factors were estimated based on ventilation rates and weighted according to the labor distribution, type of workplace and season (summer and winter). Individual exposures were then calculated in WLMs based on a 48-hour week and 48 weeks per year. Due to the unavailability of employment records for workers at Port Radium prior to 1940, estimates of exposure experienced during this period were not available for those who worked there before as well as after 1940. Therefore, cumulative exposure values reported for some workers at least, are deemed to be underestimated.

Findings from this study revealed a highly significant linear relationship between exposure and increased risk of lung cancer. The weighted linear regression estimates of relative and absolute risk coefficients of 0.27% per WLM and 3.10 per WLM per 10^6 PYR obtained in this study were thought to represent an upper limit that reflected the nature of the overall biases that could have been present in this study. The Port Radium cohort was exposed to much higher concentrations of Rn progeny than the Beaverlodge miners, and it was postulated that the much lower risk estimates obtained in this study could have been due to effect modification by high exposure rates.

IV. Ontario Studies

Mortality amongst Ontario gold and uranium miners has been studied from 1955 to 1980. Uranium mining in Ontario commenced in 1955, peaked in the late 1950's and early 1960's and declined thereafter. A pilot mortality study of Ontario uranium miners was reported in 1974 (Muller and Wheeler 1974), with several studies being reported subsequently (Ham 1976; Chovil 1981; Muller *et al.* 1981, 1983 and 1985; Kusiak *et al.* 1991).

The 1976 report by Ham (Ham 1976) on behalf of the Canadian Royal Commission on the Health and Safety of Workers in Mines, studied lung cancer mortality for 15,094 workers who had experienced at least one month of exposure in a uranium mine. The study reported 81 lung cancer deaths identified from matching against national vital statistics records; this rate of lung cancer mortality was significantly higher than the rate expected based on national vital statistics.

Results of a case-control analysis within this cohort revealed that cases had worked for a significantly longer duration in mining (average duration of employment: 43.2 months for cases, 25.6 months for controls) and had been exposed to significantly higher levels of exposure than controls (average exposure: 74.5 WLM for cases, 32.8 WLM for controls). The effects of smoking amongst subjects was not examined in this study (Hewitt 1976; IARC 1988). By examining records of the Workmen's Compensation Board of Ontario (Chovil 1981), a total of 135 lung cancer cases were identified. This report which included information on smoking histories on 64 lung cancer cases (all of whom were smokers), was however, not considered a formally designed study; results of crude analyses confirmed the excess of lung cancer amongst Ontario uranium miners and its association with estimated exposure to Rn progeny.

Subsequent studies reported by Muller *et al.*, (Muller *et al.* 1981, 1983 and 1985), were based on a cohort of 15,984 Ontario uranium miners; this cohort excluded those who may have been exposed to asbestos and those who had mined uranium in another province as an employee of Eldorado uranium mines. The cohort was selected from miners who attended a miners' chest clinic between 1955 and 1977, and had been employed for at least one month as an underground uranium miner. Vital status and cause of death ascertainment were made through linkage with the National Mortality Database of Statistics Canada for deaths between 1956 and 1986. Individual exposures were estimated using work histories assembled by the Ontario Workmen's Compensation Board and mine exposure levels estimated using various approaches for the period prior to 1968, and from personal records of exposure to Rn progeny which were available after 1967. This cohort included miners who had also worked in gold mines.

Follow-up till 1981 resulted in an excess of lung cancer mortality being identified amongst both groups, uranium miners who had and who had not worked in gold mines. After excluding prior gold mining experience, the excess relative risk per WLM for Ontario uranium miners was estimated to lie between 0.5% and 1.3%; a range of values was provided for risk estimates to account for the uncertainty of cumulative exposure estimates which averaged between 40-90 WLM per person. These findings were based on a total of 232,795 person-years-at-risk and 82 lung cancer deaths; a total of 56.9 lung cancer deaths were expected compared to the Canadian population (SMR= 1.44).

Factors modifying lung cancer risk in Ontario uranium miners were further examined by Muller *et al.* (Muller *et al.* 1989). In this report the authors examined the effect of various lag periods using relative risk models modified by time since exposure and concluded that the "*most effective exposure occurred 10 to 14 years before the time of observation*" (Muller *et al.* 1989). They also concluded using a relative risk model, that the interaction between the effects of inhaled cigarette smoke and radon progeny was multiplicative.

A more recent report on a cohort of 54,128 Ontario miners reported follow-up till the end of 1986 (Kusiak *et al.* 1991); the cohort mainly comprised gold, nickel and uranium miners. This report concentrated on lung cancer mortality amongst gold miners and excluded miners who died after beginning to work in uranium mines; of the 37,493 men who entered the study as non-uranium miners, 6,260 including 2869 gold miners went on to become uranium miners. Other criteria for inclusion into the study were similar to those reported previously.

Findings from this study showed an overall excess of lung cancer mortality amongst 13,603 Ontario gold miners (SMR=129; 95% CI: 115-145). Closer examination showed that the excess lung cancer mortality was mainly amongst those who began mining gold before 1946 (SMR=141; 95% CI: 105-184); no increase in lung cancer mortality was detected amongst those who began mining gold after the end of 1945, those who began mining nickel after 1936 or those who mined ores other than gold, nickel and uranium. The excess risk of lung cancer mortality amongst gold miners was mainly attributed to exposures to high dust concentrations and to arsenic prior to 1946 and exposure to Rn progeny. Each year of employment in gold mines prior to 1946 was associated with a 6.5% increase in lung cancer mortality after a latency period of 20 years or more from commencement of employment (95% CI: 1.6%-11.4%). Examination of the joint effect of exposures to arsenic and Rn progeny showed that they acted independently upon lung cancer risk resulting in an additive effect on the risk of lung cancer mortality; after a 20 year period from commencement of employment, the risk of lung cancer mortality associated with arsenic exposure increased by 3.1% (95% CI: 1.1%5.1%) for each year of employment; the risk of lung cancer mortality associated with Rn progeny exposure increased by 1.2% per WLM (95% CI: 0.02%-2.4%) after a period of five or more years following first exposure. Examination of smoking data on a random sample of the miners studies showed that smoking was not a likely explanation for the excess risk of lung cancer mortality observed among Ontario gold miners. Findings from this study which are pertinent to this thesis are summarized in section 2.4.8 (table 2.4.8) of this work; since mean Rn progeny exposure estimates were not reported in the most recent paper (Kusiak *et al.* 1991), table 2.4.8 presents an estimate previously (Muller *et al.* 1985) for this cohort.

V. Newfoundland Fluospar Miner Study

Fluospar was mined underground in St. Lawrence, Newfoundland from 1936 to 1978. A growing number of lung cancer deaths was noticed amongst mine workers in the early 1950s, but exposure monitoring did not commence till 1960. Subsequent investigations revealed high concentrations of Rn and Rn progeny in the mine air (in excess of 190 WL), arising from the water that seeped into the mines. Introduction of mechanical ventilation into the mines in 1960 reduced the levels of Rn progeny to well below the permissible standards of the time.

A cohort study of 1,772 underground miners and 352 surface workers from the Newfoundland fluorspar mines was reported by Morrison *et al.*, (Morrison *et al.* 1988). Occupational histories for cohort members were compiled from company records and from personal interviews. Radon progeny exposure levels were estimated from measurements of Rn and Rn progeny made since 1960, and daily exposures which were recorded for each worker since 1969. The average exposure for underground workers was 382.8 WLM, over an average of 5.7 years of exposure.

Information on smoking histories was obtained through surveys conducted in 1960, 1966, 1970 and 1980, for 48% of the underground workers. Vital status and cause of death of cohort members were ascertained through record linkage with the National Mortality Database. Persons untraced through this search were assumed to be alive for the purposes of analysis.

Follow-up to the end of 1984 resulted in a total of 113 lung cancers being observed. Since only 6 lung cancer deaths were observed among surface workers, this group could not effectively be used as a basis of reference. Attributable and relative risks were computed based on stratified analyses using indirect standardization methods.

Results showed that SMRs increased with increasing cumulative dose, the trend being highly significant. Statistically significant SMRs were reported in all but those unexposed (surface workers) and those exposed to less than 100 WLMs of Rn progeny. A simple linear model was found to fit the data adequately and no significant quadratic departures from linearity could be detected. A decline in SMRs was noted with increasing time since last exposure. Relative risk estimates declined with increasing age at observation. Though age at first exposure did not significantly improve the fit of models, the relative risk coefficients were seen to decline with increasing age at first exposure.

Compilation of smoking data revealed that only 13% of underground workers had never smoked and 70% were current smokers at the time of interview. Based on available data, the prevalence of smoking was found to be constant across exposure categories. An element of bias arising from the increased prevalence of smoking amongst underground miners as compared with the reference population (71%), was thought to possibly inflate the SMRs for smoking related cancers. Another source of possible bias was identified as the greater likelihood of lung cancer being identified as the cause of death, by the medical community of St. Lawrence who were conscious of the risk of lung cancer among fluorspar miners.

2.4.2 United States Studies

I. Colorado Plateau Uranium Miners

The uranium mining industry in the Colorado plateau region comprised approximately 2500 mines located in Colorado, Utah, New Mexico and Arizona. Investigation of the miners and millers employed in this region began in 1950 when a US Public Health Service initiated a prospective cohort study. Selection criteria for participation in this study included only workers who had undergone at least one medical examination conduced by the US Public Health Service between 1950 and 1960, and had completed at least one month of underground mining by 1 January 1964. Hence, the study cohort did not include any workers who were not exposed to Rn progeny. Two separate study cohorts - distinguished on the basis of race - were initially established, comprising approximately 3400 white miners and 740 non-white - primarily American Indian - miners (IARC 1988).

Estimates of Rn progeny concentrations were made by the US Public Health Service, state agencies and the mining companies based on nearly 43,000 measurements taken between 1951 and 1968. The sources and number of measurements varied by geographic region and calendar year. Estimates of Rn progeny concentrations were not measured directly and annually in all mines; therefore, gaps in data were filled by estimates. For subjects who had previously worked in hard rock mines, previous exposures were estimated from annual Rn concentration estimate and work histories obtained from annual census of active miners and self-completed questionnaires. Exposures experienced after 1960 were estimated from Rn concentration measurements, made primarily for control purposes which could therefore have overestimated individual exposures (Lundin *et al.* 1971). Measurement errors associated with these components constituting individual exposure estimates were studied by Hornung and Meinhardt (Hornung and Meinhardt 1987).

Cigarette smoking histories were compiled from information collected at the survey examinations, annual censuses of miners and mailed questionnaires between 1950 and 1969 (Lundin *et al.* 1971; Whittemore and McMillan 1983). Cohort mortality was determined periodically and cause of death was ascertained from death certificates. Investigation on mortality is still in progress.

The Colorado cohort has been analysed periodically at various follow-up intervals since the inception of the study, with the size of the analytical cohort varying with specific selection criteria and follow-up. Each of these analyses showed an excess of lung cancer mortality amongst white males, with relative risk of lung cancer mortality increasing with cumulative exposure; this effect was not altered by controlling for the effect of cigarette smoking through stratification (Wagoner *et al.* 1965; Lundin *et al.* 1969; Archer *et al.* 1973 1976; Hornung and Samuels 1981; Waxweiler *et al.* 1981; Whittemore and McMillan 1983; Hornung and Meinhardt 1987).

The most recent of these analyses reported in 1987 (Hornung and Meinhardt 1987) was based on a study cohort of 3346 white miners including 256 lung cancer deaths. Results showed that the dose-response relationship was best described by a power function in preference to the linear or log-linear functions. Several lag periods were examined and the optimal fit was obtained using a four year lag period followed by linear partial weighting of lag periods between four and ten years. Background levels of exposure from Rn progeny and passive smoking - obtained from estimates of background exposure in the United States of 0.2 WLM per year since birth for Rn progeny exposure and 0.005 pack of cigarettes per day for passive smoking - were introduced for each individual.

The combined effect of smoking and Rn progeny exposure provided the best fitting model for this data, with the combined effect being less than multiplication and greater than additive. Increasing the background Rn progeny estimate to 0.4 WLM was found to provide a better fitting model.

Examination of time-related factors and other modifiers of effect showed that all else being equal, for a given level of cumulative exposure, lower rates of exposure (i.e., exposures sustained over longer periods) had a larger effect than higher rates of exposure; a birth cohort effect was detected with later births having a higher risk than earlier births for a given set of other risk factors; examination of the effect of age at first exposure showed that those first exposed at older ages were at significantly higher risk than those commencing their exposures at younger ages and the risk of lung cancer mortality was found to decline dramatically with increasing time since last exposure, the decline being statistically significant.

The authors discussed the relevance of their findings in the context of the multistage theory of carcinogenesis. They concluded that their findings on temporal effects supported the theory that exposure to Rn progeny apparently acted as a late stage carcinogen primarily contributing as a promoter under the multistage theory of carcinogenesis.

Quantitative findings relevant to this thesis are summarized in section 2.4.8 (table 2.4.8).

II. New Mexico Uranium Miner Studies

A case-control study of Navajo men (a predominantly non-smoking population) conducted to examine the association between uranium mining and lung cancer studied 32 cases ascertained from the New Mexico Tumor Registry between 1969 and 1981, each being matched with two population based controls (Samet *et al.* 1984). The prevalence of uranium mining experience was 72% amongst cases, whilst none of the controls had documented uranium mining experience. Smoking data was available only for 21 cases and were supplemented by estimates derived from the British doctors' study (Doll and Peto 1978). The authors concluded that there was a strong association between uranium mining and lung cancer which could not be sufficiently attributed to selection or information bias, and that uranium mining without cigarette smoking increased the risk of lung cancer (Samet *et al.* 1984).

A cohort study was initiated in 1988 by the University of New Mexico to examine the mortality and morbidity among uranium miners in the Grants mineral belt region of New Mexico (Pathak *et al.* 1988). The study comprised 4,048 miners who had undergone at least one mining company physical examination between 1957 and 1976, at the Grants clinic and had at least one year of documented underground experience. Estimates of Rn progeny exposure were obtained from several sources including, company records on individual exposure available from 1968, mine exposure measurements made by the State Health and Environment Department prior to 1968, and exposures from work outside New Mexico as estimated by the Public Health Service for the Colorado Plateau cohort study. Vital status and cause of death ascertainment was made through the New Mexico Tumor Registry, physicians and company records, and records of the Colorado Plateau study. Follow-up to the end of 1985 revealed 66 cases of lung cancer, eligible for study; initial analyses showed that the risk of lung cancer increased with increasing exposure, duration of exposure and time since first exposure. Results of more recent analyses (Samet *et al.* 1989 and 1991; Samet 1992), based on the follow-up of 3,469 miners (87.3% of whom had complete data on cigarette smoking) are summarized in section 2.4.8 (table 2.4.8) of this thesis.

2.4.3 Czechoslovakian Studies of Joachimsthal Miners

Metal ore has been mined on the German (Schneeberg) and Czechoslovakian (Joachimsthal) sides of the Erz mountains since the 15th and 16th centuries. As early as the 1546, a high prevalence of fatal lung diseases was reported among these miners (Agricola 1597; Lundin *et al.* 1971). The disease was identified as lung cancer in 1879 (Harting and Hesse 1879). Detailed accounts of this pioneering work on lung disease in the Erz mountain regions have been presented in two recent papers (Greenberg and Selikoff 1993; Schuttmann 1993).

Mining of uranium ore commenced more recently in the Joachimsthal region; these miners comprised one of the earliest epidemiological studies on Rn exposed miners. Czechoslovakian studies of underground miners comprise of several groups studied independently of each other and recently, in combination with each other. The group most studied comprised 2,433 miners (Sevc *et al.* 1971) who had worked for an average duration of approximately 26 years (Kunz *et al.* 1979) and been exposed to an average of approximately 300 WLM. Recent reports include two additional study groups of 2,194 and 1,849 miners who commenced exposure in underground uranium mines between 1948 and 1957 and were exposed to averages of 303 and 134 WLM, respectively over 10 years; another two groups comprising 3,799 and 1,561 uranium miners were exposed to much lower levels (mean exposure: 6.1 and 3.2 WLM, experienced over 6 and 10 years respectively), between 1968 and 1975; and two other groups comprising of 1,056 East Slovak iron miners commencing exposure between 1951-1960 (mean exposure: 40 WLM over 18 years), and 916 clay shale miners who commenced exposure after 1950 and were exposed to an average of 25 WLM over 11 years. The essential characteristics of these study groups were summarized by Sevc *et al.*, (Sevc *et al.* 1988), who refer to the first group as study S - comprising groups A and B (uranium miners commencing exposure between 1948 and 1957), the second as study N (uranium miners commencing exposure after 1968) - comprising groups B and C, and the last two as study K (East Slovak iron miners) and study L (clay shale miners).

Measurements of Rn progeny levels were made from around 1960 onwards, in selected workplaces; full measurements of Rn and Rn progeny levels started in 1966. Radon progeny levels prior to 1960 were estimated from Rn gal levels which had been noted from 1946 onwards, using equilibrium factors calculated according to mining and ventilation data. Ventilation in the mines was mainly natural ventilation prior to 1952; mechanical ventilation was introduced in mines from 1953 and all mines were mechanically ventilated by 1955. After 1955, secondary auxiliary ventilation was also introduced into the mines. Individual exposures were computed using additional data on job description and hours worked obtained from pay roll data which were available since 1948. This description of the measurement of Rn gas and Rn progeny exposure in the Czechoslovakian mines was obtained from the SENES report (SENES 1990) which rates the Chezch exposure data highly, due to the extent of detail on which mine exposure levels and time spent in workplace by individuals were monitored.

The basic procedures used in all these cohort studies were uniform. Vital status and cause of death ascertainment were made through the population registry at the Ministry of Interior, examination of district death registers, registries of oncological notifications in the Ministry of Health and pathologico-anatomical records maintained in district hospitals. Follow-up was completed till the end of 1980 for studies S and N, and through to the end of 1981 for study groups K and L.

Data on cigarette smoking were obtained in studies N and L by detailed questionnaires being administered to all members of the study groups. Smoking data for study S were obtained through repeated surveys conducted in parts of the study only. Lung cancer deaths identified amongst these study cohorts comprised 484 from study S, 4 from study N, 9 from study K and 22 from study L. Analyses were based on exposures computed using five year lagged data and on non-lagged exposure data, findings of which emphasized the importance of using lagged exposures.

Results of analyses by Sevc *et al.*, (Sevc *et al.* 1988) reported that the attributable lung cancer risk per WLM of exposure increased with age at first exposure in all cohorts, and with age at death in the first three cohorts. The relative risk of lung cancer was reported as decreasing with age at first exposure and with age at death in all cohorts. Analyses included examination of the character of exposure accumulation, i.e., the time course of accumulation of exposure. The paper also contained extensive discussions of the findings in the context of "*inhibitory*" effects of radiation - the decrease in the lung cancer rate in higher accumulated exposure categories of the cohort. Overall, a statistically significant excess of lung cancer mortality was reported even at radiation exposure levels below 50 WLM, and the attributable annual cancer risk per WLM at low levels of exposure was higher for low exposures compared with the risk at higher accumulated exposures. With regards to cigarette smoking, in a previous analyses, Sevc *et al.*, (Sevc *et al.* 1976) reported a two-fold increase in the lung cancer rate amongst miners who smoked compared to those who were non-smokers. Results published in the 1988 paper (Sevc *et al.* 1988) showed that the effects of smoking and exposure to α radiation from Rn progeny were nearly additive.

More recent findings from the Czechoslovakian (Sevc *et al.* 1993) relevant to this thesis are summarized in table 2.4.8. Analyses included in this report (Sevc *et al.* 1993) included results of non-parametric evaluation of exposure-response relationships using isotonic regression and parametric modelling including fitting the BEIR IV model (BEIR IV 1988). Results showed non linearity in the dependence of the excess risk of lung cancer on the cumulative exposure; the slope being steeper for exposures below 100 WLM. However, the authors suggested that this non linearity may have been an artifact of unknown confounders and advocate that the commonly postulated linear exposure-response relationship should not be abandoned as a result of these findings. Another major finding reported was the *protracted exposure effect* similar to that reported in the Colorado Plateau study (Hornung and Meinhardt 1987); the authors concluded that a more thorough analysis was required to determine whether this was an inherent *protracted effect* or an artifact of an unrecognized confounder.

2.4.4 Swedish Studies of Malmberget Iron Miners

A study of 1415 iron miners exposed to low doses of Rn progeny in Malmberget, Sweden, was reported in 1984 (Radford and Renard 1984). The study included only miners born between 1880 and 1919, who were alive in 1930, and had worked underground in more than one calendar year between 1897 and 1976 who were considered eligible for selection into the study. These individuals were identified from mining company and union records of active and pensioned miners, available since 1900.

Vital status ascertainment in this study appears to have been considerably easier than in other studies reported, due to the Swedish systems of birth codes and parish registers. All births in Sweden are registered with a *birth code*, which incorporates the individual's birth date; the birth code appears to be used as an individual identifier and is included on all work and hospital records. Under Swedish law, it is required that all Swedish citizens must register in the local parish of the state church which includes information on each individual's occupation, transfers to other parishes and particulars of vital status including, in the case of those desceased, the principal cause of death as appearing on the death certificate. Vital status ascertainment of study cohort members was made primarily on the basis of information obtained from these sources.

Two major sources of Rn progeny exposure were identified viz., Rn dissolved in water seeping from underground springs and Rn progeny in air. Radon progeny exposure data were reconstructed based on measurements of radioactivity within the mine and in the springs around Malmberget made since 1915, and numerous measurements of Rn and Rn progeny in mine air made in 1968 and thereafter; reconstruction of Rn progeny exposure also took into consideration ventilation levels and average number of man-hours per month spent, in the underground workplace in each underground section.

Calendar years of employment for study participants were obtained from mining company and union records, and medical files. Approximate durations of employment were then computed using broad assumptions on dates of commencement and termination of employment in each calendar year, for each individual (Radford and Renard 1984). Individual exposures were based on this data, data on the number of individual man-hours spent underground in each calendar year obtained from mining company records, and estimates of Rn progeny exposure. For analytical purposes, individual exposures were computed based on five year lagged exposures.

Individual smoking histories were obtained through questionnaires administered to active miners and surface workers in 1972-1973, and to pensioners in the study population in 1977. Smoking histories were also obtained for all lung cancer deaths from the subject himself, before his death, or from relatives or co-workers, after his death. No mention was made of smoking histories being sought for cohort members who died of causes other than lung cancer.

Smoking histories were obtained for 556 miners - 388 from the 1972-1973 survey of active miners and 168 from the 1977 survey of retired miners and relatives or co-workers of deceased miners. Smokers were identified as current smokers and ex-smokers who had or had not quit at least 10 years previously. For analytical purposes, ex-smokers who had quit at least 10 years previously were combined with those who never smoked and together, classified as "*non-smokers*".

Of the 1415 cohort members originally included in the nominal roll of the study, 121 died between 1930-1950; 1294 were alive in 1951, of whom 7 subjects were lost to follow-up (6 due to emigration), leaving 1287 who were followed-up further to the end of 1976, at which time 532 were dead and 755 alive. Analyses were based only on follow-up during the period 1951-1976.

The study reported a total follow-up duration of 24,083 person years accumulated by the 1294 cohort members who were alive at the beginning of 1951. The average exposure estimate was 81.4 WLM per person year, averaging 93.7 WLM total cumulative exposure, both computed using a five year lagged exposure. Fifty lung cancer deaths that occurred more than 10 years after working at the mine were observed during 1951-1976. One further lung cancer death which occurred 9 years after commencing work at the mine was excluded from subsequent analysis.

The study also examined the contributory effects of other factors such as, other related illnesses: inactive tuberculosis and subsequent development of bronchogenic carcinoma and silicosis, which were both ruled out as factors contributing to lung cancer; exposure to diesel exhaust - ruled out because diesel equipment was not used in the mines until 1960, by which time 70% of the lung-cancer cases had left underground work or died and others spent only brief periods exposed before their death; exposure to other respiratory carcinogens e.g., arsenic, chromium and nickel which were essentially absent in the mine and therefore, ruled out; dust samples obtained from the mine showed occasional traces of serpentinite, but this factor could not be implicated as being associated with the lung cancer cases observed because x-ray diffraction studies showed no identifiable asbestos fibers.

2.4.5 Chinese Studies of Yunnan Province Tin Miners

Tin miners of Yunnan province in China comprise the largest occupational cohort of Rn exposed workers reported to date. A historical cohort study conducted among workers of the Yunnan Tin Corporation (YTC) included 17,143 male workers followed-up from 1976 to1987, during which period 981 lung cancer cases were reported (Xuan *et al.* 1993). The study cohort comprised all workers who worked at one of the five major mining units of the YTC and participated in an occupational survey conducted in 1976; the 1976 survey covered almost 20,000 of approximately 44,000 current and retired YTC employees. The YTC provides retirement and medical benefits to retired workers and identifies lung cancer as an occupational disease for which it provides compensation to cases and their families. The YTC therefore maintains medical and payroll records for all current and retired employees until their death; furthermore, the YTC also intitated a lung cancer registry in 1973 and a registry of all cancers in 1978. These records provided the source for vital status ascertainment and cause of death information for the study cohort. Data on individual smoking habits was obtained through record linkage between YTC pay roll records and the 1976 occupational survey.

Radon progeny concentrations in the mines were estimated from over 26,000 Rn concentration measurements made since 1972 and through 413 measurements obtained from simulation exercises involving the recreation of working conditions for the period 1953-1972 in the YTC mines, and from similarly configured areas in nearby non-YTC mines which used mining techniques similar to those used in YTC mines during the index years. Estimates of Rn concentration levels prior to 1953 were based on 117 measurements obtained by other simulation exercises in 13 local pits that operated before 1949 (Zhang *et al.* 1981). Individual exposures to Rn progeny were estimated from these measurements and individual work histories based on the assumption of a 7 hour working day and 285 working days per year (166 working hours per month); i.e., a worker exposed to 1 WL for one month would accumulate 166/170 = 0.98 WLM. Since YTC workers were also exposed to arsenic - an independent cause of lung cancer - individual exposures to Arsenic were also estimated.

Poisson regression techniques were used to evaluate the risk of lung cancer mortality using person-years-at-risk and event data classified according to Rn and arsenic exposure, tobacco use and categories of other confounders and modifiers of effect. Risk evaluation was based on relative risk and absolute risk models. A total of 2,591 deaths occurred in the study cohort during the follow-up period, which included 981 lung cancer deaths. Further details of results are summarized in section 2.4.8 (table 2.4.8). Risk related to Rn progeny exposure risk was made with and without controlling for arsenic exposure. Due to the quality of the data on smoking, smoking could only be used as a categorical variable (dichotomous variable - smoker and non-smoker - for analytical purposes); smoking data was not available for 4,088 subjects, including 74 lung cancer cases; the prevalence of non-smokers was 7%. Results showed a 2 to 3 fold excess of lung cancer risk amongst smokers compared to non-smokers. The interaction between smoking and Rn exposure was evaluated as being between multiplicative and additive, with the multiplicative model providing a better fit than the additive model.

2.4.6 French Studies of Uranium Miners

A retrospective cohort study of underground uranium miners who worked in one of 12 French mines between 1947 - 1972, was reported by Tirmarche *et al.* in 1984 (Tirmarche *et al.* 1984). The study comprised 2,442 underground miners who were followed through to 1983. Ascertainment of individual vital status and cause of death cannot be made through the National Registry of Death Causes in France, because information maintained is anonymous and does not allow individual identification. Vital status and cause of death ascertainment in this cohort study was therefore performed through alternative means - a collaborative effort with the Institute Gustave Roussy, a cancer institute with access to national statistics which enabled periodic extraction of data on miners' life status (dead or alive). Estimates of Rn progeny levels from 1947 - 1955 were made by an expert committee on the basis of a few Rn measurements and details of ventilation conditions, ore characteristics and work practices. Estimates of exposures experienced thereafter, were based on extensive measurements of individual miners. A total of 36 lung cancer deaths were observed, revealing an elevated rate of lung cancer mortality amongst this cohort. Smoking histories were not available in this study, and the report did not address exposure-response relationships (IARC 1988). Verification of data on exposure estimates and further analyses were reported to be continuing.

Further findings from the most recent report on this study (Tirmarche *et al.* 1993) are summarized in section 2.4.8 (table 2.4.8).

2.4.7 United Kingdom Studies of Tin Miners

A study of mortality based on the 1939 population register in the United Kingdom, showed clear evidence of excess lung cancer mortality amongst male Cornish tin miners (Fox *et al.* 1981). The scope of this 1939 population register study was extended in 1983 by an occupational cohort study of miners from two remaining operational tin mines, identified as mines A and B (Hodgson and Jones 1990). Study participants were all men who had worked in either one of these tin mines for at least one year between 1941 and 1984, and for whom complete work histories could be reconstructed from mine records. Criteria for selection into the study further excluded all those who were aged over 60 at the commencement of mining work, those with birth dates earlier than 1880, those for whom necessary personal particulars were not available and those who had any exposure to arsenic. The cohort thus determined comprised 3,082 workers who were studied from 1941 to 1986. Vital status determination was made through the central register of the National Health Service with a race rate of 97.6% (72 workers could not be traced). For those deceased during this time, cause of death identification was made through the Office of Population Census and Surveys. No data on smoking was available for this cohort.

Exposures to Rn progeny were assessed by the National Radiological Protection Board (Strong *et al.* 1975) based on measurements of Rn concentrations made from 1967. Unlike in most other mines studied, exposures experienced in mine A were identified as having been constant over the period of time study; consisting of a complicated network of interconnected shafts and workings that had been developed over 200 years or more, Rn concentration in this mine could not be reduced by additional ventilation. In mine B, reduced Rn concentrations were registered from 1967. On these bases, annual exposures were estimated as being 10 WLM throughout the period of study in mine A, and 20 WLM upto 1970 and 15 WLM thereafter for mine B.

Using individual work histories periods of employment were classified into three categories viz., *underground*, *surface* and *intermediate*. Individual exposures were estimated based on duration of employment underground with broad weightings being used for prior underground mining experience and *intermediate* jobs. Only those who strictly had worked *underground* were included in the analyses; those having worked in *intermediate* jobs alone were excluded from analyses.

Results showed a highly significant excess in lung cancer deaths (57.7%) with a strong increase in lung cancer mortality with increasing exposure. Poisson regression modelling was used to examine the dose-response relationship and the roles of effect modifiers, which resulted in highly significant dose-response relationship, but failed to identify any significant modifiers of effect. Latency models provided marginal improvement in fit; though the effect of latency was statistically non-significant there was some suggestion of excess risk varying with time.

An approach different to those adopted in other studies was used to further characterize temporal variations in the dose-response relationship, which was modelled differently during three stages of time since exposure periods with the excess risk being zero during a *lag period*, linear increasing during the *rising time periods* and then exponentially decreasing during the *effect half life period.* Using maximum likelihood techniques, parameter estimates for these periods were estimated as: a lag period of 10.5 years (95% CI: 8.3-13.5 years), a rising time period of 0.8 years (95% CI: 0-7.8 years), and an effect half life period of 4.3 years (95% CI: 1.3-15.5 years). The dose function thus defined - which the authors refer to as *effective dose* - was a significantly better predictor of lung cancer mortality than cumulative dose, whilst embracing all the explanatory power of cumulative dose.

2.4.8 Summary of Overall Findings: Individual and Joint Analyses

Major findings of specific relevance to this thesis from the studies reviewed above are summarized in table 2.4.8. Table 2.4.8 shows that the estimates of ERR/WLM obtained from the various studies reported range from 0.16% to 9.4%. The lowest estimates of cumulative Rn progeny exposure were experienced by the Canadian Beaverlodge and Ontario cohorts; apart from these two cohorts and the Swedish Malmberget cohort who were all exposed to mean cumulative exposures of less than 100 WLM and the Port Radium cohort who experienced an average of less than 200 WLM, all other cohorts had mean cumulative exposures in excess of 200 WLM; the highest mean cumulative exposures - over 800 WLM - were registered amongst the Colorado miners. The largest cohorts comprised the Ontario and Chinese miners, the latter cohort having the shortest follow-up period with the largest number of lung cancer deaths (981) ever reported in a single cohort. The largest number of lung cancer deaths in other cohorts included 574 amongst the Czech miners and 378 amongst the Ontario miners. Standardized Mortality Ratios (SMR) for these cohorts showed that the Canadian studies recorded the highest and lowest excess lung cancer mortality compared to the relevant national populations - 29% (Ontario) and 425% (Newfoundland). In the absence of the corresponding statistic for the Chinese cohort, the Colorado and Czech cohorts had SMRs greater than 4.0. Findings from these individual studies also revealed the need for further studies of populations exposed to lower levels of Rn progeny, so as to enable the specific determination of risk at lower exposures.

Table 2.4.8 shows that despite some similarity, due to the basic differences in methods of study design and analyses the findings from these studies are not strictly comparable which makes direct comparisons difficult. However, general inferences may be still be drawn to help formulate some broad perspectives of the extent of risk related to Rn progeny exposures. In order to address these issues several efforts have been made in recent years to reanalyses of some of these cohorts in a strictly comparable sense, and at joint analyses including both meta analyses and combined analyses.

Table 2.4.8: Summary of Findings from Reported Studies

Study	Follow up Period	Cohort Size	Cases Ca. Lung	PYRS	Mean Exposure (WLM)	Doubling Dose (WLM)	SMR Ca. Lung Mortality	ERR/WLM Ca. Lung Mortality (%)	Comments
Canada:									
Beaverlodge (Howe *et al.* 1986; SENES 1991)	1948-80	8,487	54[e] 65	56,942[e] [114,170] (BEIR IV 1988)	20.2 [44][s]	31[c]	1.90 (1.43-2.49)	3.28 (2.08-4.48) [1.3][s] (0.59-2.98)[s]	D-R: Linear Lag:10 yrs. Smk: Not Rep. [s]SENES 1991.
Port Radium (Howe *et al.* 1987)	1942-80	2,103	48[e] 57	34,673[e]	183.3	370[c]	2.10 (1.55-2.79)	0.27 (0.11-1.43)	D-R: Linear Lag: 10 yrs. Smk: Not Rep.
Newfoundland (Morrison *et al.* 1988)	1933-84	2,124	110[e] 113	38,509[e]	382.4	112	5.25[u] (4.33-6.32)	0.9 (0.6-1.2)	D-R: Linear Lag[e]:10 yrs. Smk: Risk Rep.
Ontario (Kusiak *et al.* 1991)	1955-86	54,128	378	1,706,103	[40-90[r]] (Muller *et al.* 1985)	83[c]	1.29 (1.15-1.45)	1.2 (0.02-2.4)	Lag: 5 yrs. Smk: Assessed. Mean Exp. NR.
United States:									
Colorado (Hornung and Meinhardt 1987)	1950-82	3,346	256	[73,642] (BEIR IV 1988)	834.0	71-111[cr]	[4.33][c] (BEIR IV 1988)	0.9-1.4[r]	Lag: Estimated. Smk: Risk Rep.
New Mexico (Samet *et al.* 1991)	1957-85	3,469	65	NR	111	55.6[c]	4.0 (3.1-5.1)	1.8 (0.7-5.4)	Smk: Synergistic with Rn Exp.
Czech.: Study S (Sevc *et al.* 1993)	1948-80	4,042	574	97,913	227.0	95	4.7 NR	0.6 NR	Lag: 5 years. D-R: Supra-linear, linear
Sweden: (Radford and Renard 1984)	1951-76	1,294	50	24,083	93.7	28[c]	3.9 (3.0-4.9)	3.6 (2.5-4.8)	Lag: 5 years. CI: 90%. Smk: Risk Rep.
France: (Tirmarche *et al.* 1993)	1946-85	1,785	45	44,005	70	58.8[c]	1.91 NR	0.59 (0-1.6)	Interim Report; Analyses Cont. Smk: No Data.
UK: (Hodgson and Jones 1990)	1941-86	3,082 3,010[t]	105	NR	NR	NR	1.58 NR	9.4 NR	Lag: Not Used; Est. 10 yrs. Smk: No Data.
China: (Xuan *et al.* 1993)	1976-87	17,143	981	175,405	275.4	161[c] 625[ca]	NR NR	0.62 (0.5-0.8) 0.16[a] (0.1-0.2)[a]	Lag: 5 yrs. Smk: Risk Rep. [a] Adjusted for Arsenic Exp.

Note: NR Not Reported in main reference.
e Excluding first 10 years of follow-up.
t Traced Cohort: only these analysed.
r Range covers uncertainty in estimates.
D-R Dose-Response Relationship.
u Unlagged data.
c Computed from published findings.
Smk. Smoking; Rep.: Reported.

I. The BEIR IV Analyses

One of the most well known attempts at combined analyses was performed by the Committee on the Biological Effects of Ionizing Radiation (BEIR) in its 1988 report (BEIR IV 1988). Though the BEIR IV report addresses the health risks of Radon and other internally deposited α-emitters, the largest part of the report concentrated on the health outcomes due to exposure to Rn and its progeny. The BEIR IV committee reviewed four major cohort studies of radon-exposed miners: the Canadian Eldorado (Beaverlodge) and Ontario uranium miners, U.S. Colorado Plateau uranium miners and Swedish Malmberget metal miners, independently reanalysed each cohort using the same broad principles in an effort to resolve differences in study design and previously adopted analytical methods so as to make these studies comparable, and then performed combined analyses of the individual cohorts. All analyses performed by the BEIR IV committee were based on 5 year lagged individual cumulative exposures. The committee first fitted a constant relative risk model and then proposed a modified relative risk model - the BEIR IV model - which incorporated a dependence of the relative risk of lung cancer mortality on both time since exposure and age at risk. These models were fitted using Poisson regression techniques for categorical data based on the time dependent allocation of person-years-at-risk and events across predetermined categories of various factors relevant to the study of lung cancer mortality relating to Rn progeny exposure.

BEIR IV results of ERR/WLM from analysing individual cohorts using internal comparisons and a constant relative risk model were: 0.6% for Colorado Plateau (95% CI: 0.3%-1.3%), 1.4% for Ontario (95% CI: 0.6%-3.3%), 1.4% for Malmberget (95% CI: 0.3%-8.9%) and 2.6% for Eldorado (95% CI: 1.3%-6.0%). The risk estimate for the Swedish Malmberget miners was subject to greater variability than other estimates (multiplicative standard error of 2.6% compared with 1.5%-1.6%). The Colorado cohort used in these analyses excluded all those who were exposed to more than 2000 WLMs, due to the lack of linearity in the exposure-response relationship beyond this limit; this excluded 371 persons of whom 99 died of lung cancer.

Joint analyses showed that the ERR/WLM estimated from a constant relative risk model using internal comparisons for the data pooled from all four studies was 1.34% (95% CI: 0.8%-2.3%). It is interesting to note at this point, that the recent reanalyses of the Ontario and Beaverlodge studies produced estimates that correspond very closely with these BEIR IV estimates; estimates previously reported for the other cohorts are considerably higher than these BEIR IV estimates.

The BEIR IV analyses then proceed by examining the modifying effects of age at risk, age at first exposure, duration of exposure and time since exposure; each of these factors was examined in turn for each individual cohort by comparison with the corresponding constant relative risk model and the significance of each effect was tested using likelihood ratio tests for improvement in model fit. Conclusions drawn from these analyses were as follows:

- Analyses based on internal comparisons showed no definite dependence of the relative risk on age at risk, whereas externally adjusted analyses showed some indication of a decline in excess relative risk with increasing age.

- Findings on the effect of age at first exposure were not consistent between the cohorts; in the internal analyses; though, age at first exposure had a significant effect on the excess relative risk for the Eldorado and Colorado cohorts with the risk increasing with increasing age at first exposure, these findings were not supported by the external analyses. However, in view of findings from studies on Japanese atomic-bomb survivors amongst whom there appeared to be no age at first exposure effect before the age of 20 (Preston *et al.* 1986), the committee stressed the importance of the age at first exposure effect, stating that "*it is widely held that such an effect should be expected more generally (NIH 1985)*" (BEIR IV 1988). The BEIR IV committee further noted that the range of ages at first exposure observed "*may not have been great enough to give sufficient power to detect a real effect*" (BEIR IV 1988), and that age at first exposure was not a critical factor in relation to prolonged exposures such as those observed in the cohorts analysed.

- A protracted effect of duration of exposure was observed only in the Colorado data; no consistent patterns in excess risk estimates were observed with relation to duration of exposure in any of the other cohorts analysed. The BEIR IV committee noted that since the analyses adjusted for the effect of cumulative exposure in assessing the effect of duration of exposure effect, a duration of exposure effect would corresponded to an exposure rate effect and that it would be important to examine the protracted effect of duration in future work.

- A consistent decline in excess relative risk was observed with increasing time since exposure in all other than the Colorado cohort, with the effect being statistically significant in the Eldorado and Ontario cohorts.

The BEIR IV committee concluded that: *"there is substantial evidence that the relative risk of lung cancer mortality depended on age at risk and/or time since exposure, and there is little clear evidence that it depends on the other factors considered"*. It was on these grounds that the modified relative risk model was proposed by the BEIR IV committee.

The excess relative risk function of the BEIR IV modified relative risk model comprised two components - one of age at risk and the other expressing cumulative exposure as a function of time since exposure. The initial model fitted to each cohort was of the general form:

$$r(a) = r_0(a) [\, 1 + \gamma(a)\, \beta\, (W_1 + \theta_2 W_2 + \theta_3 W_3)] \qquad (2.1)$$

where, a - age at risk,
$r(a)$ - age-specific lung-cancer mortality rate,
$r_0(a)$ - age-specific background lung-cancer mortality,
$\gamma(a)$ - a function of age at risk categorized as:
$a < 55$ years, $55 \leq a < 65$ years, and $a \geq 65$ years
W_1 - lagged cumulative exposure incurred 5-10 years prior to age a,
W_2 - lagged cumulative exposure incurred 10-15 years prior to age a,
W_3 - lagged cumulative exposure incurred $\geq$15 years prior to age a.

The BEIR IV committee described the term within parentheses in equation 2.1 as an *"effective cumulative dose at age a"*. When this model was fitted to each individual dataset analysed, it was found that there was very little difference between parameter estimates q1 and q2; the committee therefore decided to combine these two windows of time since exposure into a single window spanning 5-15 years prior to age a. This reduced form of the model was then fitted to the combined data using maximum likelihood methods. Maximum likelihood estimates were obtained for each of the parameters with the analyses being repeated using internally and externally referenced data. The *best* estimate for each parameter was then chosen from the range of values obtained from the internal and external analyses which were both in close conformity. The final form of the BEIR IV modified relative risk model (including its *best* parameter estimates) was defined as follows:

$$r(a) = r_0(a) \, [\, 1 + 0.025 \, \gamma(a) \, (W_1 + 0.5 \, W_2) \,] \qquad (2.2)$$

where, a - age at risk,

$r(a)$ - age-specific lung-cancer mortality rate,

$r_0(a)$ - age-specific background lung-cancer mortality,

$\gamma(a)$ = 1.2 when $a < 55$ years

= 1 when $55 \leq a < 65$ years

= 0.4 when $a \geq 65$ years

W_1 - lagged cumulative exposure incurred 5-15 years prior to age a,

W_2 - lagged cumulative exposure incurred ≥15 years prior to age a.

Model (2.2) was then fitted to each individual cohort (with parameter estimates constrained at the given values) to estimate the ERR/WLM. Estimates of ERR/WLM obtained from internal analyses using the BEIR IV model were: 0.9% for Colorado, 1.8% for Ontario, 3.6% for Malmberget and 5.1% for Eldorado. The estimated ERR/WLM for the combined data from all these cohorts obtained from internal analyses was 2.2%.

The rationale behind the approach of a combined analysis undertaken by the BEIR IV committee was that although each independent epidemiological study will have its own limitations, the pooling of data sets will permit a comprehensive assessment of the health risks associated with Rn progeny exposure and of other factors that influence the risk, such as age and time since exposure. The BEIR IV committee also observed that quite apart from the increased statistical power of a number of studies regarded in conjunction, such a combined exercise provides a greater degree of variation in environmental settings, individual susceptibility and dosimetry of individuals and cohorts to varying degrees of exposure.

In presenting its findings the BEIR IV committee identified several uncertainties that may have affected its risk estimates, including random and possibly systematic errors in the individual datasets, appropriateness of the statistical models used, sampling variation, wrong characterization of the interaction between Rn exposure and smoking and uncertainties arising from the use of different reference populations in risk projection.

The BEIR IV committee did not regard smoking as a confounding factor in their cohort analyses. However, it examined the combined effects of Rn progeny and smoking and concluded that their interaction was sub-multiplicative and supra-additive.

II. Findings from Other Recent Joint Analyses

A meta analysis presented in a recent ICRP draft (ICRP Draft 1993) summarized the excess relative risk coefficients from uranium miner studies of the Colorado, New Mexico, Ontario, Beaverlodge, Bohemian and French cohorts and the Malmberget iron miners and arrived at a weighted average (weighted by person-years-at-risk) estimate of ERR/WLM: 1.33% (95% CI: 0.8% - 2.1%); these estimates bore a remarkable closeness to the BEIR IV estimate of ERR/WLM derived from the constant relative risk model - 1.34% (95% CI: 0.8% - 2.3%) (BEIR IV 1988).

The 1990 (SENES 1990) update to the 1984 SENES (SENES 1984) report focused on risk projection models from epidemiological data while paying special attention to the derivation of Rn progeny exposure estimates. This report reviewed several risk projection models - the BEIR IV model (BEIR IV 1988), NCRP 78 absolute risk model (NCRP 78 1984), the Ontario model - which included an expression of cumulative exposure in the three time since exposure windows defined in model 2.1, but excluded the age at exposure function $\gamma(a)$ in model 2.1 - from which the BEIR IV model evolved (BEIR IV 1988), and the ICRP 50 model (ICRP 50 1987). The SENES report concluded from its appraisal that the BEIR IV model overestimated the life-time risk from Rn progeny exposure to individuals.

2.5 Occupational Radon Epidemiology: Current Needs and The Role of This Work

Recommendations made by the BEIR IV committee for further research included the following: several current underground-miner surveys could provide a more extensive data base with increased person-years of follow-up and help to refine lung-cancer risk coefficients; provide more information on the interaction between smoking and Rn exposure; and, with improved dosimetry, narrow the uncertainties in the application of lung-cancer risk data derived from miners to the estimation of risk in the general population. Specific emphasis was laid on the need for improved and in-depth multivariate statistical analyses of the available data. Further recommendations included the need for continued epidemiological study with multivariate analysis of the temporal expression of lung cancer in underground miners exposed to Rn progeny. It was also stressed that collecting and reporting smoking data on miners should be an essential part of future study designs.

The roles of various temporal modifying factors - age at risk, age at first exposure, exposure rate and duration of exposure and time since exposure - have to be further examined to enable more thorough characterization of their roles in the risk of lung cancer mortality related to Rn exposure. Continued research is also needed to quantify the extent to which the carcinogenic effectiveness of low-LET radiation may be reduced by fractionation or protraction of exposure (BEIR V 1990).

Due to the lack of data on the lifetime cancer experience of exposed populations, overall risk estimates have to be obtained by means of models which extrapolate over time; hence, risk estimates obtained from epidemiological studies cannot be considered precise (BEIR V 1990). Studies of Rn exposed populations therefore must continue with the aim of following up the lifetime experience of these populations.

The BEIR IV committee concluded that there was further need to examine populations exposed to low-levels of ionizing radiation, to comprehensively characterize the lung cancer risk associated with exposure to Rn and its short-lived progeny. This was stressed further in the BEIR V report which found that studies of populations chronically exposed to low-level radiation did not show consistent or conclusive evidence of increased risk (BEIR V 1990).

To date, high dose epidemiology has been the main source of knowledge on the effects of ionizing radiation. With high exposures being well controlled against, overwhelming concern rests over the effects of small or very small doses (Anthony 1988). The BEIR V committee recognized that its risk estimates become more uncertain when applied to very low doses and that departures from a linear model at low doses, could either increase or decrease the risk per unit dose. The consensus therefore is a wide recommendation that epidemiological investigations to measure the cancer risk related to low doses must continue with some "statistically respectable" (Anthony 1988) populations being studied. The BEIR V committee recommends that studies on low as well as large dose of high and low LET radiation must continue and that low-dose epidemiological studies may be able to supply information on the extent to which effects observed at high dose rates can be relied on to estimate the effects due to chronic exposures such as occur in occupational environments (BEIR V 1990).

Findings from the individual studies reported previously show that the lowest exposures studied so far were amongst populations exposed to average cumulative exposures in the order of above 40 WLM; thus, revealing the need to study populations exposed to lower levels of Rn progeny, so as to enable the specific determination of risk at lower exposures. The Radium Hill cohort from South Australia was exposed to much lower cumulative levels of exposure to Rn progeny than those reported in any other study. Apart from its limited power in examining the roles of smoking and other exposures, this cohort could provide a rich basis for the addressing and examining each of the issues identified above as *current needs in radon epidemiology*. As for *statistical respectability*, it is stressed that this is yet a relatively *young cohort* capable of becoming more *statistically respectable* with further follow-up.

CHAPTER 3

3. Study Design and Implementation

Page

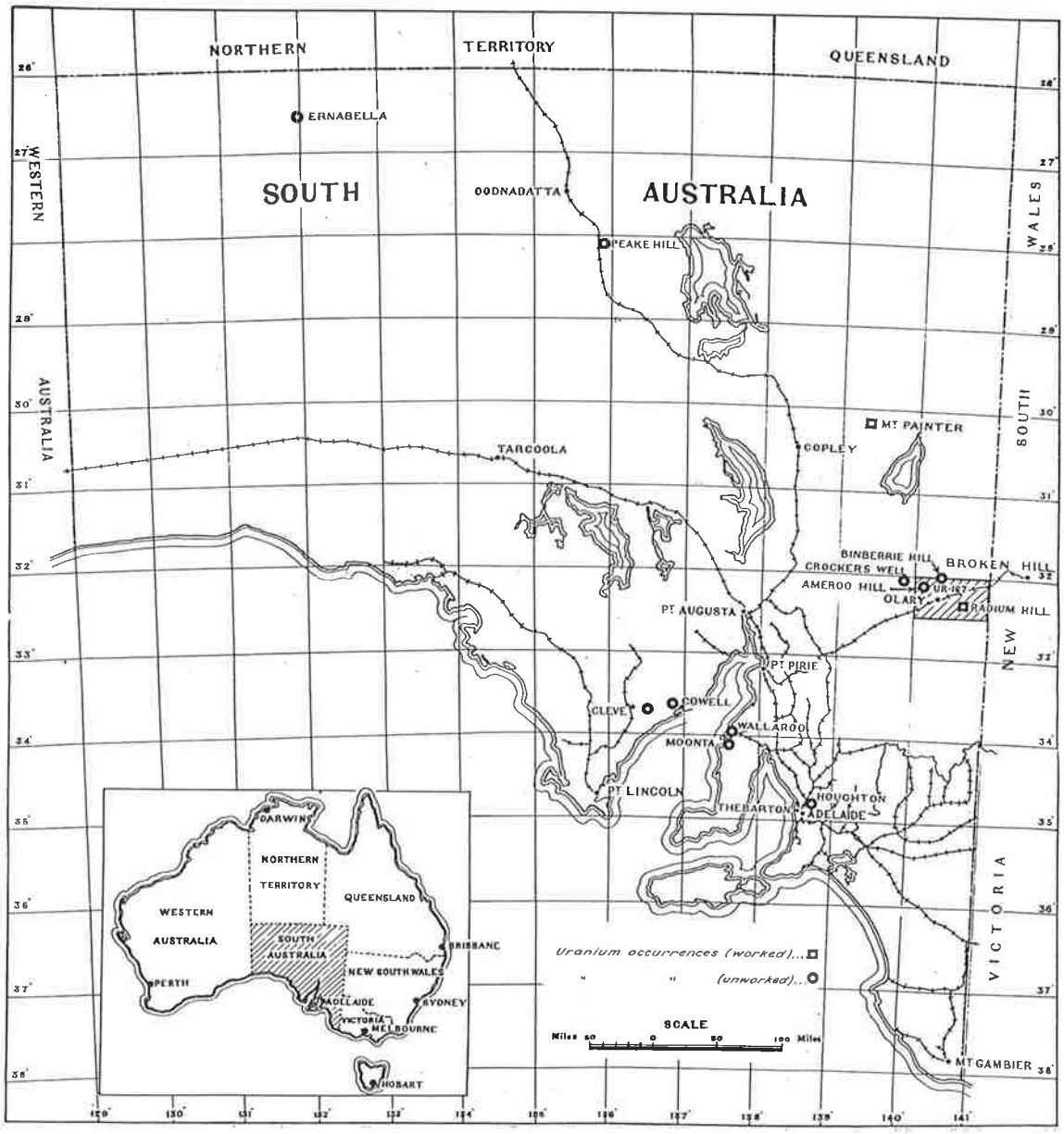
Tables and Figures

3. Study Design and Implementation

3.1 Historical Background to Uranium Mining at Radium Hill

Radium Hill (RH) is situated in a remote location of South Australia, approximately 460 kilometers north-east of Adelaide - the State Capital - and 100 kilometers south-west of Broken Hill - an established mining town on the border of the two states, South Australia and New South Wales (Figure 3.1a).

Figure 3.1a: Geographic Location of Radium Hill (SADM 1952)



Uranium was discovered at RH in 1906 by A. J. Smith, a prospector, who found a heavy dark rock - later named davidite (Crouch and Corani 1986). The uranium ore found in RH was an unusual iron-titanium rich mineral that was described in 1954, as having comprised a highly complex mixture of up to 9 per cent uranium oxide, and approximately, 50 per cent titanium oxide, 30 per cent iron oxide, up to 8 per cent rare earths and minor amounts of chromium and vanadium (SADM 1954).

The discovery of uranium at RH preceded the findings of uranium deposits in the Belgian Congo and Eldorado, Canada. The main producer of uranium in the world at the time was the Joachimsthal deposits in Czechoslovakia. At this time the main importance of uranium was as a source of radium, used chiefly in medicine; uranium therefore, had strictly limited demand.

The extraction of radium from the complex uranium ore found in RH was extremely difficult and expensive (SADM 1954). Early attempts were made at producing radium for the medical market from the RH ore; all these attempts were financially unsuccessful and therefore abandoned. Furthermore, in the light of competition posed by subsequent discoveries of uranium deposits in the Belgian Congo and Eldorado, in Canada, and the resulting drop in the price of radium, it was found that uranium mining at RH was not financially viable. Early uranium mining at RH terminated in 1930, after the production of a few hundred milligrams of radium and a few hundredweights of uranium oxide by-products.

Mining interests at RH were renewed during the mid 1940's, when uranium was identified as a source of nuclear power. Investigations into the feasibility of uranium mining at RH commenced in 1944 at the request of the British government. War-time investigations at RH comprised *dewatering* - draining of water -, the collection of bulk samples of ore from the old mine workings for treatment tests, and geological and geophysical surveys.

Pilot mining activities commenced in 1948. Early activities concentrated on securing the old mine shafts and providing other necessary preliminaries. Meanwhile, laboratory research was also undertaken by the South Australian Department of Mines and the Commonwealth Scientific and Industrial Research Organization, with a view to developing an efficient and economical process for uranium extraction from the complex and unusual ore found at RH.

By 1951, preliminary investigations at the RH mine site indicated that the magnitude and grade of the lodes were comparable to many of those being exploited in other countries. New techniques of uranium extraction had been perfected, and pilot plants were setup to check the concentration and ore-dressing process and the chemical extraction process. The feasibility of further uranium mining was confirmed by the end of 1951. Planning and preparation for ore production and uranium extraction commenced in early 1952 and the RH mine was formally opened on 10 November 1954, by which time the mine was fully operational.

The *'Radium Hill uranium mine'* was owned by the South Australian Government and operated by the South Australian Department of Mines from 1948 to 1962. The mine produced ore from 1954 to 1961 on a seven year contract with the United States of America and the United Kingdom. The ore produced at the mine was of a relatively low grade - an average concentration of 2.65 pounds of Uranium Oxide (U_3O_8) per ton - and was exported to the United States and the United Kingdom (SADM 1952). Uranium production activities were terminated at the end of 1961 with the expiration of the seven year production contract, during which time the concentration of ore had steadily deteriorated from 3.6 pounds U_3O_8 per ton to 2.3 pounds U_3O_8 per ton. As a consequence of the discovery of other richer deposits and the resulting drop in world market prices for uranium, the RH uranium mine could not compete efficiently on the world market and was closed. In 1962 the last of the employees left the mine site after the completion of stope filling and mine sealing activities.

Figure 3.1b shows Radium Hill as it is today. A compilation of photographs of the life and times at Radium Hill are included in appendix A.

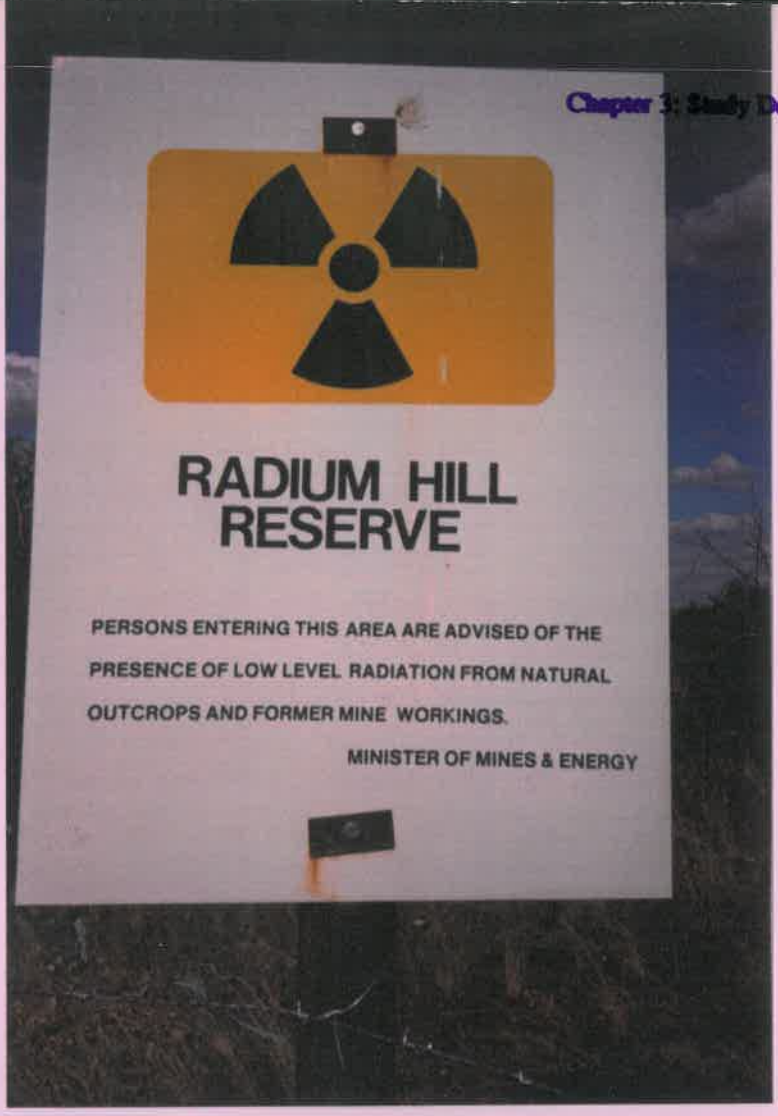


Figure 3.1b: Radium Hill - 1990

What Remains of a Once Busy Mining Town

3.2 The Radium Hill Study

3.2.1 Introduction

Uranium was mined at RH during two different periods this century; first, in the early nineteen hundreds and more recently during the fifties and sixties. My research is based on a study of the latter period. During the early 1980's, the South Australian Health Commission (SAHC) in conjunction with the University of Adelaide initiated a study to evaluate the risk of exposure to radiation amongst former workers at the '*Radium Hill uranium mine*'.

The initial phase of the study included (i) identification of the study cohort through the compilation of a nominal roll of ex-RH workers; (ii) determination of vital status of workers; and (iii) ascertainment of cause of death for those who died since leaving RH. Members of the cohort alive and living in South Australia were then asked to participate in a survey of their smoking habits and work histories. The findings of this study - referred to as the '*pilot study*' - were reported in an internal working paper entitled *'Deaths from Lung Cancer and Other Causes Among Former Workers at Radium Hill - An Interim Assessment'* (SAHC 1986).

The RH pilot study reported *"an elevation in the proportion of deaths attributed to lung cancer"*, and that *"lung cancer risk tended to correlate with the actual duration of time spent in ore-contact work underground"* (SAHC 1986). However, anecdotal evidence gathered from some respondents indicated that underground workers at RH were *"probably relatively heavy smokers"*. The excess of lung cancer deaths identified in this study could therefore, not confidently be attributed to radon alone. The pilot study also showed that the average duration from commencement of employment at RH to detection of lung cancer was approximately 21 years.

Meanwhile, a long-term Swedish study reported excesses of lung cancer cases even 40-50 years after initial employment (Radford and Renard 1984). In view of these findings, investigators of the pilot study recommended further long term follow-up of the entire cohort of former RH workers resident throughout Australia. The *'Radium Hill pilot study'* was then expanded to the entire cohort of former RH workers and called the *'main-study'* - henceforth, referred to as the *'Radium Hill study'* or the 'RH *study*'.

3.2.2 Aims

The primary aims of the RH Study as listed in the original study protocol were:

1. *"to determine whether uranium mine workers employed at the South Australian* RH *mine during 1952 to 1961 are experiencing an increased incidence of lung cancer as a result of radiation exposure at the mine";*
2. *"to determine the nature of any dose-response relationship between exposure to radiation at* RH *and the incidence of lung cancer."*

In keeping within the broad framework of these aims, the emphasis in this thesis is given to:

i. evaluating the lung cancer risk experienced by former workers at the RH uranium mine in South Australia;
ii. describing the exposure-response relationships pertaining to radon progeny related lung cancer mortality amongst former RH workers;
iii. studying the effects of potential confounders, effect modifiers and temporal factors on radon progeny related lung cancer mortality in the RH study cohort;
iv. interpreting the findings from RH in a global perspective .

3.2.3 My Contributions

My participation in the RH study began in early 1988, when I joined the Department of Community Medicine at the University of Adelaide as a Research Officer. At this time, a summary of individual work histories derived from archieved RH pay-roll data had been obtained on computer disks from the South Australian Health Commission; estimates had been made of radon progeny exposures, by job classification and calendar year, based on historical records of radon gas levels; a questionnaire previously used in the RH pilot study had been circulated to all those on the nominal roll of the study, resident in Australia; and requests for *cause of death traces* had been made to all death registries and cancer registries in Australia.

My contributions to the RH study commenced with the compilation and processing of the data on individual work histories and Rn exposure histories, the questionnaire survey and the death trace activities; I was responsible for the design and implementation of the data processing systems and all subsequent statistical analyses arising from this data set. Furthermore, because initial evaluation of the RH data set showed that the trace rate achieved by the above activities needed to be improved to achieve greater statistical power in the study, from 1989 onwards, I organized *secondary trace activities* to improve the trace rate and the quality of data available to the RH study; these activities are detailed in section 3.5 of this work. All analytical work presented in this thesis, excluding the estimation of Rn exposure, is a result of my own work. Components of my contributions to the RH study have also been presented at several conferences and meetings and in publications by myself and others (Woodward *et al.* 1991; Mylvaganam *et al.* 1992; Mylvaganam 1993; Mylvaganam and Woodward 1993a; Mylvaganam and Woodward 1993b; Mylvaganam *et al.* 1993a; Mylvaganam *et al.* 1993b). The aim of this work is to provide a substantive analysis of the RH study. The current chapter outlines the study design and follow-up characteristics. Descriptive epidemiological findings from the RH study population are presented in chapter 4. Analytical epidemiological findings are reported in chapters 5, 6 and 7. Implications of the findings from the RH study are discussed in a global perspective in the final chapter.

3.3 Study Design

The RH main study was a historical cohort study that included a concurrent prospective follow-up. Comprehensive records of ex-RH workers' identification particulars and job histories were sought at the commencement of the pilot study. The most comprehensive of such records available at that time consisted of pay-roll information that had been maintained by the mine management during the operation of the mine. The pay-roll information consisted of a card index that summarized employees' identification particulars and work histories pertinent to the processing of their remuneration. Individual work histories included calendar time and duration of work in each of a worker's jobs at the RH mine.

Estimates of individual exposure were based on periodic measurements of radon gas levels at various mine locations. Estimates of radon progeny were made using knowledge of ore bodies, mining information and mining activities. Reconstruction of radon progeny levels at the RH mine was undertaken by a team of health physicists.

The RH study therefore, consisted of two main phases viz.,

1. the estimation of radon daughter levels and a job-exposure matrix during the index years of the mines operations; and,
2. the conduct of cohort follow-up, including ascertainment of vital status and cause of death, work histories and smoking histories.

These phases are outlined in the forthcoming sections 3.4 and 3.5.

3.4 Evaluation of Radon and Radon Progeny Exposure

No measurements of radon progeny concentrations were known to have been made at the mine during the time of its operations. However, periodic measurements of radon gas concentrations were made, enabling estimation of radon progeny concentrations from mine ventilation and airflow levels. This work which was carried out by Dr. Phil Crouch and Ms. Claire Corani of the Radiation Protection Branch, South Australian Health Commission (Crouch and Corani 1986), is summarized in the following sections (3.4.1 - 3.4.3).

3.4.1 Mine Ventilation

Ventilation and air flow levels in the mine were known to have varied over time. At the commencement of mining activities, only natural ventilation estimated at 2 m^3 per second was known to have existed. The installation of a venturi in mid-1954 increased ventilation to 5 m^3 per second, which was further augmented to 45 m^3 per second in March 1955, when a main fan was installed.

Prior to the installation of the main fan, it was assumed that the mine was uniformly ventilated. Calculations of mean air-residence time were then made from the mine void and ventilation rate. Estimates thus obtained were used to compute *equilibrium factors* for the conversion of radon concentration estimates to estimates of radon daughter concentrations. The derivation of these equilibrium factors was based on the *'tunnel model'* of Beckman and Holub (Beckman and Holub 1979).

A separate mine ventilation model was developed for the period following the installation of the main fan. This model was based on contemporary mining practices of the time and recollections of workers from the mine. The *age of ventilation air* - mean transit time - was then estimated from mine voids and ventilation rates derived from the ventilation model. The *age* of air thus computed, was assigned to the level 4 drive of the mine from which the main fan exhausted. Thereafter, the *ages* of air in drives and stopes at other levels of the mine were estimated. *Equilibrium factors* were derived from these estimates for each location of the mine.

3.4.2 Measurements of Radon Concentrations

Radon gas was measured at the RH mine from 1954 to 1961. Air sampling was irregular at first. After 1955, a regular measurement protocol of radon concentration was started; this corresponded with concern raised worldwide over ventilation in mines (chapter 2, this work).

The installation of the main fan in March 1955 was reported to have resulted in considerably improved mining conditions. The calendar year of 1955 is therefore demarcated into two exposure periods - those prior and subsequent to the installation of fan - viz., 1st January to 31st March 1955 and 1st April to 31 December 1955..

A total of 721 routine radon measurements were made at the mine from 1954 to 1961. Estimates of radon concentrations prior to 1954 were based on radon monitoring results that were averaged for each level of the mine. After the installation of the main fan estimates of mean radon concentrations were made separately, for each *level* and three *'regions'* of mining operations. These *'regions'* comprised drives, stopes and shafts.

The number of locations measured at each level and region of the mine and the number of measurements made at each location are given in table 3.4.2a. Mean estimates of radon concentrations derived from the methods described above are presented in table 3.4.2b in the form of a *mine locality exposure matrix.*

Table 3.4.2a: Number of Locations Measured and Measurements By Mine Level

Level	Through March 1955		After March 1955					
			Drives		Stopes		Shafts	
	Locations	Measures	Locations	Measures	Locations	Measures	Locations	Measures
1	6	8	4	4	1	1		
2	2	3	4	4	3	4		
3	18	40	26	163	9	12	8	70
4	1	1	14	109	8	14		
5	2	2	19	88	6	9	8	69
6	2	2	11	63			1	1
7			4	46			1	1
8			2	2	1	1		
9								
11							1	4

Table 3.4.2b: Average Radon Concentrations By Mine Level

Level	Through March 1955	After March 1955		
		Drives	Stopes	Shafts
1	18.0	4.8	2.4	
2	15.5	3.3	7.9	
3	6.9	3.9	4.9	
4	3.7	3.1	7.1	1.8
5	6.3	2.9	3.1	
6	2.1	2.3		1.4
7		1.9		0.1
8		1.3		0.4
9			1.1	
11				3.4

3.4.3 Derivation of Radon Progeny Concentrations

The derivation of radon daughter concentrations was based on the conversion of radon concentration estimates using the estimated *equilibrium factors*. Since there were no measurements of radon concentration available prior to 1953, radon daughter concentrations for this period were assumed the same as 1953. Though radon monitoring ceased when the mine stopped producing ore in 1961, several workers remained employed at the mine during 1962. Radon daughter concentrations 1962 were assumed the same as in 1961. The result of Rn exposure estimation was an annual estimate of Rn progeny concentration for each level, and after April 1955, each *'region'* of the mine.

3.4.4 Job Exposure Matrix

The proportion of time spent in each mine region (shafts, drives, and stopes) was estimated for each job category from job descriptions and knowledge of mining methods. Mining plans were examined to determine the calendar periods during which mining activities had taken place at each level of the mine. Based on these factors, estimates were made of the proportion of time spent in the various locations (regions specific to levels) of the mine for each job category by calendar year. The proportions thus obtained were used as job-time-location occupancy weighting factors in the derivation of a two dimensional job exposure matrix.

For each job category, an estimate of annual exposure estimate was calculated by summing the time-location specific radon daughter concentration estimates weighted by the job-time-location occupancy factors. Estimates of average radon progeny concentration by underground job category and calendar time derived by the health physicists are presented in table 3.4.4a. This table was used to compute estimates of individual exposures in chapter 4.

Table 3.4.4a: The Job Exposure Matrix - Average Radon Progeny Concentrations By Underground Job Categories and Calendar Time (1953-1961)

Underground Job Category	1953	1954	1955 <Mar	1955 Mar-	1956	1957	1958	1959	1960	1961
Miner	2.1	1.2	1.4	0.34	0.39	0.48	0.55	0.62	0.58	0.59
U-G Mine Labourer	2.1	1.2	1.4	0.26	0.30	0.38	0.41	0.50	0.47	0.46
U-G Shift Foreman	1.3	0.7	0.8	0.18	0.20	0.25	0.28	0.33	0.31	0.31
U-G Shift Boss	1.7	1.0	1.1	0.23	0.27	0.33	0.37	0.44	0.41	0.41
Loader Operator	2.1	1.2	1.4	0.23	0.28	0.32	0.35	0.49	0.49	0.43
Scraper Operator	2.1	1.2	1.4	0.32	0.37	0.46	0.52	0.60	0.56	0.56
U-G Timberman	2.1	1.2	1.4	0.25	0.30	0.36	0.40	0.51	0.50	0.46
Trucker / Loco Operator	2.1	1.2	1.4	0.20	0.25	0.30	0.31	0.44	0.44	0.39
U-G Sampler	1.3	0.7	0.8	0.17	0.20	0.25	0.28	0.33	0.31	0.30
U-G Pipe Fitter	2.1	1.2	1.4	0.26	0.30	0.38	0.41	0.51	0.49	0.47
Greaser	2.1	1.2	1.4	0.21	0.24	0.32	0.33	0.41	0.40	0.38
U-G Air Hoist Operator	2.1	1.2	1.4	0.23	0.28	0.32	0.35	0.49	0.49	0.43
U-G Chainman / Surveyor	1.1	0.6	0.7	0.16	0.18	0.23	0.36	0.30	0.28	0.28
U-G Sanitary Man	1.5	0.8	1.0	0.13	0.16	0.20	0.20	0.28	0.28	0.25
Platelayer	2.1	1.2	1.4	0.23	0.28	0.32	0.35	0.49	0.49	0.43
Driller / Driller's Assistant	2.1	1.2	1.4	0.23	0.28	0.32	0.35	0.49	0.49	0.43
Snapman	2.1	1.2	1.4	0.23	0.28	0.32	0.35	0.49	0.49	0.43
U-G Storeman	2.1	1.2	1.4	0.11	0.14	0.20	0.17	0.25	0.25	0.21
U-G Fitter	2.1	1.2	1.4	0.15	0.18	0.24	0.22	0.32	0.32	0.28
Platman	2.1	1.2	1.4	0.11	0.14	0.20	0.17	0.25	0.25	0.21
Pumpman	2.1	1.2	1.4	0.11	0.14	0.20	0.17	0.25	0.25	0.21
Sinking Miner	2.1	1.2	1.4	0.11	0.14	0.20	0.17	0.25	0.25	0.21
Skipman	2.1	1.2	1.4	0.11	0.14	0.20	0.17	0.25	0.25	0.21
Geologist	0.6	0.4	0.4	0.10	0.11	0.14	0.16	0.18	0.17	0.17
Mining Engineer	1.1	0.6	0.7	0.16	0.18	0.23	0.26	0.30	0.28	0.28

3.5 Trace Activities

3.5.1 Trace procedures

Trace procedures utilised during the course of the RH pilot study were further extended during the main study. Linkage of members of the cohort to the tracing sources depended on the matching of birth dates and names. Perfect matches were generally sought, but criteria were made flexible to allow alterations in the sequence of given names and Anglicised or shortened versions of names and transcription errors in matching fields. Occasionally, matches were accepted even when these criteria were not satisfied, if the discrepancies could be plausibly explained. For example, the misrepresentation of age for the purposes of securing employment was known to have been possible. Consequently, matches were accepted even when birth dates or ages differed, if names were unusual or matched exactly and other available evidence from records traced (e.g., employment history) confirmed the individuals as being members of the study cohort.

3.5.2 Primary Trace Sources

Ascertainment of vital status included searches of electoral rolls in all Australian States and Territories, drivers' license records and motor vehicle registrations, Australian Workers Union files, and Health Insurance registers. Other sources included Department of Immigration files, State and Territorial death records and cancer registries. Telephone directories in all Australian States and Territories were also searched for those who had unusual names and were untraced through other sources. In the absence of identical name matches contact was also made with phone book entrants who had only surnames that matched (different initials), in the hope that they may have known the person being sought.

Trace procedures pertaining to the above mentioned sources are reported in further detail elsewhere (Woodward *et al.* 1988; Woodward *et al.* 1991; SAHC 1986). These sources yielded a trace rate of 60% during the pilot study and to 66% during the course of primary trace activities in the main study. As a result of my own follow-up of workers (*"secondary trace activities"*), the trace rate was further improved to 74% with secondary trace activities. Analyses of the characteristics of traced and untraced workers are detailed in the results section below.

3.5.3 Difficulties Encountered in Primary Trace Activities

The RH study cohort contained a large number of post-war European migrants who were newly arrived in Australia. Anecdotal information suggested that some of these migrants registered under their full names on the RH pay-roll, but subsequently Anglicised or shortened their names. Other information indicated that migrant workers who broke their contract of employment at RH may have changed their names to avoid identification. Name changes also occurred amongst women who married subsequent to their registration on the RH pay-roll. Reports from ex-RH workers revealed that some members of the study population had emigrated sometime after their leaving RH.

Anecdotal information thus obtained suggested that there were sources other than those consulted, capable of improving the trace rate. It was therefore thought possible that a systematic effort to consult these sources would reveal information that may otherwise have been more difficult, if not impossible, to obtain. Hence, additional procedures were undertaken to improve vital status ascertainment.

3.5.4 Secondary Trace Activities

Additional tracing efforts were conducted between 1989 and 1991. These secondary trace activities included: matching of the RH study cohort with the Wittenoom Gorge (WG) study cohort in Western Australia; searches at the Perth Chest Clinic (PCC); and contact with former employees of the RH mine who volunteered information on workmates. Cohort members were also approached for information through requests that were attached to *'thank you letters'* sent out to respondents with notices announcing the *'Radium Hill reunion'* notices; and, at the RH reunion at Easter, 1991.

These activities not only improved the trace rate and strengthened the statistical power of the study, but also enhanced our knowledge of mining activities at RH. Furthermore, these activities also provided us with an insight into the personal feelings and views of some members of the study cohort, on life at RH, their attitudes towards the occupational health risks of uranium mining and their impressions of our study. Each of these secondary trace activities are described below.

I. Wittenoom Gorge Study Cohort Comparisons

The Australian Blue Asbestos (ABA) mine at Wittenoom Gorge (WG) operated in Western Australia from 1943 to 1966 and employed 6,916 employees - 6,505 men and 411 women - (Armstrong *et al.* 1988; De Klerk *et al.* 1989). The mine was owned and operated by the Colonial Sugar Refining Company which was the only major employer in the area. The Wittenoom Gorge study was initiated to study the effects of asbestos exposure amongst workers at the ABA mine.

Comparisons between the nominal rolls of the WG study and the RH study revealed that 99 members of the Wittenoom Gorge study had also worked at RH. Electronic matching of nominal rolls resulted in 77 successful matches on surname and date of birth. The 22 contradictions identified during the comparison of nominal roll details included discrepancies in dates of birth and differences in the spelling of surnames.

For 72 of the 99 workers common to both studies, the WG study provided the initial contact source for the RH study. Furthermore, for 28 of these workers the WG study was the sole source of follow-up information for the RH study. For another 3 of these workers additional information was also obtained subsequently, from other proxy sources.

II. Perth Chest Clinic Search

In accordance with the law in Western Australia, all workers at mines were required to register with the PCC and undergo a medical examination or at least provide a recent medical report. Records maintained at the PCC included mining employment histories and particulars of individual tobacco use. Data on mining related occupational histories maintained at the PCC included detailed history of employment in mines prior to initial registration and were then updated at every follow-up visit to the PCC. Data on smoking however, were limited to smoking status at initial registration with the PCC. Furthermore, data on smoking were only available for individuals who had miners' tickets issued prior to the mid 1950's, after which the format of the registration cards was altered and data on smoking habits excluded.

III. Consultation with Ex-Radium Hill Workers

In mid-1990, we learned that an association of former RH workers - *'The Radium Hill Association'* - had been formed, that the *'Radium Hill newsletter'* (a popular newsletter during the operation of the RH mine) was revived, and that a reunion of former RH workers was being planned to commemorate the 30th anniversary of the closure of the RH mine. Secondary trace activities for the RH study were pursued through these avenues as described below.

Personal contact was also made with two former employees of the South Australian Department of Mines and Energy who had worked at the RH. One of them was a geologist who was still attached to the South Australian Department of Mines and Energy and the other, a former pay-roll clerk who was employed at the RH mine. Having been salaried employees, they were not included in the nominal roll compiled for the study. These sources were able to provide additional information on the RH mine operations, activities, record maintainance, pay-roll processing procedures and overtime, and other information that proved useful for the study.

IV. Thank You Letters to Study Participants

A reunion of former RH workers offered an opportunity to expand tracing coverage. The RH Reunion Committee had access only to the limited network of ex-RH residents who kept in contact with each other, whereas the RH study maintained an address database of traced members of the study cohort. A collaborative effort of assisting the RH Reunion Committee in mailing notices to study participants with the forthcoming comunique on the RH study was undertaken.

In early 1990, letters were mailed out to members of the study cohort for whom addresses were known, thanking them for their participation in the study, informing them of the progress made in the study and of problems faced in trace activities. Also enclosed were a reply slip to update their mailing address or volunteer further information that may be useful for the study and a self addressed stamped envelope for the return of this information; a formal notice from the RH reunion committee announcing plans for the reunion and a reunion attendance form; and, a news letter from the RH committee.

The goodwill fostered by this collaboration with the RH reunion committee enabled vital improvements in the study trace rate, and to previously collected data that were incomplete.

V. Radium Hill Reunion

The 1991 Radium Hill reunion was held during the Easter long-weekend with two gatherings taking place, one at the old mine site in RH on Friday the 29th of March and the other in Adelaide on Sunday the 31st of March. The reunion coincided with the release of a book entitled *'We Were Radium Hill'* (Harrington and Kakoshcke 1991) and the unveiling of a monument at the RH cemetery, erected in memory of RH residents who died during the time of the mines operations. Over 400 people including ex-RH workers, residents and their relatives attended the reunion at RH. The Adelaide reunion was attended by over 600 people including ex-RH residents and workers who were unable to attend the reunion at RH.

Also in attendance at both reunions were investigators from the RH study, who provided general assistance to the organizers and manned a *'Missing Person's Bureau'* (appendix A) which sought information on untraced workers and deceased workers who lacked full information. Figure 3.5.4 depicts a most unusual attempt at gathering epidemiological data - the identification and transcription of data from gravestones. Other photographs taken at the RH reunions are presented in appendix A.

Figure 3.5.4: Vital Status Ascertainment - An Unusual Attempt

3.6 The Collection and Compilation of Data

3.6.1 Follow-up Activities

Data collection activities commenced with the identification of the status of cohort members through the methods previously outlined. Follow-up activities included vital status ascertainment, and the collection of data on smoking, occupational histories, and cause of death. Instruments of data collection used in the study included questionnaires (appendix B) and summary information sheets and copies of death certificates from death registries.

I. Retrieval of Cause of Death Data

Cause of death data were obtained for cohort members who died prior to the commencement of the study or during the course of the study. Death searches were performed through death registries in all Australian states and territories. These searches for ex-RH worker deaths registered in Australia spanned the calendar periods from 1960 to 1987 (inclusive). In the event of matches and possible matches, copies of death certificates or extracts of death certificates were obtained.

Further information on vital status was also sought from relatives or friends of untraced cohort members, during secondary trace activities.

II. Questionnaire Survey

The purpose of the questionnaire survey was to ascertain the smoking and occupational histories of cohort members both prior to and following their employment at RH.

Questionnaires were administered to cohort members who were known to be alive and to next-of-kin of those who were known to have died South Australia alone, to collect data on individual smoking and occupational histories. Contact was not attempted for deaths traced in other Australian States and Territories, where death records were released on the explicit condition that no contact would be made with families of the deceased.

For deaths registered outside South Australia information on questionnaires was sought during secondary follow-up activities which included the collection of information from proxy sources.

3.6.2 Follow-up Procedures

When possible cohort members were telephoned and questionnaires were administered over the telephone by trained personnel.

Where telephone interviews were not possible, questionnaires enclosed with covering letters explaining the purpose of the questionnaire and stamped self addressed return envelopes were mailed out to traced cohort members. In the case of mail questionnaires, if no replies were obtained after ten days, reminders were made. This procedure was repeated if no replies were received after a further 10 days.

3.6.3 The Compilation of Data

Collected data were entered into an information system that was specifically designed and implemented by myself for this study. The *'Radium Hill study information system'* consisted of subsystems and data sets that could be linked by unique identifiers. Each subsystem was designed for the capture and compilation of specific data sets in the study. These data sets - *the raw data sets* - comprised the *'employee file'*, the *'questionnaire dataset'*, the *'death dataset'* and the *'job-exposure matrix'*. These raw data sets were used to derive several analytical data sets which will be described in conjunction with their use. The most basic analytical dataset used in this work was the *job exposure dataset* containing job-specific individual exposures i.e., exposures specific to each job held at RH; these individual exposures were derived by merging individual work histories from the employee dataset with the job-exposure matrix. Individual cumulative exposures obtained by summing individual exposures in the job exposure data set were then extracted and combined with the questionnaire data set and the death data set to form combined data sets which were used to produce the results on study participation and follow-up characteristics presented in section 3.7 and demographic characteristics summarized in chapter 4.

Separate analytical data sets were then derived from these basic data sets for risk evaluation using analytical approaches based on person-years at risk (chapters 5 and 6) and nested case control analyses (chapter 7). All analytical data sets used in this work were derived from algorithms and programs written by me during the course of my doctoral research. These algorithms and programs were first tested using internal validation criteria, then validated against other available programs and finally, presented for peer review at presentations made by me at the University of Southern California, USA, Epidemiology Resource Incorporated, Boston USA, and the MRC Unit of Biostatistics, Cambridge UK. Descriptions of these routines are provided in conjunction with their application; derivation of the data set for person-years based analyses is described in chapter 5 and that for nested case control analyses is described in chapter 7.

3.6.4 Data Processing Tools

Data processing was mainly performed on IBM compatible personal computers. Other computer hardware used included IBM and VAX mainframe computing systems, SUN mini-computer systems and Macintosh personal computers. The need for using this array of hardware arose due to the nature of the various aspects of data processing and statistical analyses involved in this work and the restricted availability of specialised analytical software.

Most of the data processing and statistical analyses presented in this work were performed using the Statistical Analysis System (SAS) software package and the statistical modelling package - EPICURE (EPICURE 1992). The SAS package provided a powerful programming language for data manipulation, extensive procedures for data processing and modules for full-screen data entry, statistical analyses, statistical graphics, matrix manipulations and report generation (SAS 1990). The use of EPICURE is described in conjunction with its applications in the forthcoming chapters of this thesis. Other software used in this work included, the spreadsheet packages Excel - which was used on both, IBM and Macintosh PCs - and LOTUS 1-2-3; the Fortran 77 programming language which was used on the mainframe and the SUN minicomputer systems and on IBM PCs; and, the generalized linear modelling package GLIM which was initially used for analyses before the availability of EPICURE. The use of several other epidemiological packages was also explored prior to the final choice of software used in this work; these software are referred to in relevant sections of this thesis.

3.7 Results: Study Participation and Follow-up Characteristics

3.7.1 Cohort Definition

The study population was based on a nominal roll that was compiled from payroll records maintained by the South Australian Department of Mines and Energy of its employees at the RH mine during the operation of the mine. These pay-roll records had only been maintained for wage earners employed at RH by the South Australian Department of Mines and Energy; they therefore did not include salaried employees of the South Australian Department of Mines and Energy (e.g., management, and professional staff such as geologists) or contractors and their staff. It should therefore be noted that this study population consists only of wage earners employed at RH by the South Australian Department of Mines and Energy, and not of the entire cohort of ex-RH workers. The nominal roll thus compiled included the records of 2,574 ex-RH workers, 2,521 (98%) males and 53 (2%) females.

All analyses pertaining to Rn induced lung cancer risk evaluation undertaken in this work are based on the only cohort of male workers. Critical appraisal of trace rates is therefore restricted to the male cohort. Females in the study cohort are only included for purposes of summarizing the basic descriptive and demographic characteristics of the RH cohort - reported in this chapter and the next (chapters 3 and 4).

3.7.2 Trace Rates

Search procedures were carried out sequentially. The various sources were systematically searched, with the search list being shortened at each stage when successful matches were identified. Therefore, results presented in this section must be considered in the light of the fact that trace statistics on trace sources need not be mutually independent; i.e., once an individual was identified through one source, further searches were not attempted for this individual through any other source.

Trace activities continued to the end of December 1991 by which time some information was available for about three-quarters (74%) of the study cohort; there was no vital status information on the remaining quarter. The 1,894 individuals for whom information was available comprised 1,850 (98%) men and 44 (2%) women. A similar sex-ratio existed amongst those untraced to this date, 671 (99%) men and 9 (1%) women. The trace rate was 73% amongst males and 83% amongst females. These figures are summarized in table 3.7.2.

Since systematic death searches had been made only to the end of 1987, the analytical dataset is based on a follow-up to 31 December 1987. Of those traced through 1987, 615 (32%) cohort members were known to be dead, 1,113 (59%) were known to be alive and 137 (7%) were censored for other reasons at various times during the follow-up period of the study. The status of a further 29 (2%) was unknown, though proxy questionnaire information was obtained from former RH workmates. The deaths included 45 who were alive at the time of initial trace and died during the follow-up period. At the end of study trace activities, a total of 645 deaths was recorded amongst members of the study cohort who were traced; 615 of these deaths occurred during the follow-up period; the other 30 deaths occurred between 1988 and 1991. These persons were included in the dataset as *'alive at the end of the study period'*.

Table 3.7.2: Distribution of Trace Rates by Gender

Trace Status		Gender		
		Males	Females	All
Untraced	Number	671	9	680
	Overall %	26.07	0.35	26.42
	Row %	98.68	1.32	
	Column %	26.62	16.98	
Traced	Number	1850	44	1894
	Overall %	71.87	1.71	73.58
	Row %	97.68	2.32	
	Column %	73.38	83.02	
All	Number	2521	53	2574
	%	97.94	2.06	100.00

3.7.3 Questionnaire Survey Data

Some questionnaire data were obtained for 1,582 individuals who comprised 84% of those traced and 61% of the entire cohort. Of these, 1,136 were known to be alive and 260 known to be dead at the time of questionnaire completion. Though the status of the remaining 166 were unknown, some smoking and work history particulars were obtained through third parties. For analytical purposes these individuals were treated as lost-to-follow-up from the last known date of survival and were censored at this date; these individuals were classified as *censored prior to completion of follow-up*.

Questionnaire data were obtained in varying degrees of completion with 1,115 being fully completed, 456 partly completed, and no information at all being available on 11 cohort members traced who comprised 8 refusals to participate and 3 non-responses. Though there was a total of 14 refusals to participate in the study, some questionnaire information was available from proxy sources on 6 of them, which enabled the partial completion of questionnaires. The refusals comprised 13 cohort members and relatives of another who died after the follow-up cut-off period.

Respondents to questionnaires included 1,100 cohort members answering for themselves - 4 of whom were resident overseas - and 104 relatives answering on behalf of cohort members who could not be contacted personally. Some questionnaire information was obtained from proxy sources for a further 367 cohort members. Proxy sources providing questionnaire information primarily comprised 119 work mates and friends, 3 death certificates, the Wittenoom Gorge (WG) study records for 31 others and Perth Chest Clinic (PCC) records for 214 others.

Initial contact particulars for the majority (936) of respondents were obtained through electoral roll searches. One hundred and fifteen others were located through Health Insurance Commission records, 75 others through searches of motor vehicle registrations, 8 more through phone book searches for unusual untraced names, another cohort member was located through Australian Workers Union files and one other from a death certificate; the initial contact source for 35 others traced during the RH pilot study was unknown. Secondary trace activities conducted thereafter, contributed further information on 187 workers through the PCC, 72 through the WG study, 85 through the RH reunion and 67 others through other workers and friends.

3.7.4 Death Trace Data

A total of 615 deaths amongst cohort members was confirmed to have occurred during the study period. This included 606 males (98%) and 9 females. Six hundred of these deaths were known to have occurred in Australia, two deaths were reported to have occurred overseas (one each in Italy and the United Kingdom), and the places of death for the others were unknown. The majority of deaths traced in Australia were from South Australia, where the RH mine was situated.

Death certificates or certified abstracts were obtained for 425 (69%) of these individuals. Detailed death information - extracted from death certificates - was obtained in the case of a further 133 (22%) cohort members from data collected during the RH pilot study. Particulars of the remaining 57 (9%) deaths were obtained from other sources including friends and relatives. Details of cause of death were available in 600 (98%) instances. Underlying cause of death was coded by a medically qualified epidemiologist (Dr. A. Woodward) using the ICD 9 classification (ICD 9 1978).

No detailed information was available in the case of 15 (2%) individuals who were reported as being deceased by former RH workmates. A further 11 deaths were also reported where dates of death were unavailable. For analytical purposes, where dates of death was not available, individuals were censored at the last known date of survival. Unknown cause of death was coded with ICD 9 (ICD 9 1978) code 799.9.

3.7.5 Questionnaire and Death Data

Of the 615 deaths recorded amongst cohort members, some questionnaire data were obtained for 303 (49%) of them. This included 44 cohort members who were alive at the time of the questionnaire survey and were known to have died subsequently.

A total of 290 (47%) deaths were registered in South Australia, where explicit permission was obtained from the Registrar General, to use particulars provided in the death certificate for the purposes of contacting relatives of the deceased. Follow-up of a sample of deaths from recent years resulted in 22 successful contacts. Questionnaire data were also obtained for a further 32 of those who died in South Australia through other contact procedures; in one unusual case, both death and questionnaire data were obtained through information that appeared in an obituary notice.

3.7.6 Analyses of Trace Rates for Male Workers

I. Trace Rates By Vital Status

A summary of questionnaire and death data obtained for male workers is presented in table 3.7.6a. Some questionnaire data were available on 61% of the male cohort. This included about half of the 606 (24%) males who were know to have died during the study follow-up period. Vital status ascertainment could not be made on 671 (27%) males who were considered lost-to-follow-up from their last date of employment at RH.

Table 3.7.6a: Trace Characteristics for Traced Male Workers

Death	Questionnaire Data			
Data	Yes	%	No	Total
Yes	295	49%	311	606
No	1233	65%	671	1904
Total	1528	61%	982	2510*

NOTE: * Figure excludes 11 male workers on whom some information was available but dates of death were missing.

II. Birth Cohort Profile By Trace Status

The analyses data sets included only the 2,521 males in the study cohort, of whom 671 (26.6%) were untraced. The 1,849 males traced comprised 1,112 (60%) who were alive, 606 (33%) who were dead by the end of the study, and 131 (7%) individuals on whom some additional information was available but, who were lost-to-follow-up and therefore, censored-prior-to-the-completion-of-follow-up. Those referred to as censored prior to the completion of follow-up (in table 3.7.6b) do not include the untraced members of the cohort who were lost-to-follow-up from the date their employment at RH was terminated. Table 3.7.6b which summarizes the characteristics of this population, shows that the subset of those known to be dead are (not surprisingly) significantly older than those known to be alive. Furthermore, it can also be seen that the untraced and censored fractions of the male cohort - together comprising the cohort fraction *lost-to-follow-up* - would be of about the same age as those traced and known to be alive.

Table 3.7.6b: Birth Cohort Profile For Males in Respect of Study Trace Status

Year of Birth	Alive	Dead	Censored-Prior-To-Completion of-Follow-up	All Traced	Untraced
Mean	1928	1916	1926	1924	1926
Median	1929	1918	1929	1926	1929
Range	1898-1946	1888-1945	1894-1942	1888-1946	1883-1946
I-Q Range	1924-1933	1909-1925	1922-1933	1919-1931	1923-1932
SD	7.6 years	11 years	10 years	10 years	9 years
N	1112	606	131	1849	667
%	44.2	24.1	5.2	73.5	26.5

Note: I-Q = Inter-Quartile Range; SD = Standard Deviation; N = Number of Observations.
* Traced beyond employment at RH but censored prior to completion of follow-up, for reasons other than death.

III. Trace Rates By Radon Exposure Status

Of the male workers at RH 59% worked underground sometime during their employment at RH (and hence are defined as "*ever exposed*"). Table 3.7.6c summarizes follow-up trace rates by the exposure status of individual male workers. This classification of exposure status differs from exposure categorizations used in the analytical epidemiological chapters where all analyses are based on *relevant exposures* - defined with a lag period taken into consideration.

Table 3.7.6c: Trace Rates for Male Workers By Exposure Status

Ever Exposed	Traced			
	Yes	%	No	Total
Yes	1100	75%	359	1459
(Mean BC)*	(1925)		(1928)	(1926)
No	750	71%	312	1062
(Mean BC)*	(1923)		(1925)	(1923)
Total	1850	74%	671	2521
	(1924)		(1927)	

Note: * Mean Birth Cohort Year

The trace rate was higher among those ever exposed to Rn at RH (75%) compared to those never exposed (71%); 59% of those traced had worked underground at RH. Table 3.7.6c also lists the mean birth cohort year for workers by trace status and exposure status. These results show that the untraced cohort fraction was slightly older (born earlier) than that traced, and that the cohort fraction who worked underground (*exposed to* Rn*)* at RH was younger than the cohort fraction of surface workers (*unexposed*).

IV. Death Trace By Radon Exposure Status

Examination of the RH cohort by exposure status and mortality status (table 3.7.6d) showed that compared to surface workers, a smaller proportion of underground workers died before the end of the study period. Results also showed that deceased underground workers comprised a younger birth cohort than deceased surface workers. The fraction of the cohort who died during the study period was on average 10 years older than those who were not know to be dead, i.e., survivors and those untraced.

Table 3.7.6d: Deaths Traced for Male Workers By Exposure Status

Ever	Dead			
Exposed	Yes	%	No	Total
Yes	327	22%	1132	1459
(Mean BC)*	(1919)		(1928)	(1926)
No	279	26%	783	1062
(Mean BC)*	(1915)		(1926)	(1923)
Total	606	24%	1915	2521
	(1917)		(1927)	

Note: * Mean Birth Cohort Year

It was also found that 54% of those who died during the study period had worked underground at RH.

3.7.7 Summary of Findings on Study Participation and Follow-up

An overall trace rate of 74% was achieved in the follow-up of the RH study cohort. At the end of the study period, 31 December 1987, 24% of the cohort (33% of those traced) were dead and 5% of the cohort (7% of those traced) were censored-prior-to-completion-of-follow-up. Approximately 44% of the RH cohort were eligible for further follow-up, should the study be extended beyond the cut-off date identified. Though the term *'completed follow-up'* is used in epidemiological terms to refer to who died or were lost-to-follow-up at the end of the study period, vital information on smoking and other occupational exposures was not available for about half of those who had *completed* their follow-up. This information could still be sought through proxy sources if the period of study was to be extended in the future.

Analyses of trace rates showed a higher trace rate amongst those who were exposed - worked underground - at RH; the heterogeneity in trace rates for surface (75%) and underground (71%) workers was statistically significant ($p < 0.05$) based on a χ^2 test for homogeneity in table 3.7.6c. This may indicate bias in trace rate, suggesting that those who were exposed to Rn progeny at RH were more likely to respond to the questionnaire survey, but this finding has to be addressed with some caution. Statistical significance based on the a χ^2 statistic tends to be influenced by large cell frequencies such as those in table 3.7.6c, and may result in significance with little meaning. Two components of bias may have been possible: in addition to the possibility of *response bias* mentioned above, bias could have been possible on the part of study organizers placing greater emphasis on tracing those who were exposed. Since the people who carried out the trace activities were on the whole, not aware of the exposure status of the individuals being sought, such a bias is unlikely; individual exposure histories were not directly available on the nominal roll used in study trace activities.

The issue of *bias* in epidemiological study design, selection, losses-to-follow-up and information has been widely discussed in the literature (Greenland 1977; Criqui *et al.* 1978; Kleinbaum *et. al.* 1981; Checkoway *et al.* 1989). Kelsey *et al.*, (Kelsey *et al.* 1986) summarize that: with reference to prospective studies it appears more likely that participating cohort fractions may differ with respect to exposure than with respect to disease; if the *selection fractions* differ with respect to disease only, then the risk ratio is biased with the magnitude of the bias being small for plausible differences in selection fractions if the disease affects only a small proportion of the population; if the fractions differ according to the presence or absence of a risk factor (exposure) the resulting bias may be controlled by methods similar to those used in controlling confounding (Miettenen and Cook 1981); the more serious bias results when selection fractions vary according to specific combinations of exposure and disease status. Greenland (Greenland 1977) examined bias in cohort studies where he defined *study loss* as comprising the proportion of persons refusing to participate in the study - non response - and the proportion of persons who enter the study but, dropped out before completion of their follow-up; *study participation* was defined as the complement of *study loss,* and comprised *response bias* and *follow-up bias*. Greenland (Greenland 1977) refers to some of the earliest reports on bias in cohort studies (Berkson 1955; White and Bailar 1956; US PHS 1964) and the possible misinterpretation of concepts on bias (MacMahon and Pugh 1970), examines the effects of bias using a hypothetical example and concludes that selection bias is a theoretical possibility whenever correlates of the outcome capable of influencing study participation exist for some participants at the beginning of the study; in evaluating the likelihood of an important bias in a study situation he states that *"many experienced epidemiologists tend to regard selection bias in cohort studies as generally unimportant"* and concludes that the intended use of study results must play an important role in determining the degree of bias necessary to be considered significant.

Empirical data are seldom available for full evaluation of the extent of these biases; therefore, the common view is that everything possible should be done to attain high participation rates. Sources of bias that may have arisen in this work include bias in trace rates - particularly response bias. Important questions that may consequently arise in the RH study include:

i. could trace rates have been influenced by exposure status? - This question has already been addressed in the preceding paragraph (the answer is, most probably not);

ii. could trace rates have been influenced by the disease outcome being studied? - Several issues arise here; firstly, there is a greater likelihood that those who died during the study period may have been traced because of the systematic searches of death registries; therefore, it is very likely that mortality is over-estimated in this cohort, as deceased workers were more likely to be traced than were workers alive; the extent of this bias cannot be measured because those who were identified as being deceased would have been classified as *traced* whereas those who were *untraced* may be dead or alive with the latter possibility being more likely; secondly, by searching the records of cancer registries and chest clinics there may have been a greater likelihood of tracing cancer deaths including lung cancer deaths.

It is possible that such biases may exist in the RH study; taken individually, specific trace procedures undertaken in this study may be associated with specific types of response bias. However when combined, these trace procedures contribute towards an attempt at tracing the entire cohort. Ideally, if this were possible, the avenues by which such a *complete trace* was achieved, even if they were biased, would have little overall impact; in other words, this may be described as an instance where a compromise is made between *local bias* and *global unbiasedness,* where even *locally biased avenues of tracing* could produce *globally unbiased results* if they eventually enable complete trace.

Concerns may be raised about bias resulting from secondary trace activities. Though our findings show considerable improvement in trace rates, person-years of follow-up and the amount of data on individual smoking and occupational histories, the findings cannot quantify the bias these activities may have contributed; qualitative arguments and conjectures abound.

The secondary trace activities which have formed a part of this work are by no means complete; they were embarked upon when conventional trace activities had reached a point of saturation in trace rates during primary trace activities. This first attempt at secondary trace activities has brought results that show promise. In the absence of other avenues of tracing the complete cohort, it appears reasonable that methods such as these, however innovative, should be carefully explored and if satisfactory, executed towards the ultimate goal of *complete trace* - an end that justifies the means.

CHAPTER 4

4. Descriptive Epidemiological Findings

Tables and Figures

Page

4. Descriptive Epidemiological Findings

4.1 Overview: Population Profile

The average male worker at RH was aged 31 at the time of entry into the cohort, worked a total of 17 months at RH, held 2.4 different jobs, was exposed to 4.25 WLMs and aged 32 at the end of employment at RH; the mean duration of follow-up was 19.5 years and the mean age at the end of follow-up was 50 years. A total of 85% (1,121 out of 1,320) of male workers used tobacco products sometime in their lives. Detailed smoking information was obtained for 84% of those who ever smoked regularly. Amongst males for whom detailed smoking information was available, 510 (54%) reported they were current smokers and 433 (46%) were ex-smokers at the time of interview.

The average female employee at RH was 27 at the commencement of work at RH, worked a total of 12 months at RH, held 1.3 jobs at RH, was not exposed to any radiation whilst working at RH and was aged 28 at the end of employment at RH. The average duration of follow-up for women was 21.5 years and their mean age at the end of follow-up, 49 years. The prevalence of *ever-smokers* amongst women at RH was 60% (25 out of 42).

This chapter provides a comprehensive description of the study cohort. Critical appraisal of demographic characteristics is restricted to male workers, who are the subjects of the major analyses in this study. A broad summary of demographic characteristics of the female cohort is provided where ever appropriate.

4.2 Population Characteristics

4.2.1 Commencement of Employment at Radium Hill

Entry into the study cohort was defined as the date of first inclusion on the RH payroll. The annual intake of workers at RH is summarized in table 4.2.1. One hundred and twenty three members of this cohort (all males), worked at RH prior to 1952, when the mine commenced production of ore. These workers were involved with the initial preparatory activities that were required at the mine site. An additional 244 members of the cohort, which included 1 woman, commenced work at RH in 1952 when the mine commenced ore production.

The highest intake of workers occurred during 1955 - the year in which additional mechanical ventilation was introduced underground - when 453 men (18% of total male intake) and 10 women (19% of total female intake) commenced employment. Nineteen fifty five was also the median year of employment for men with 53% of all males being employed by the end of that year. Approximately half the intake of males occurred prior to the installation of additional ventilation, a time when exposure levels were substantially higher.

Table 4.2.1: Commencement of Employment at Radium Hill

Calendar	Males		Females		Total Cohort	
Year	N	Cum %	N	Cum %	N	Cum %
1948	5	+	-	-	5	+
1949	3	+	-	-	3	+
1950	35	2%	-	-	35	2%
1951	80	5%	-	-	80	5%
1952	234	14%	1	2%	235	14%
1953	246	24%	1	4%	247	24%
1954	280	35%	8	19%	288	35%
1955	453	53%	10	38%	463	53%
1956	226	62%	6	49%	232	62%
1957	136	67%	6	60%	142	67%
1958	224	76%	7	74%	231	76%
1959	222	85%	3	79%	225	85%
1960	253	95%	6	91%	259	95%
1961	122	100%	5	100%	127	100%
1962	2	100%	-	100%	2	100%
Overall	**2521**	**100%**	**53**	**100%**	**2574**	**100%**

4.2.2 Termination of Employment at Radium Hill

Terminations of employment at RH occurred from 1951 through till 1962; one male employee left the occupational cohort in 1951. By 1958, about half the cohort (53%) had ended their employment in RH. Annual exits from the RH workforce are summarized in table 4.2.2a.

Table 4.2.2a: Termination of Employment at Radium Hill

Calendar	Males		Females		Total Cohort	
Year	N	Cum %	N	Cum %	N	Cum %
1951	1	+	-	-	1	+
1952	31	1%	-	-	31	1%
1953	338	15%	1	2%	339	14%
1954	225	24%	4	9%	229	23%
1955	353	38%	9	26%	362	37%
1956	236	47%	4	34%	240	47%
1957	162	53%	6	45%	168	53%
1958	201	61%	6	57%	207	61%
1959	234	71%	6	68%	240	71%
1960	290	82%	7	81%	297	82%
1961	415	99%	9	98%	424	99%
1962	35	100%	1	100%	36	100%
Overall	**2521**	**100%**	**53**	**100%**	**2574**	**100%**

The largest exits from employment were registered during 1953 (13%) - soon after the mine commenced its production activities, in 1955 (14%) - when the new ventilation system was installed, and in 1961 (16%) - when the mine stopped production of ore. The final exodus in 1962 comprised 35 male workers and 1 female worker.

Employment termination codes obtained from payroll information are summarized in table 4.2.2b. Eleven members of the RH cohort, all of whom were men, died during the term of their employment at the mine. The majority of RH employees (94%) terminated their employment with the South Australian Department of Mines and Energy at the end of their time at RH. Three percent of the cohort were transferred out of RH. Two percent of the RH study cohort, comprising 46 men and 2 women, were transferred to other sections of the South Australian Department of Mines and Energy. A further 32 males (1%) were transferred to other government departments. Anomalies in payroll record included 64 employment records (3%) for which termination codes were not available. Furthermore, five men who were included in the RH payroll did not actually commence work at RH and were therefore classified as *'non-starters'*.

Table 4.2.2b: Radium Hill Employment Termination Status

	Males		Females		Total	
Termination Status:	**N**	**%**	**N**	**%**	**N**	**%**
Death	11	+	-	-	11	+
Discharged	2365	94%	49	92%	2414	94%
Transferred To:						
Dept. of Mines	46	2%	2	4%	48	2%
Other Govt. Depts.	32	1%	-	-	32	1%
MM - Missing	62	3%	2	4%	64	3%
Did not start work	5	+	-	-	5	+
Overall	**2521**	**100%**	**53**	**100%**	**2574**	**100%**

4.2.3 Size of Employed Cohort

Table 4.2.3 summarizes the available information on the size of the RH workforce. A total of 357 (14%) workers, including 1 woman, were employed at RH in 1952 - the first year of uranium ore production. During 1961, the last year when uranium ore was produced at RH, 458 (18%) - 448 men and 10 women - worked at the RH. A total of 36 workers - 35 men and 1 woman - remained employed sometime in 1962, when the mine was finally closed. Exact termination dates were unavailable for 64 cohort members.

Table 4.2.3: Number of Employees and Cohort Fraction Employed

Year	Males		Females	Total Population	
	N	Cohort Fraction	N	N	Cohort Fraction
1948	5	+	-	5	+
1949	8	+	-	8	+
1950	43	2%	-	43	2%
1951	123	5%	-	123	5%
1952	356	14%	1	357	14%
1953	572	23%	2	574	22%
1954	513	21%	9	522	20%
1955	741	30%	15	756	29%
1956	614	25%	12	626	24%
1957	514	21%	14	528	21%
1958	576	23%	15	591	23%
1959	596	24%	12	608	24%
1960	616	25%	12	628	24%
1961	448	18%	10	458	18%
1962	35	1%	1	36	1%
Overall	**2521**	**100%**	**53**	**2574**	**100%**

Figure 4.2.3a shows the cumulative percentages of male workers commencing and terminating employment at the RH mine over time; the vertical distance between the two curves give the cohort fraction employed at that time. Variations in the cohort fractions employed are further described in figure 4.2.3b.

Figure 4.2.3a: Commencement and Termination of Employment at Radium Hill (For Males Only)

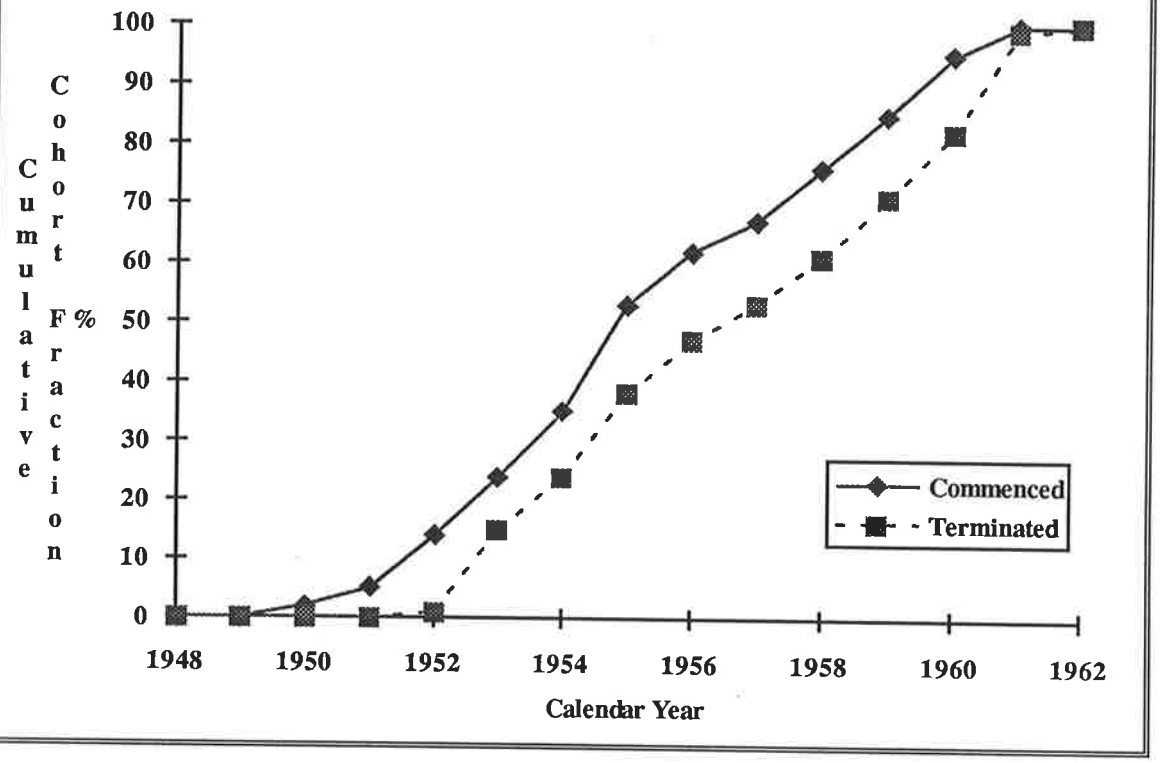
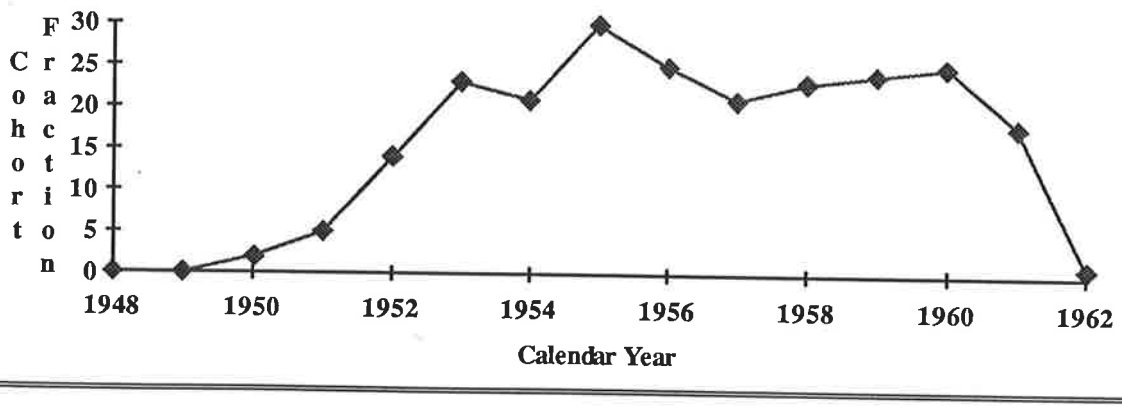


Figure 4.2.3b: Cohort Fraction Employed at Radium Hill (For Males Only)



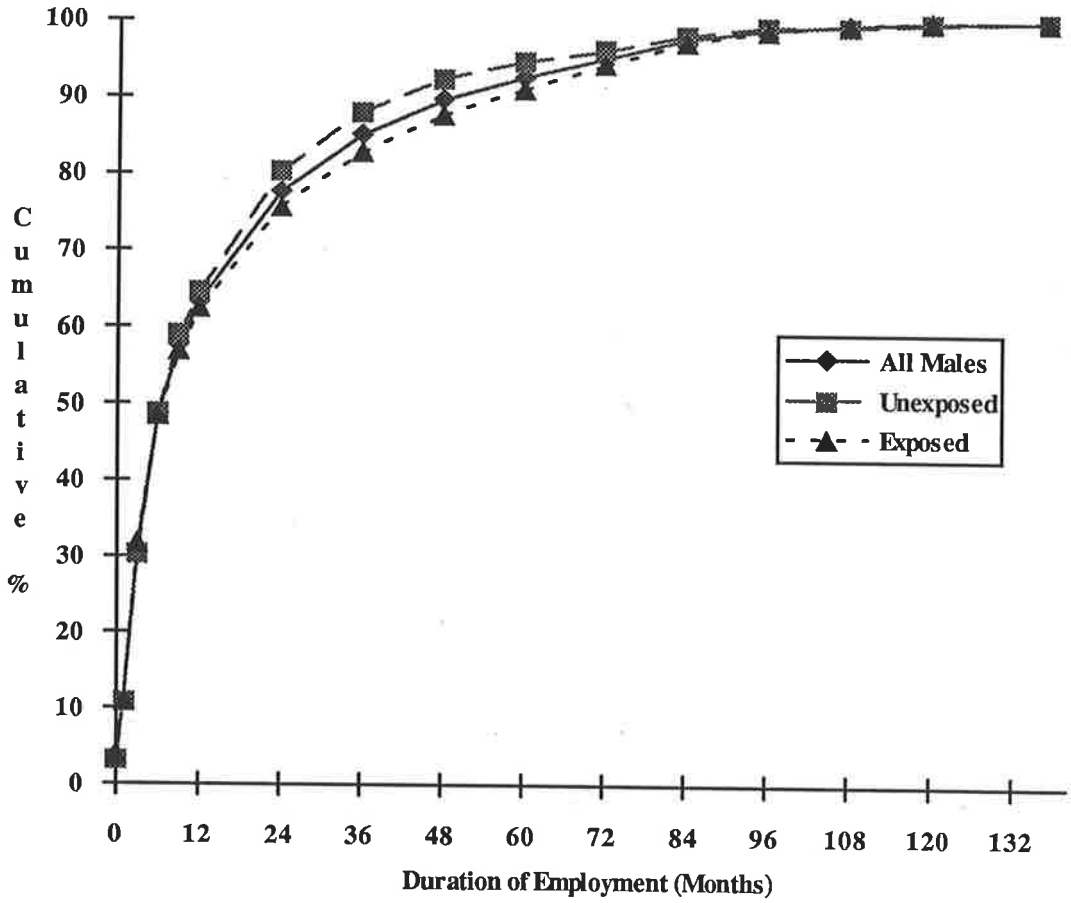
4.2.4 Duration of Employment

Job classifications showed that a total of 401 person years was allocated to '*inactive*' job categories i.e., job codes used when individuals were not actively employed at RH. Therefore, employment time was computed on the basis of active employment time and not on the interval of time between first commencement and final termination of employment at RH.

The distribution of duration of employment for male workers was extremely skewed; duration ranged from a fortnight to 11.5 years. The mean duration was 17 (95% CI: 16-18) months, and the median duration was 7 months. The inter-quartile range was 3 - 21 months. Only 5% of the male cohort were known to have worked at RH for over 6 years.

Distributions of duration of employment for RH workers are presented in figure 4.2.4. Plots of duration of employment by exposure status at RH show very little difference between those exposed and unexposed to Rn at RH.

Figure 4.2.4: Distribution of Cumulative Duration of Employment at Radium Hill (For Males Only)



4.2.5 Termination of Follow-up

Table 4.2.5 shows the annual distribution of the termination of follow-up. The 1,170 who completed their follow-up in 1987 (the last year of follow-up) included 1 cohort member who migrated from Australia and thereafter was considered lost to follow-up and 26 deaths that occurred during that calendar year. Consequently, 44% of the study cohort (comprising 60% of those traced) were known to be alive at the end of the study period. Those censored before the end of the study included 24% who were censored at the time of their death.

Table 4.2.5: End of Study Follow-Up

Calendar Year	Males		Females		Total Cohort	
	N	Cum %	N	Cum %	N	Cum %
1951	1	+			1	+
1952	12	1%			12	1%
1953	94	4%			94	4%
1954	56	7%	1	2%	57	6%
1955	95	10%	3	8%	98	10%
1956	73	13%	2	11%	75	13%
1957	62	16%			62	16%
1958	73	19%	1	13%	74	18%
1959	89	22%	2	17%	91	22%
1960	100	26%	1	19%	101	26%
1961	127	31%	3	25%	130	31%
1962	28	32%			28	32%
1963	14	33%			14	33%
1964	10	33%			10	33%
1965	11	34%			11	33%
1966	23	34%			23	34%
1967	18	35%	1	27%	19	35%
1968	16	36%			16	36%
1969	24	37%			24	37%
1970	26	38%			26	38%
1971	11	38%	2	30%	13	38%
1972	22	39%	1	32%	23	39%
1973	23	40%			23	40%
1974	38	42%	1	34%	39	41%
1975	28	43%			28	42%
1976	22	44%			22	43%
1977	29	45%			29	44%
1978	20	45%	1	36%	21	45%
1979	29	47%			29	46%
1980	28	48%			28	47%
1981	28	49%	1	38%	29	49%
1982	34	50%			34	50%
1983	37	52%			37	51%
1984	26	53%			26	52%
1985	33	54%	1	40%	34	54%
1986	22	55%	1	42%	23	55%
1987	1139	100%	31	100%	1170	100%

4.2.6 Duration of Follow-Up

Duration of follow-up ranged from a fortnight to 39.8 years. The mean duration of follow-up was 19.6 years (95% CI: 19.1-20.1 years) - 19.5 years for men and 21.5 years for women. The median duration of follow-up was 26.7 years - 26.7 years for men and 27.8 years for women.

The mean duration of follow-up amongst males who were exposed to Rn at RH was 19.9 (95% CI: 19.2 - 20.6) years and 19.0 (95% CI: 18.2 - 19.8) years amongst those unexposed to Rn at RH. Distributions of follow-up duration (figures 4.2.6a and 4.2.6b) showed very little difference between those exposed and unexposed to Rn at RH.

Figure 4.2.6a: Cumulative Frequency Distribution of Duration of Follow-up (For Males Only)

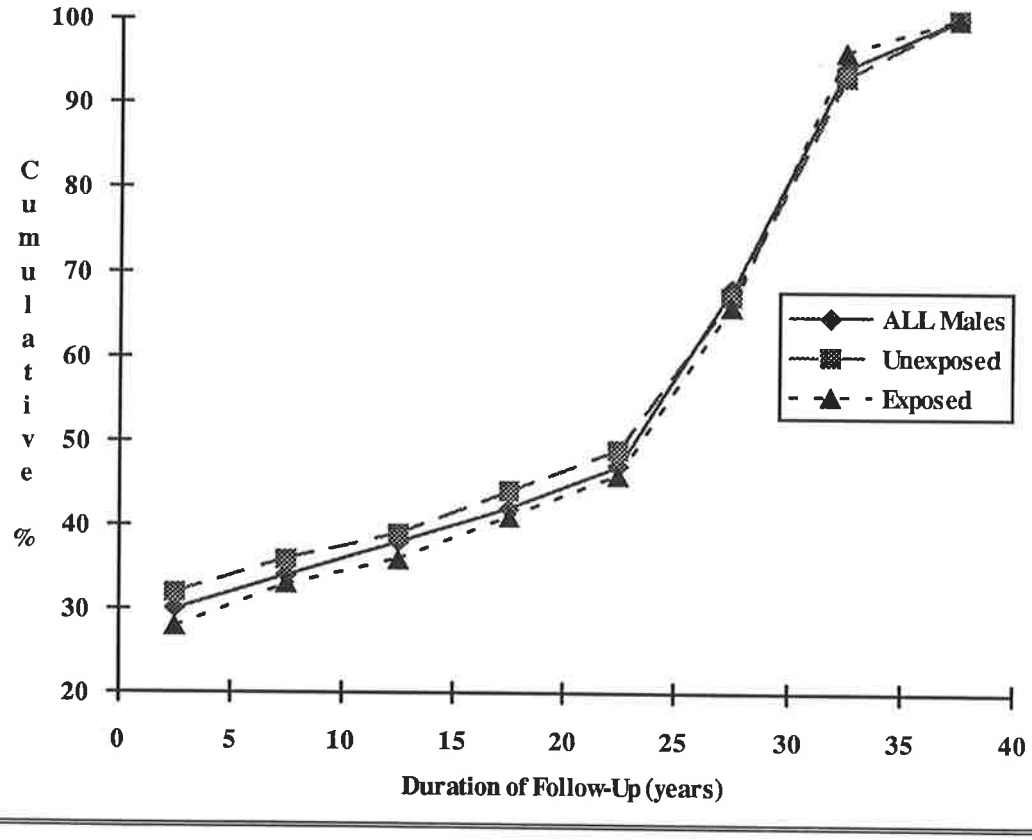
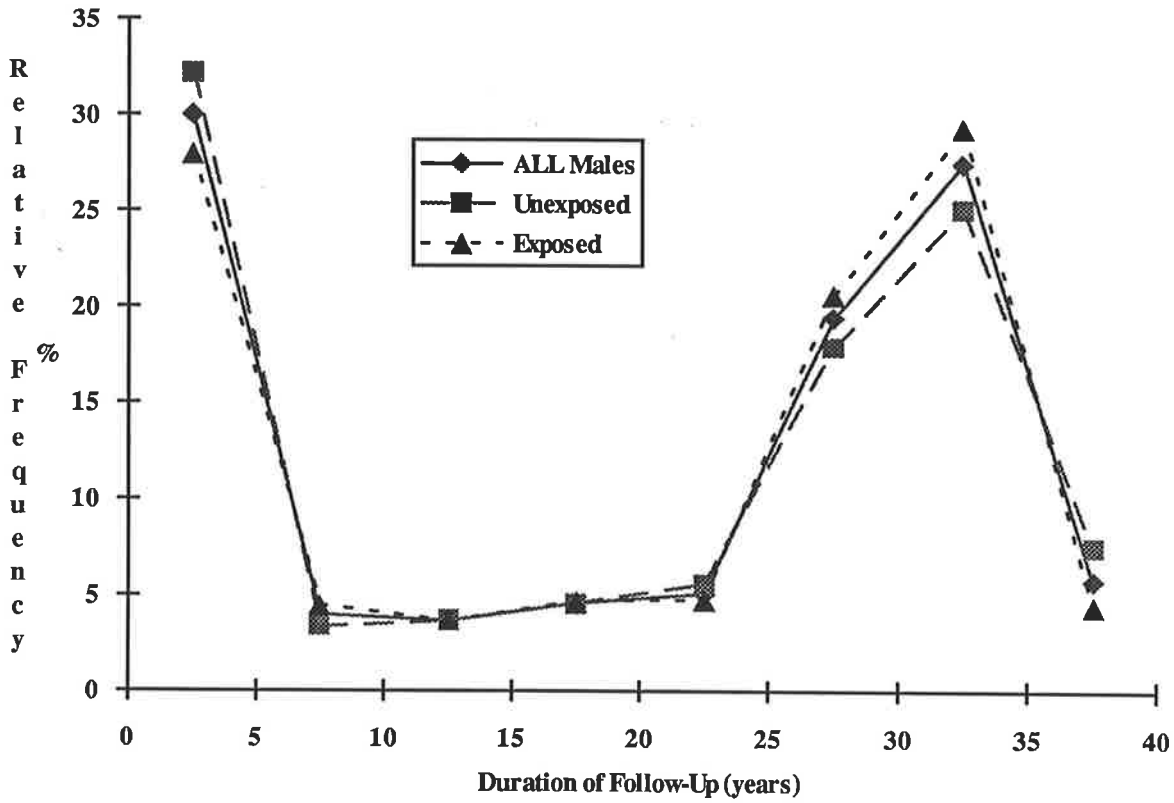


Figure 4.2.6b: Distribution of Duration of Follow-up
(For Males Only)



4.3 Demographic Characteristics

4.3.1 Age at Commencement of Employment at Radium Hill

The mean age of males at commencement of work in RH was 31 years (95% CI: 30.6-31.4). The distribution of age at commencement was skewed with half the cohort commencing employment by the age of 29 and three-quarters of the male cohort by the age of 36 (inter quartile range 24-36) years. This pattern was however not consistent over time. As may be seen from table 4.3.1, the age of males at hire was higher during the first three years of the mine's operation (range 19-60); perhaps reflecting more experienced workers being hired in earlier years during the preparation of the mine.

Anecdotal information obtained from former RH residents revealed that teenage students living in RH were employed at the mines on an ad hoc basis during the school holiday seasons. This may account for the presence of younger members of the cohort from 1951.

Female members of the RH workforce commenced work at the mean age of 28 years (standard deviation: 12 years). Ages at commencement for females ranged from 14 to 54 years; half the female work force commenced employment at RH by the age of 27 years (median = 27 years), and the inter-quartile range of age at commencement for females was 16 to 38 years.

Table 4.3.1: Age at Commencement of Work at RH By Calendar Year (For Males Only).

Year	N	Mean & 95% CI (years)	Median (years)	Range (years)	I-Q Range (years)
1948	5	34 ± 9	36	19-42	29-42
1949	3	33 ± 9	29	28-43	28-43
1950	35	32 ± 4	27	20-60	23-42
1951	80	34 ± 3	29	14-62	25-43
1952	233	33 ± 1	30	16-63	26-39
1953	245	33 ± 1	31	15-59	26-40
1954	279	31 ± 1	29	17-71	25-37
1955	453	30 ± 1	28	16-63	24-35
1956	226	30 ± 1	27	15-60	23-34
1957	135	31 ± 1	30	15-63	25-35
1958	224	29 ± 1	28	14-58	23-33
1959	222	29 ± 1	26	14-62	22-33
1960	253	30 ± 1	28	14-58	22-34
1961	122	29 ± 1	28	14-63	23-34
1962	1	48	-	-	-
Overall	**2516**	**31 ± 0.4**	**29**	**14-71**	**24-36**

Note: N - Frequency; CI - Confidence Interval; I-Q - Inter-Quartile

4.3.2 Age at Termination of Employment at Radium Hill

At the termination of employment at RH the average age of males was 31.8 (95% CI: 30.4 - 31.2) years, and 29 (95% CI: 24.6 - 30.8) years for females. Half the male cohort had completed their employment at RH by the age of 30. The annual distribution of age at termination of employment (table 4.3.2) shows that most males employed at RH were aged in their 30's when they left their employment at RH (mean ages at termination of employment ranged from 30 to 39 years).

Table 4.3.2: Age at Termination of Work at RH - Males Only

Year	N	Mean & 95% CI (years)*	Median (years)*	Range (year*	I-Q Range (years)*
1951	1	59	59	59 - 59	59 - 59
1952	31	35 ± 4	29	22 - 63	26 - 43
1953	336	34 ± 1	32	18 - 60	27 - 41
1954	224	33 ± 1	31	19 - 71	26 - 38
1955	353	31 ± 1	29	18 - 64	25 - 36
1956	236	32 ± 1	29	18 - 63	24 - 37
1957	161	33 ± 2	30	15 - 65	26 - 37
1958	201	32 ± 1	30	18 - 69	26 - 36
1959	234	30 ± 1	28	14 - 61	23 - 35
1960	290	32 ± 1	30	15 - 62	24 - 37
1961	415	32 ± 1	31	14 - 64	26 - 37
1962	34	39 ± 4	38	17 - 65	32 - 48
Overall	**2516**	**31.8 ± 0.4#**	**30**	**14 - 71**	**25 - 37**

Note: * Rounded to the nearest integer
Rounded to the nearest decimal place.

4.3.3 Age of Population Employed at Radium Hill

The average age of the cohort employed at RH was computed using the duration of employment as weights. The duration weighted average age of the cohort of workers was 32.6 (range: 13 - 70) years for males and 27.7 years for females.

Table 4.3.3: Age Distribution of Working Cohort at Radium Hill (For Males Only).

Year	N	Mean (years)	Median (years)	Range (years)	I-Q Range (years)
1948	5	34	36	18-43	30-42
1949	8	35	37	19-44	29-43
1950	43	34	29	20-61	22-44
1951	123	33	28	13-63	24-43
1952	356	33	29	14-64	25-40
1953	572	34	31	15-65	26-40
1954	513	34	30	16-70	25-39
1955	741	32	29	16-67	24-37
1956	614	32	29	14-68	24-37
1957	514	33	30	15-69	25-36
1958	576	32	30	14-69	25-35
1959	596	32	30	14-62	24-36
1960	616	32	30	14-63	24-36
1961	448	33	31	14-64	26-36
1962	35	39	36	16-64	31-47
Overall	**2516**	**32.7**	**30**	**13-70**	**25-37**

4.4 Radium Hill Work and Radon Progeny Exposure Histories

4.4.1 Categories of Employment at Radium Hill

Categories of employment at the RH mine are listed in tables 4.4.1a and 4.4.1b according to their exposure status. Employment categories which only involved work on the surface where workers were not exposed to Rn above background levels are referred to as job categories *unexposed* to Rn. For each of these job categories unexposed to Rn, the cumulative duration of employment (in person years), the total number of episodes of employment and the mean age of workers (obtained as a weighted average of duration of employment and age), are also summarized in table 4.4.1a. Results showed that the greatest cumulative duration of employment in surface job categories constituted work by fitters and turners (159 person years) and tradesmen's assistants (131 person years) which together contributed 13% of the total duration of employment above ground; on an average, workers in these job categories tended to be younger than other surface workers (mean age: 34 years).

Employment categories which involved work underground with exposure to Rn are referred to as job categories *exposed* to Rn. All reference of *exposures* are made only to the male cohort of workers. Table 4.4.1b summarizes the cumulative duration of employment (in person years), the cumulative exposure to Rn progeny measured in working level months (in WLM), the number of episodes of employment and the employment duration weighted mean age of workers in each of the job categories exposed to Rn. Job categories contributing the highest employment at RH were Miners and Sinkers (581 person years of employment), and under-ground Labourers and Pumpmen (315 person years of cumulative employment). Miners and Sinkers contributed 39% of the total person years of employment under-ground and 42% of the cumulative exposure to Rn progeny experienced at RH; their average age was 34 years. Underground Labourers and Pumpmen contributed 21% of the total person years of employment under-ground and 21% of the cumulative Rn progeny exposure experienced at RH; they were aged 31 years, on an average.

Table 4.4.1a: Radium Hill Job Categories Unexposed to Radon Progeny

Job Description		Cumulative Employment			Weighted
Code	Category	Duration		Episodes	Mean Age*
		(Years)	%	N	Years
01	Motor Mechanic	38	2%	83	38
02	Carpenter	84	4%	146	38
03	Builder's Labourer	93	4%	232	40
04	Welder, Spec. Welder	45	2%	99	33
05	Fitter, Fitter & Turner	159	7%	286	34
06	Fitter Operator	1	+	5	48
07	Electrical Fitter, Radio T/Man	34	2%	81	33
08	Electrician, Linesman	44	2%	87	34
09	Pipe Fitter	10	+	37	35
11	Plumber	22	1%	42	42
12	Painter	34	2%	67	39
13	Maintenance Worker	6	+	14	41
15	Blacksmith, Metal Worker	35	2%	58	37
16	Blacksmith's Assistant	4	+	10	31
17	Sampler, Radiometric Tech.	36	2%	103	40
18	Surface Sampler	26	1%	107	36
19	Cook	74	3%	154	45
20	Cook's Offsider, Mess Sup.	75	3%	197	47
22	Owner Driver	4	+	9	31
23	Driver	27	1%	71	35
24	DMT Driver	16	1%	39	37
25	Grader, Tractor/HMV Driver	26	1%	66	35
26	Bus Driver	7	+	20	38
27	Engine Driver	30	1%	57	42
28	Barman	17	1%	31	45
29	Storeman	69	3%	158	44
30	Charge Storeman	35	2%	68	45
31	Store Assistant	3	+	5	54
32	First Aid attendant	18	1%	34	41
33	Night Watchman	4	+	6	42
34	Butcher	4	+	10	49
35	Camp Attendant	3	+	11	63
36	General Labourer, Handyman	89	4%	343	40
37	Construction Worker	8	+	37	42
38	Crane Driver	12	1%	29	39
39	Tradesman's Assistant	131	6%	340	34

Note: + Less than 1%
* Duration Weighted Mean Age

Table 4.4.1a: Radium Hill Job Categories Unexposed to Radon Progeny (Continued)

Job Description		Cumulative Employment			Weighted
Code	Category	Duration		Episodes	Mean Age*
		(Years)	%		Years
40	Mill Sampler	41	2%	214	35
41	Flotation Operator	46	2%	90	36
42	Fitter Operator & Attendant	27	1%	108	31
43	Heavy Media Operator	30	1%	66	35
44	Bituminous Operator	2	+	13	37
45	Crusher Operator	10	+	27	48
46	Crusher Attendant	14	1%	34	44
47	Ball Mill Operator	26	1%	89	34
48	Mill Labourer or Hand	5	+	40	37
49	Youth Labourer or Apprentice	41	2%	112	19
50	Storeman Improver	7	+	16	22
51	Winding Engine Driver	36	2%	71	43
57	Lamp Room Attendant	13	1%	24	40
58	Change Room Attendant	18	1%	31	56
61	Wood Machinist	4	+	5	38
66	Braceman	32	1%	95	33
75	Shop Assistant	5	+	11	29
76	Shift Foreman - Surface	8	+	19	46
77	Motor Lorry Driver	20	1%	58	38
80	Machinist & Turner	17	1%	33	49
81	Lamproom Electrical Fitter	+	+	1	27
83	Tool Storeman	6	+	12	64
84	Diesel Fitter	7	+	15	38
86	Block Attendant	3	+	6	66
87	Brick Layer	3	+	9	43
89	Plant Operator	2	+	10	34
90	Magnetic Separator Operator	1	+	9	36
91	Canvas Worker	2	+	2	34
92	Rigger	5	+	12	43
93	Mullocker	2	+	9	30
97	Transferred to Mines Dept.	26	1%	33	39
98	Inactive	401	18%	875	37
	Overall	**2184**	**100%**	**5291**	

Note: + Less than 1%
* Duration Weighted Mean Age

Table 4.4.1b: Radium Hill Job Categories Exposed to Radon Progeny

Job Description		Radon Progeny Exposure at Radium Hill					Mean
Code	Category	Duration		Quantity		Episodes	Age*
		(Years)	%	WLM	%		Years
10	U/G Pipe Fitter	20	1%	129	1%	52	30
14	Greaser	7	+	41	+	25	34
52	U/G Air Hoist Operator	44	3%	367	3%	139	39
53	Miner & Sinker	581	39%	4462	42%	1202	34
54	Trucker	83	6%	360	3%	219	30
55	U/G Mine Labourer, Pumpman	315	21%	2250	21%	1365	31
56	Driller, Assistant	14	1%	300	3%	39	39
59	U/G Chainman	30	2%	114	1%	63	30
60	U/G Timberman	65	4%	352	3%	147	35
62	U/G Storeman	47	3%	208	2%	84	35
63	Loco Operator	28	2%	111	1%	77	30
64	Loader Operator	41	3%	176	2%	113	32
65	Platman	22	1%	124	1%	58	35
67	Skipman	11	1%	32	+	23	37
68	Platelayer	24	2%	125	1%	62	38
69	U/G Sampler	27	2%	94	1%	62	28
70	U/G Sanitory Man	10	1%	47	+	26	48
71	Snapman	19	1%	310	3%	70	38
72	U/G Shift Boss, Trainee	20	1%	232	2%	43	50
73	Winchman	24	2%	273	3%	69	48
74	U/G Fitter	29	2%	384	4%	70	36
78	Winch Driver	3	+	60	1%	15	47
79	U/G Machinist	2	+	22	+	5	34
85	Scraper Operator	4	+	21	+	17	33
88	Mine Foreman	11	1%	110	1%	18	46
	Overall	**1480**	**100%**	**10703**	**100%**	**4063**	

Note: + Less than 1%

* Duration Weighted Mean Age

4.4.2 Changes in Job Categories at Radium Hill

Pay-roll information showed that many of the RH employees had worked in multiple jobs during their employment. The maximum number of changes in job categories was 19 amongst males workers and 4 amongst female workers. Job changes were more common for male workers with 52% working in more than one job category - this included 498 (20%), 292 (12%) and 180 (7%) who changed job category two, three or more times. Those having held just one job at RH comprised 1,220 males and 42 females workers. Only 21% of the female workers held more than one job at the mine.

In addition to career enhancement, job changes partly reflect the variation in the emphasis of mining activities. In addition, anecdotal evidence suggests that some amount of job rotation may have occurred as a means of limiting individual exposure to Rn. Surface workers had fewer job changes than underground workers. Unexposed workers held 2 jobs on an average whilst exposed workers held 3 jobs.

4.4.3 Exposure to Radon Progeny at Radium Hill

A total of 1,459 (59%) male workers experienced some exposure to Rn progeny at RH by being employed in exposed job categories. Ages at hire of exposed workers ranged from 13 to 70 years; average age at hire was 30 years (standard deviation: 8 years) and half the male workers were employed before the age of 28.

On the whole, the average age at hire amongst the 1,062 (41%) unexposed workers was higher than those who were exposed. Ages at hire for unexposed workers ranged from 14 to 63 years and averaged 32 years (standard deviation: 10 years); median age at hire was 30 years.

Compared to those unexposed, exposed workers had a longer duration of employment. The duration of employment amongst the unexposed workers ranged from a fortnight to 126 months (10.5 years) and averaged 16 months (standard deviation: 21 months). Half the unexposed were employed for 7 months or less, with half of them being employed for 3 months or less; only one quarter of the unexposed male population was employed for periods above 19 months (inter quartile range 3 - 19 months). The duration of employment for the exposed workers ranged from 1 month to 138 months (11.5 years). The mean duration was 18 months (standard deviation: 24 months), median 7 months and I-Q range 3-24 months.

The mean duration of exposure to Rn progeny at RH was 12 months. Duration of exposure for individuals varied considerably, ranging from 10 days to 9 years, with the median being 5 months and an inter-quartile range of 2 to 14 months. The mean cumulative exposure to Rn progeny at RH was 7.34 WLM and the median exposure was 3 WLM; individual cumulative exposures ranged from 0.06 - 111.59 WLM and the inter-quartile range was 1.0 to 8.0 WLM. The average concentration of Rn progeny exposure experienced at RH was 0.83 WL; individual exposure concentrations ranged from 0.14 to 2.14 WL with a median of 0.50 WL. The person-year weighted mean cumulative exposure to Rn progeny was 7.7 WLM and the person-year weighted mean concentration of exposure was 0.88 WL. Person-year weighted means will be used in all further analytical work, including the risk evaluation presented in chapters 5 and 6 of this thesis.

Distributions of the cumulative duration of exposure and the cumulative levels of exposure to Rn progeny experienced by workers at RH are represented in figures 4.4.3a and 4.4.3b.

Figure 4.4.3a: Distribution of Cumulative Duration of Exposure (Males Only)

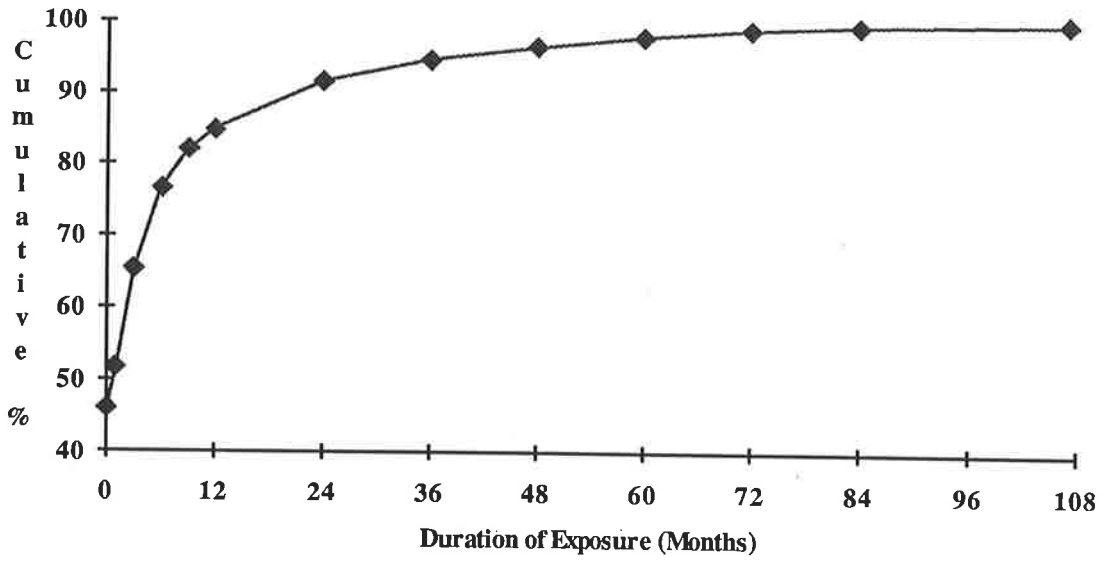
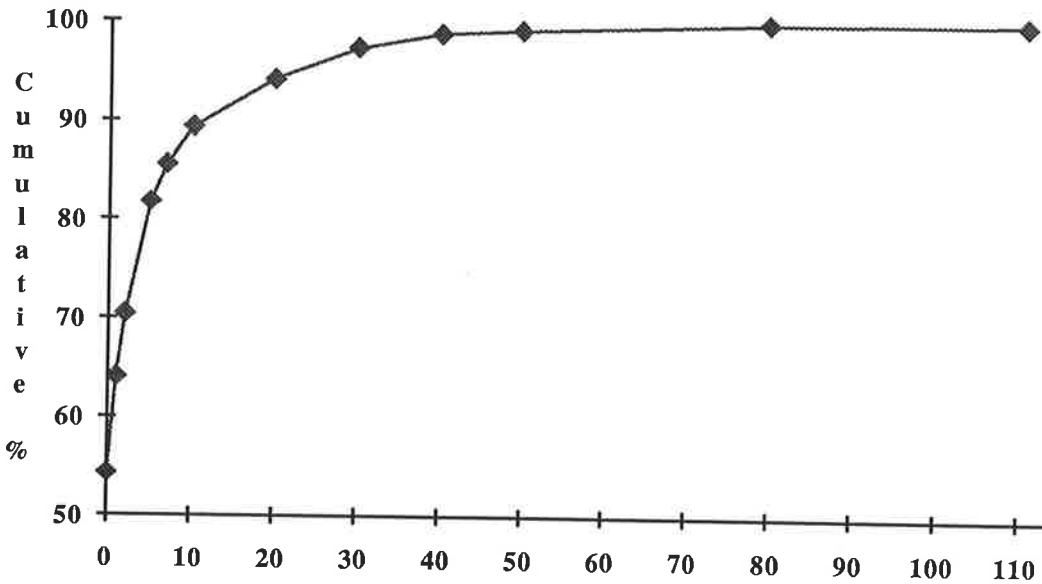


Figure 4.4.3b: Distribution of Cumulative Exposures (Males Only)



4.5 Other Occupational Exposures

The questionnaire survey which formed part of the Radium Hill study provided a means of collecting data on other occupational exposures of possible relevance to the outcome (lung cancer) studied. Questions on other occupational exposures comprised specific queries on exposure to radioactive material and asbestos. In many of the self administered questionnaires, these questions were either left blank or somewhat incomplete. For the purposes of this work, all responses were broadly classified on the basis of affirmative responses and non-responses were taken as indicating a negative response.

Findings from sections of the questionnaire survey on other relevant occupational exposures are summarized in tables 4.5a-f. Questions on other occupational exposure to radioactive materials included specific queries on exposures prior to (table 4.5a) and following (table 4.5b) employment at Radium Hill. These questions were combined to form a new composite variable on *any other occupational exposure to radioactive material* (table 4.5c). Questions on occupational exposures to asbestos were structured to cover responses on exposure to asbestos mining and milling (table 4.5d) and any other occupational exposure to asbestos (table 4.5e). These two questions were combined to form another composite variable representing *any occupational exposure to asbestos* (table 4.5f). Finally the two composite variables on other occupational exposure to radioactive material and any asbestos exposure were combined to create a new variable representing *any other relevant occupational exposure* (table 4.5g).

A total of 169 cohort members - constituting 12% of the respondents to this query - had worked with radioactive material other than at RH (table 4.5c); 57 individuals worked with radioactive material prior to their employment at RH and 128 subsequent to their employment at RH (tables 4.5a-b). Occupational exposure to radioactive material outside of RH was equally prevalent amongst those exposed and unexposed to Rn progeny at RH.

Table 4.5a: Occupational Exposure to Radiation Prior to Employment at Radium Hill

Prior	Radon Progeny Exposure at Radium Hill					
Exposure	Unexposed		Exposed		Total	
Yes	21	4%	36	4%	57	(4%)
No	516	96%	830	96%	1346	(96%)
Total	**537**	**100%**	**830**	**100%**	**1403**	**100%**

Table 4.5b: Occupational Exposure to Radiation After Employment at Radium Hill

Subsequent	Radon Progeny Exposure at Radium Hill					
Exposure	Unexposed		Exposed		Total	
Yes	57	12%	71	10%	128	11%
No	425	88%	628	90%	1053	89%
Total	**482**	**100%**	**699**	**100%**	**1181**	**100%**

Table 4.5c: Any Occupational Exposure to Radiation Apart From Radium Hill

Any Other	Radon Progeny Exposure at Radium Hill					
Exposure	Unexposed		Exposed		Total	
Yes	68	13%	101	12%	169	12%
No	470	87%	765	88%	1235	88%
Total	**538**	**100%**	**866**	**100%**	**1404**	**100%**

Table 4.5d: Occupational Exposure to Asbestos Mining or Milling

Asbestos	Exposure to Radon Progeny at Radium Hill					
Exposure	Unexposed		Exposed		Total	
Yes	24	5%	100	13%	124	10%
No	464	95%	644	87%	1108	90%
Total	**488**	**100%**	**744**	**100%**	**1181**	**100%**

Table 4.5e: Other Occupational Exposure to Asbestos

Asbestos	Exposure to Radon Progeny at Radium Hill					
Exposure	Unexposed		Exposed		Total	
Yes	38	13%	96	18%	134	16%
No	258	87%	441	82%	699	84%
Total	**296**	**100%**	**537**	**100%**	**833**	**100%**

Table 4.5f: Any Occupational Exposure to Asbestos (Mining, Milling and Other Asbestos)

Asbestos	Exposure to Radon Progeny at Radium Hill					
Exposure	Unexposed		Exposed		Total	
Yes	42	9%	122	16%	164	13%
No	446	91%	622	84%	1068	87%
Total	**488**	**100%**	**744**	**100%**	**1232**	**100%**

Table 4.5g: Any Relevant Occupational Exposure Other Than at Radium Hill (Radioactive Material or Asbestos)

Other	Exposure to Radon at Radium Hill					
Exposure	Unexposed		Exposed		Total	
Yes	102	19%	205	24%	307	22%
No	437	81%	655	76%	1102	78%
Total	**488**	**100%**	**870**	**100%**	**1409**	**100%**

Some occupational exposure to asbestos was reported by 164 individuals who constituted 13% of the respondents to this question (table 4.5f). A total of 124 individuals reported having worked in asbestos mining or milling and 134 reported other occupational exposure to asbestos, respectively constituting 10% and 16% of the respondents to these questions (tables 4.5d-e). The prevalence of occupational exposure to asbestos was higher amongst those who had been exposed to Rn at RH. This increase in prevalence of occupational exposure to asbestos mining and milling amongst underground workers at RH indicates the possibility that these workers may have been professional miners who may have moved from asbestos mining to uranium mining at RH or worked in asbestos mines after their employment in RH. Several of these individuals were identified by secondary trace activities in Western Australia where asbestos was mined (i.e., through record linkage with records maintained at the Perth Chest Clinic and with the nominal roll of the Wittenoom Gorge cohort study of asbestos miners, outlined in chapter 3).

Results summarized in table 4.5g show that 370 male workers from the RH cohort were occupationally exposed to other radioactive materials or asbestos; these individuals comprised 22% of the respondents to the questionnaire survey component on other relevant occupational exposure. The prevalence of other relevant occupational exposures was higher amongst those who had been exposed to Rn progeny at RH (24% amongst those exposed versus 19% amongst those unexposed).

4.6 Other Relevant Exposures: Smoking

Some information on smoking habits was obtained for 1,320 (52%) males and 42 (79%) females. Those who ever smoked - *ever smokers* - comprised 1,121 (85%) of the males and 25 (60%) of the females.

Amongst male cohort members some smoking information was available on 54% (787 out of 1,457) of those exposed and 50% (533 out of 1,062) of those unexposed. Sixty-one percent of the smokers were exposed to some Rn (679 out of 1,121).

The prevalence of smoking amongst the RH cohort is summarized in table 4.6a according to Rn progeny exposure status at RH. The prevalence of smokers was 86% amongst those ever exposed to Rn progeny at RH and 83% amongst those never exposed to Rn progeny at RH. The overall prevalence of smoking amongst the study cohort was 85%.

Table 4.6a: Prevalence of Smoking By Exposure to Radon at Radium Hill (For Males Only; Excluding Workers With Missing Information)

Ever Smoker	Radon Progeny Exposure at Radium Hill		
	Unexposed	Exposed	Total
Yes	442	679	1121
	(83%)	(86%)	(85%)
No	91	108	199
	(17%)	(14%)	(15%)
Total	533	787	1320
	(100%)	(100%)	(100%)

Detailed smoking histories were available in varying degrees of completion for 534 males who were current smokers and 433 ex-smokers at the time of the questionnaire survey. Quantitative estimates of smoking characteristics were sought from these individuals (table 4.6b).

Table 4.6b: Quantitative Estimates of Smoking Characteristics for Male Respondents

Smoking Characteristic		Radon Progeny Exposure at Radium Hill		
		Unexposed	Exposed	All
Current Cigarette Smokers:				
Age Started Smoking	Mean ± 1.96SE	19 ± 1	18 ± 0.5	18 ± 0.5
	N	186	324	510
Duration Smoked Regularly	Mean ± 1.96SE	37 ± 2	36 ± 1	37 ± 1
	N	188	328	516
Average Quantity Smoked / Day (Currently)	Mean ± 1.96SE	24 ± 2	23 ± 1	24 ± 1
	N	185	323	508
Average Quantity Smoked / Day (At Radium Hill)	Mean ± 1.96SE	18 ± 2	18 ± 1	17 ± 2
	N	183	317	500
Ex-Cigarette Smokers:				
Age Started Smoking	Mean ± 1.96SE	19 ± 1	19 ± 1	19 + 0.5
	N	184	249	433
Duration Smoked Regularly	Mean ± 1.96SE	25 ± 2	26 ± 2	26 ± 1
	N	184	249	433
Average Quantity Smoked / Day	Mean ± 1.96SE	21 ± 2	22 ± 2	22 ± 1
	N	185	246	431
Average Quantity Smoked / Day (At Radium Hill)	Mean ± 1.96SE	15 ± 2	18 ± 2	17 ± 1
	N	183	246	429
Other Tobacco Products:				
Cigars:				
Age Started Smoking	Mean ± 1.96SE	30 ± 3	27 ± 2	28 ± 2
	N	59	77	136
Pipes:				
Age Started Smoking	Mean ± 1.96SE	26 ± 4	27 ± 3	27 ± 3
	N	29	39	68
Estimated Quantity Smoked: (Obtained from Proxy Sources)				
Cigarettes	Mean ± 1.96SE	27 ± 4	29 ± 3	28 ± 2
	N	27	18	45
Perth Chest Clinic Estimates:				
Ounces / Week	Mean ± 1.96SE	1.75 ± 0.4	2 ± 0.3	1.9 ± 0.3
	N	14	36	50

Note: SE - Standard Error; N - Number of Respondents.

The reported average number of cigarettes smoked by respondents during their employment at RH was considerably lower than their current average consumption or the average quantity smoked most recently by ex-smokers. Current smokers had smoked regularly for an average of 37 years whilst ex-smokers had smoked regularly for 26 years. Both current smokers and ex-smokers commenced smoking around the ages of 18 or 19. Overall, the majority of users and ex-users of tobacco products (approximately, 64% of the current smokers and 58% of the ex-smokers) had been exposed to Rn progeny at RH.

Respondents to the questionnaire survey also included 136 cigar smokers and 68 pipe smokers who were older when commencing smoking pipes or cigars than cigarette smokers. The prevalence of individuals ever smoking cigars was 6% overall, 3% during employment at RH and 2% at the time of the questionnaire survey; the prevalence of individuals ever smoking pipes was 13% overall, 2% during employment at RH and 2% at the time of the survey (table 4.6c). Approximately 59% of pipe or cigar smokers had been exposed to Rn progeny at RH; this percentage was constant across current smokers, ever smokers and smokers at RH.

Table 4.6c: Prevalence of Other Smoking (Men Only)

Smoking Characteristic			Radon Progeny Exposure at Radium Hill			
			Unexposed	Exposed	Total	
Cigars:						
	Ever	Yes	30	43	73	(6%)
		No	448	648	1096	
		Total	**478**	**691**	**1169**	
	At RH	Yes	11	20	31	(3%)
		No	456	663	1119	
		Total	**467**	**683**	**1150**	
	Current	Yes	8	13	21	(2%)
		No	455	669	1124	
		Total	**463**	**682**	**1145**	
Pipes:						
	Ever	Yes	65	83	148	(13%)
		No	414	608	1022	
		Total	**479**	**691**	**1170**	
	At RH	Yes	11	14	25	(2%)
		No	458	666	1124	
		Total	**469**	**680**	**1149**	
	Current	Yes	11	9	20	(2%)
		No	458	670	1128	
		Total	**469**	**679**	**1148**	

4.7 Mortality Profile

A total of 615 cohort members (606 males and 9 females) died during the study period. The underlying cause of death for these individuals is summarized in table 4.7a. Circulatory disorders comprised the leading cause of death (44%) amongst male cohort members. Malignant neoplasms were the second leading cause of death among males (22%) and the leading cause of death amongst females (45%). Of the deaths due to malignant neoplasms, 54 were attributed to lung cancer; lung cancer deaths comprised 9% of all deaths and 43% of the malignant neoplasms amongst male cohort members. The average age at death was 58 years for males dying of all causes and 61 years for males dying from malignant neoplasms or circulatory disorders.

Table 4.7a: Underlying Cause of Death Amongst Radium Hill Workers

Underlying Cause of Death	Men			Women		Total	
	N	%	Mean Age	N	%	N	%
Malignant neoplasms	127	21%	61	4	45%	131	21%
Circulatory disorders	264	44%	61	2	22%	266	43%
Respiratory conditions	45	7%	65	-	-	45	7%
Injuries & poisoning	95	16%	45	2	22%	97	16%
Other causes	75	12%	58	1	1%	76	13%
All	**606**	**100%**	**58**	**9**	**100**	**615**	**100%**

Tabulations of underlying cause of death for males by their status of exposure to Rn progeny at RH (table 4.7b) showed a slightly higher proportion of malignant neoplasms amongst those exposed. A higher proportion of deaths resulting from injuries and poisonings was also observed amongst those who worked under-ground at RH compared to surface workers. Those who had worked underground at RH and died of malignant neoplasms, circulatory disorders or injuries and poisonings, died younger - than surface workers - of the same cause.

Table 4.7b: Underlying Cause of Death Amongst Males By Radon Progeny Exposure

Underlying Cause of Death	Exposure to Radon Progeny at Radium Hill					
	Unexposed			Exposed		
	N	%	Mean Age	N	%	Mean Age
Malignant neoplasms	55	20%	63	72	22%	60
Circulatory disorders	130	47%	61	134	41%	60
Respiratory conditions	26	9%	64	19	6%	65
Injuries & poisoning	34	12%	48	61	19%	43
Other causes	34	12%	58	41	12%	59
All	**279**	**100%**	**60**	**327**	**100%**	**57**

Data on smoking characteristics were available for 228 deceased male cohort members of whom only 14 (6%) had never smoked tobacco products. The proportion of smokers dying of malignant neoplasms (24%) was higher than the overall proportion of males dying from malignant neoplasms (21%). These findings are summarized in table 4.7c.

Table 4.7c: Underlying Cause of Death Amongst Males By Smoking Status

Underlying Cause of Death	Smoker			Non Smoker		
	N	%	Mean Age	N	%	Mean Age
Malignant neoplasms	52	24%	63	2	14%	64
Circulatory disorders	90	42%	63	7	50%	63
Respiratory conditions	21	10%	66	-	-	-
Injuries & poisoning	23	11%	45	1	7%	35
Other causes	28	13%	64	4	29%	54
All	**214**	**100%**	**61**	**14**	**100%**	**60**

4.8 Summary of Basic Demographic Characteristics of the Radium Hill Study Cohort

Basic demographic characteristics of the RH study cohort are summarized in table 4.8.1.

Table 4.8: Demographic Profile of Male Cohort

Characteristic	ALL	Trace Status		Exposure Status	
		Untraced	Traced	Unexposed	Exposed
Mean Year of Birth	1925	1927	1924	1923	1926
Mean Age at:					
Start of Employment (Years)	31	30	31	32	30
End of Employment (Years)	32	31	33	34	31
End of Follow-up (Years)	50	-	58	51	50
Mean Duration of:					
Employment (Months)	17	12	19	16	19
Exposure (Months)	7	5	8	-	12
Follow-up (Years)	19.5	1	26.3	19	20
Mean Rn Exposure:					
Mean (WLM)	4.2	3.0	4.7	-	7.3 (7.7)^{W}
Intensity (WL)	0.83	0.79	0.84	-	0.83 (0.88)^{W}

Note: ^{W} Person-Year Weighted Mean

Inferences of particular relevance to the RH study, drawn from the characteristics summarized in table 4.8.1 include:

- on average, members of the study cohort commenced employment at the age of 31, completed their employment at the age of 32 and were aged 50 at the end of follow-up. The average age at commencement of employment at RH was slightly higher for surface workers than for under-ground workers.

- the mean duration of follow-up and age at end of follow-up for those in the traced column of table 4.8.1 was higher than the corresponding values in other columns which included untraced workers who had extremely short durations of follow-up accumulated during their employment at RH.

- compared to traced workers, untraced workers in the RH study cohort tended to have shorter durations of employment, shorter durations of exposure and to have experienced lower cumulative exposures to Rn progeny at lower concentrations at RH.

Other major descriptive epidemiological findings from this chapter include:

- The highest cumulative exposures to Rn progeny experienced in RH (60%), were by Miners, Sinkers, Under-Ground Labourers and Pumpmen. Workers employed in these categories also contributed the longest cumulative duration (person years) under-ground. Amongst job surface job categories, the longest cumulative duration was worked by Fitters, Turners and Tradesmans' Assistants. Job rotations were common at RH; the average numbers of jobs worked in were 3 for under-ground workers and 2 for surface workers.

- Fifty nine percent of the male workforce at RH experienced some exposure to Rn progeny at RH. Approximately 90% of all male workers at RH were exposed to less than 10 WLMs of Rn progeny exposure.

- Of the respondents to the questionnaire survey queries on other occupational exposure, 12% reported occupational exposed to radioactive material other than at RH, and 13% reported occupational exposure to asbestos. Overall, 307 of the respondents (22%) had experienced some occupational exposure - radioactive material and/or asbestos - relevant the risk of lung cancer mortality, other than at RH; two-thirds of these workers had been exposed to Rn progeny at RH.

- Eighty five percent of the respondents to the questionnaire survey queries on smoking reported having smoked sometime in their lives; the prevalence of *ever smoking* was 83% amongst surface workers and 86% amongst under-ground workers. The average cigarette consumption for smokers was estimated as: 22 to 24 cigarettes/day most recently, and 17 cigarettes/day while working at RH.

- A total of 606 male workers in the study cohort died during the period of the study. The leading causes of death amongst them were circulatory disorders or malignant neoplasms. Twenty-two percent of the deaths amongst the male cohort were due to malignant neoplasms; these included 54 lung cancer deaths which comprised 9% of all deaths and 43% of the deaths from malignant neoplasms. The proportion of deaths from malignant neoplasms was slightly higher amongst those who had been exposed to Rn progeny at RH (22%) compared to surface workers at RH (20%). The proportion of deaths from malignant neoplasms was observed to be higher (24%) amongst workers who had smoked sometime during their lives, compared to the 21% observed amongst all deceased males. Overall, workers exposed to Rn progeny at RH died younger (mean 57 years) than surface workers (mean 60 years). Similarly, the mean age at death from malignant neoplasms was 60 years amongst exposed workers and 63 years amongst surface workers at RH.

CHAPTER 5

5. Identification and Control of Confounders

Tables and Figures

5. Identification and Control for Confounders

5.1 Aims

The aims of the analyses presented in this chapter are as follows:

1. To obtain crude estimates of lung cancer mortality amongst the RH cohort.
2. To examine the roles of potential confounders of the Rn related lung cancer mortality effect in the RH cohort.
3. To obtain estimates of Rn related lung cancer risk amongst the RH cohort after controlling for confounders using methods based on internal and external references.
4. To examine the efficacy of different methods of estimating lung cancer risk based on various internal and external references.

All analytical findings presented in this chapter are based only on the male cohort of RH workers. The cohort analysed comprised 2,516 male workers. Five other males for whom birth dates were unavailable were excluded from this analyses.

5.2 Methodological Overview

5.2.1 Definition of Radon Daughter Exposure Variables

For the purposes of analyses, the study of all exposure effects was based on cumulative *'relevant exposures'* (hereafter referred to as cumulative exposure or CE). Here, I define *relevant exposure* as exposure accumulated during the period of time relevant to the outcome - lung cancer mortality. Relevant exposure therefore refers to the amount of exposure that could plausibly contribute to the biological outcome being studied. This implies that there exists a *lag period* during which exposures do not contribute to the risk of disease at that given point of time. All analyses in this chapter will be based on the simple assumption of relevant exposure being defined as cumulative exposures pertaining to a five year lag period. This choice of five year lagged exposures is made on the assumption that *"increments in exposure have no substantial effect on the risk of lung cancer mortality for at least five years"* (BEIR IV 1988). In later sections of this thesis, the more complex assumptions where relevant exposure is defined under multiple exposure lag windows known as *time since exposure (TSE) windows* will be introduced and examined.

Exposure categorization was performed in two stages. Firstly, workers were broadly classified as those *unexposed* and *exposed* to *relevant* Rn *exposure* at the RH mine (hereafter, referred to as unexposed and exposed). In the second stage, workers who received any relevant Rn progeny exposure at RH were further categorized into finer exposure groups viz., 0<-1, 1<-10, 10<-20, 20<-40 WML. The choice of exposure categories could either have been arbitrary or been based on biological plausibility and statistical sensibility. Exposure categorization adopted here was made after reviewing contemporary radon literature and extensive examination of the data. Further details of the basis of this categorization are provided in section 5.2.2.

5.2.2 Construction of Analyses Dataset: PYRS and Event Tables

All analyses presented in this chapter are based on *person years at risk (*PYRS*)* data compiled categorically as PYRS and event tables. These tables were specified in multiple dimensions. Each dimension was defined by a basic variable of interest including, potential confounders or modifiers of Rn progeny exposure effect. Variables of interest as potential confounders were age and calendar year of observation. Other covariates were age at first exposure, time since last exposure, duration of exposure and intensity of exposure. These variables were categorized according to the same principles of categorization as outlined in section 5.2.1 for categorization of the cumulative exposure variable.

To facilitate trend modelling, averages of the various continuous variables that were categorized were required within each of their categories. Since these variables were not symmetrically distributed within each of their categories, the mid-point of each category could not be used as a suitable average. Therefore, person year weighted averages were used to quantitatively represent each category; i.e. as the appropriate *class mark* of each *class interval*. For example, the distribution of cumulative exposures was downwardly skewed within each category; PYRS weighted cumulative exposures were therefore used as the appropriate average representing each exposure category.

PYRS and event tables were constructed by tabulating this dataset over the dimensions described above and summing PYRS and events. Each quantitative variable that was categorized to form a table dimension was then assigned its the PYRS weighted mean as the *class mark* within each of its *class intervals*.

In formulating categories of cumulative exposure, efforts were made to maintain adequate numbers of PYRS and events in each exposure category whilst allowing finer classification at lower levels of exposure to enable the examination of Rn effect at lower levels of cumulative exposures.

5.2.3 Analytical Methods

Poisson regression modelling techniques were used for all PYRS based analyses presented in this work (Frome and Checkoway 1985; Pearce and Checkoway 1987). This section describes the specific models used to estimate risk, after identifying and controlling for confounders. Relevant epidemiological background underlying the concept of confounding and methods of identifying and controlling confounding are summarized in appendix C.

Risk estimation was based on *relative risk* (RR) models defined below (models 1 - 3). Unless otherwise specified, multiplicative log-linear models (model 1) were used for RR evaluation when exposure was treated as a categorical variable. Variations in *excess relative risk* (ERR) were modelled using the *linear additive excess relative risk model* (model 2) and a time-since-exposure model (model 3). These models are defined as follows:

Multiplicative Log-Linear RR Models: $RR = e^{\beta w}$ (1)

where, w - is a measure of Rn progeny exposure

β - measures the RR of exposure

Linear Additive ERR Models: $RR = [1 + \beta w]$ (2)

Time-Since-Exposure Model: $RR = [1 + \beta_1 w_{5\text{-}15} + \beta_2 w_{15\text{-}}]$ (3)

where, β - measures the ERR/WLM

$w_{5\text{-}15}$ - measures Rn progeny exposures accumulated 5-15 years previously

$w_{15\text{-}}$ - measure Rn progeny exposures accumulated prior to 15 years before

These models were fitted under the assumption that the cell specific mortality rates - based on PYRS denominators - were Poisson distributed. Poisson regression models were fitted using maximum likelihood estimation methods. Maximum likelihood estimates (MLE) of parameters, likelihood based confidence intervals and model deviances were obtained for all models fitted. Overall goodness of fit of models was gauged from model deviances and their degrees of freedom. Nested models were compared using likelihood ratio tests (LRT).

5.2.4 Analyses Plan

Statistical assessment of the frequency and risk of lung cancer mortality amongst the Radium Hill cohort commenced with the application of the simplest methods of risk assessment and then progressed to the application of analytical methods of increasing degrees of complexity. A crude estimate of mortality were first obtained for the entire male cohort (Rothman 1986). Crude estimates of mortality rates and effect were then obtained for various categories of exposure to Rn progeny and rate ratios were used as measures of effect. Results of the analyses of crude data are presented in section 5.3 below.

The analysis of crude data was followed by the simplest method of controlling for potential confounders - stratified analyses. Stratified analyses were used to examine the confounding effect of age and calendar time. Findings from the stratified analyses of the RH data are presented in section 5.4 of this chapter.

Stratified analyses were followed by analyses based on internal adjustment for confounders. Internal adjustment for confounders was performed using the simplest Poisson regression models in EPICURE, based on PYRS denominators with the confounders being treated as stratification variables (EPICURE 1992). Summary estimates of disease frequency and effect were obtained in turn whilst controlling for confounders individually and then simultaneously. These analyses were then extended to evaluate the comparative risk of Rn progeny exposure and lung cancer mortality amongst various categories of exposure described above. Results of internal adjustment for confounding are reported in section 5.5.

Internal adjustment for confounders was followed by methods of external adjustment for confounders. External adjustment for confounders was based on standardization techniques using the Australian National Population as the source of external reference.

Standardized analyses were performed using both indirect and direct methods of standardization. Indirect standardization was based on Poisson regression modelling in EPICURE. Standardized Mortality Rates (SMR) were obtained as estimates of disease frequency. Age and calendar year specific SMR were examined for any residual heterogeneity after adjustment to the external reference population. SMR ratios were then used to obtain estimates of the RR of exposure after controlling for the confounding effects of age and calendar year through indirect standardization for age and calendar year. Estimates of RR were first obtained for the overall cohort and then, in turn for the various exposure categories, using log-linear models in EPICURE and maximum likelihood evaluation techniques. Results of indirect standardization are presented in section 5.6.

Indirect standardization was followed by the application of direct standardization methods. Directly standardized rates (DSR) and estimates of risk based on the Comparative Mortality Figure (CMF) were obtained by using simple programming algorithms based on methods outlined by Breslow and Day (Breslow and Day 1987). The results of these analyses are presented in sections 5.7.

In each section of the analyses, tests for homogeneity and trend were performed using likelihood ratio tests (LRT) and score tests for trend in EPICURE. Estimates of ERR per unit exposure were obtained by using linear excess relative risk models defined in section 5.2.3 (model 2). Effects of departures from linearity were tested by comparison with models with non-linearity parameters. Unless otherwise specified, categorical modelling was based on log-linear models.

The basic concepts underlying methods of stratification, standardization and statistical modelling are summarized in appendix C. A general description of modelling with EPICURE - as applied to this work - is provided in appendix D.

5.3 Crude Estimates of Lung Cancer Mortality Rates and Risk

Amongst the 2,521 males in the study population, 606 were known to have died by the end of 1987. These deaths included 54 lung cancer deaths which comprised 9% of all deaths and 43% of deaths from malignant neoplasms. The total duration of follow-up in person-time was 49,240.59 person-years.

The crude lung cancer mortality rate amongst the study cohort was 1,097 per 10^6 person-years-at-risk which lay within a 95% confidence interval of 840 - 1,432 per 10^6 person-years-at-risk. Based on relevant exposures defined with a 5 year lag, the overall person year weighted mean cumulative Rn progeny exposure level was 3.68 WML. Estimates of crude *excess relative risk* (ERR) of lung cancer mortality per unit exposure were obtained by Poisson regression modelling of person-year weighted mean cumulative exposure in each cell of the categorically grouped data. By fitting a linear ERR model, a crude ERR effect of 8% per WML exposure was estimated within a 95% CI of 0.02% - 16% per WML. This crude estimate of ERR per WML implied a doubling of risk at a exposure of 12.5 WML.

When exposure was modelled as a log linear function of lung cancer mortality the ERR increased by 3.7% per WML within a 95% CI of 2.2% - 5.3%. The effect of exposure was found to be significant in both the linear ERR and the log ERR models (p = 0.0004 and p = 0.0001). Though these are not nested models and cannot therefore be compared directly, the log-linear model had a slightly smaller deviance than the linear model. Discussion of the appropriateness of these two models is delayed till the concluding section of this chapter. Statistically significant trends were also detected in the crude exposure-response relationship in both models using score tests for trend ($p < 0.05$).

Exposures to radon progeny at RH was seen to have a significant effect on crude lung cancer mortality at the 10% level of significance (p = 0.0942). This result was obtained from a likelihood ratio (LR) test based on modelling relevant exposure as a dichotomous variable using Poisson regression techniques. These results are given at the foot of table 5.3a.

Table 5.3a: Crude Estimates of Lung Cancer Mortality

EXPOSURE GROUP	MEAN (*)	PYRS (&)	CASES	RATE (#)	95% CI (FOR RATE) (#)	RR (@)	95% CI (FOR RR) (@)
Overall	3.68	49241	54	1097	840 -1432		
Unexposed	0.00	25657	22	858	565 - 1302	1	Fixed
Exposed	7.69	23584	32	1358	960 -1919	1.59	.92 - 2.72

NOTE: * PYRS Weighted; & rounded to the nearest integer;
Per 10^6 PYRS; @ rounded to the nearest decimal
LR test for Exposure Effect:
LR Statistic = 0.801; DF = 1; P = 0.0942

The crude mortality rate amongst those unexposed was 858 (95% CI 565 - 1,302) per 10^6 person years at risk, and 1358 (95% CI 960 - 1,919) amongst those exposed to radiation at RH. This difference in crude mortality rates amongst those unexposed and exposed was not statistically significant (p = 0.5). The crude RR of exposure was 1.59 which was found not to be significantly different from a RR of 1; the 95% confidence interval for the crude RR was 0.92 - 2.72. The person-year weighted mean exposure amongst those exposed was 7.69 WML.

Estimates of crude lung cancer mortality rates and risk by further categorizations of cumulative exposure levels are given in table 5.3b. Compared to those unexposed, significant increases were seen in the crude RR of exposure in the higher exposure categories. These increases were greater than three fold amongst those in the 20 to 40 WML group and eight fold amongst those exposed to over 40 cumulative WML. Exposure group specific estimates of crude mortality and risk were found to be significantly heterogeneous. A significant trend was also noted with increasing exposure, in the crude estimates of RR.

Table 5.3b: Crude Lung Cancer Mortality by Exposure Categories

EXPOSURE GROUP	EXPOSURE MEAN (*)	PYRS (&)	CASES	RATE (#)	95% CI (FOR RATE) (#)	RR (@)	95% CI (FOR RR) (@)
0	0.00	25657	22	858	565 - 1302	1	-
0<-1	0.59	4537	8	1763	882 - 3526	2.1	0.9 - 4.6
1<-10	3.84	13921	10	718	387 - 1335	0.84	0.4 - 1.8
10<-20	13.94	2463	4	1624	610 - 4327	1.90	0.7 - 5.5
20<-40	27.53	2080	6	2885	1296 - 6421	3.36	1.4 - 8.3
>40	57.46	583	4	6861	2575 - 18280	8.00	2.8 - 23.2

NOTE: * PYRS Weighted; & rounded to the nearest integer;
Per 10^6 PYRS; @ rounded to the nearest decimal
Test for Homogeneity of Stratum Specific Exposure Effects:
LRT Statistic 17.73; DF = 5; $p = 0.0033$

5.4 Examining the Role of Basic Confounders Through Stratification

Tables 5.4a - 5.4b show results of stratification by potential confounders.

Table 5.4a: Age Specific Lung Cancer Mortality

AGE GROUP	MEAN AGE (*)	PYRS (@)	CASES	RATE (#)	95% CI (FOR RATE) (#)	RR (@)	95% CI (FOR RR) (@)
<15	14.0	10	-	-	-	-	
15 -< 20	17.6	256	-	-	-	-	
20 -< 25	22.6	1691	-	-	-	-	
25 -< 30	27.2	3920	-	-	-	-	
30 -< 35	32.1	5588	-	-	-	-	
35 -< 40	37.0	6613	-	-	-	-	
40 -< 45	42.0	7105	1	141	20 - 999	1	Fixed
45 -< 50	47.0	7218	6	831	374 - 1850	16	2 - 133
50 -< 55	51.9	6433	7	1088	519 - 2283	21	3 - 171
55 -< 60	56.8	4858	6	1235	555 - 2749	24	3 - 198
60 -< 65	61.7	2872	15	5223	3149 - 8664	101	13 - 763
65 -< 70	66.7	1589	10	6716	3614 - 12480	130	17 - 1013
70 -< 75	71.7	737	8	10853	5428 - 21700	210	26 - 1676
75 -< 80	76.6	318	1	3148	444 - 22350	61	4 - 972
80 -< 85	81.5	109	-	-	-	-	-
> 85	86.6	24	-	-	-	-	-

NOTE: * PYRS Weighted; @ nearest integer approx.; # Per 10^6 PYRS;
χ^2 Test for Homogeneity: LR Statistic = 115.7; DF = 15; $p = 0.0000$

Table 5.4b: Calendar Year Specific Lung Cancer Mortality

YEAR	MEAN YEAR (*)	PYRS	CASES	RATE (#)	95% CI (FOR RATE) (#)	RR	95% CI (FOR RR) (#)
1948 - 1952	1951.6	365	-	-	-	-	-
1953 - 1957	1955.5	4527	-	-	-	-	-
1958 - 1962	1960.1	8170	2	245	61 - 979	1	Fixed
1963 - 1967	1965.0	8382	2	239	60 - 954	2.7	.4 - 18.8
1968 - 1972	1970.0	7924	6	757	340 - 1686	8.4	1.7 - 41.7
19 73 - 1977	1975.0	7318	13	1776	1031 - 3059	19.7	4.5 - 87.4
1978 - 1982	1980.0	6659	14	2102	1245 - 3550	23.3	5.3 - 102.7
1983 - 1987	1985.0	5897	17	2883	1792 - 4637	32.0	7.4 - 138.5

NOTE: * - PYRS Weighted; # - Per 10^6 PYRS
χ^2 Test for Homogeneity: Deviance = 47.08; DF = 7; $p = 0.0000$

Lung cancer mortality amongst the study cohort occurred between the ages of 40 and 80 years and during the calendar years of 1958 to 1987. Both attained age and calendar year specific rates were found to be heterogeneous. Lung cancer mortality was seen to increase with attained age and calendar time until the age of 75 years. Considering the end-point of lung cancer, this is an expected phenomenon in an aging population. Furthermore, calendar year of observation acts as a surrogate for several other occupational and environmental factors that may have varied over the time of observation. Hence, controlling for calendar year is a means of controlling for the possible confounding effects of these unidentified factors.

In the preceding chapter it was shown that age and calendar year were associated with Rn progeny exposure. Here we see the heterogeneity in attained age and calendar year specific rates which confirm their roles as confounders. Therefore, the effects of attained age and calendar year will be controlled for in all further analyses (also see appendix C section C.4.2; Breslow and Day 1987).

5.5 Internal Adjustment for Confounders

5.5.1 Evaluation of Continuous Exposure Effect

Assessments of ERR per unit exposure were performed prior to detailed categorical examination of confounding effects. The results of these analyses (table 5.5.1) showed that the excess RR per unit exposure changed markedly with the adjustment for the confounding effects of age and calendar year. Based on linear additive ERR models (model 2), the crude ERR was seen to increase by 8.0% for each increasing WML of exposure. After controlling for the background effect of age only, the estimated increase in ERR was 4.8% per WML. When controlled for the background effect of calendar year alone, the increase in excess RR per WML was estimated as 5.3%. After controlling for the confounding effects of both, age and calendar year, the excess RR was seen to increase by 4.3% per WML. Likelihood ratio bounds (LRB) (table 5.5.1) confirmed that these estimates of ERR/WLM were significantly greater than 0. Likelihood ratio test results showed that at each stage of adjustment, the effect of Rn progeny exposure was also highly significant. Results of score tests for trend revealed highly significant linear trends in the effect of exposure on lung cancer mortality.

Table 5.5.1: Percentage Increase in Excess RR per WML of Exposure

EFFECT	ERR/WLM (%)	95% LRB (%)	LRT P-VALUE	SCORE TEST P-VALUE
Crude	8.04	2.4 - 18.2	0.0004	0.0000
Age Adjusted	4.78	0.9 - 11.9	0.0064	0.0003
Calendar Year Adjusted	5.30	1.1 - 12.9	0.0041	0.0000
Age and Cal. Year Adjusted	4.34	0.7 - 11.1	0.0102	0.0007

Risk evaluation based on a log-linear model showed that the ERR of lung cancer mortality increased exponential by 2.6% with each increasing WLM after controlling for the confounding effects of attained age and calendar year.

5.5.2 Evaluation of Exposure Effect Using Dichotomous Exposure Categories

Estimates of the RR of exposure to Rn progeny obtained from log-linear modelling (model 1), before and after adjustment for confounders based on dichotomous classification of exposure are presented in table 5.5.2.

Table 5.5.2: RR of Lung Cancer Mortality From Radon Exposure After Internal Adjustment for Confounders

EFFECT	RR OF EXPOSURE	95% CI FOR RR	LRT P-VALUE
Crude	1.58	0.92 - 2.72	0.0942
Age Adjusted	1.36	0.79 - 2.34	0.2661
Calendar Year Adjusted	1.04	0.60 - 1.80	0.8856
Age and Cal. Year Adjusted	1.25	0.72 - 2.16	0.4344

In comparing the risk of lung cancer mortality between two broad categories of exposure - viz., those unexposed and exposed, though the RR was higher amongst those exposed even after controlling for confounders, this increase was not statistically significant. However, it was possible that the classification of all those exposed into one group could have hidden trends in risk with increasing exposure. This analysis between broad categories of exposure was therefore followed by a detailed analysis by finer exposure categories to identify and examine any hidden trends in risk with increasing exposure.

5.5.3 Evaluation of Exposure Effect Using Multiple Exposure Categories

Analyses by multiple exposure categories showed that the RR of exposure was inflated by the confounding effects of age and calendar year. When controlled for these confounding effects, crude estimates of RR were reduced within each exposure category. Though all exposure group estimates were heterogeneous, and revealed an increasing trend, only those in the highest exposure category were seen to be significantly at higher risk than those unexposed.

Table 5.5.3: Internal Adjustment for Confounders By Radon Exposure Categories

EXPOSURE CATEGORY IN WMLS	CRUDE UNADJUSTED RR (95% CI)	AGE ADJUSTED RR (95% CI)	CAL. YEAR ADJUSTED RR (95% CI)	AGE & CAL. YEAR ADJUSTED RR (95% CI)
0	1 (Fixed)	1 (Fixed)	1 (Fixed)	1 (Fixed)
0<-1	2.06 (0.02 - 4.62)	1.95 (0.87 - 4.40))	1.33 (0.59 - 3.00)	1.75 (0.77 - 3.97)
1<-10	0.84 (0.40 - 1.77)	0.73 (0.35 - 1.55)	0.55 (0.26 - 1.17)	0.66 (0.31 - 1.40)
10<-20	1.89 (0.65 - 5.50)	1.59 (0.55 - 4.62)	1.27 (0.44 - 3.70)	1.56 (0.53 - 4.56)
20<-40	3.36 (1.36 - 8.30)	2.34 (0.95 - 5.78)	2.16 (0.87 - 5.33)	2.09 (0.84 - 5.18)
>40	8.00 (2.76 - 23.22)	4.87 (1.68 - 14.13)	5.47 (1.88 - 15.88)	4.38 (1.46 - 13.16)
LRT P-VALUE	0.0033	0.0246	0.0106	0.0336

5.5.4 Evaluation of Exposure Effect Under Time Since Exposure Windows

Exposure was constructed in *'exposure windows'* corresponding to the time elapsed since exposure. These periods of elapsed time were termed *'time since exposure'* (TSE) and the constructions, *'time since exposure windows'* (TSE windows). Exposures in TSE windows were computed based on lagged cumulative exposure variables. For example, exposure accumulated in the TSE window of 5-15 years was obtained from the difference between the five year and fifteen year lagged cumulative exposures.

Various classifications of TSE windows were examined in these analyses. However, due to the sparseness of the data only one particularly broad classification of TSE windows could be feasibly modelled. This classification was similar to that adopted by the BEIR IV committee viz., two windows of 5-15 years and 15 years and over and was especially useful because it enabled comparisons with BEIR IV findings. Other classifications explored included two separate categorizations of three TSE windows: 5-10, 10-15 and 15 years or more; and 5-15, 15-25 and 25 years or more. Estimates of ERR/WLM resulting from exposures in the two TSE windows - 5-15 years and 15 year or more - obtained from a linear additive ERR model (model 3) are summarized in table 5.5.4.

Table 5.5.4: Percentage Increase in ERR/WLM Based on TSE Windows

EFFECT EVALUATED : TSE WINDOWS	ERR/WLM (%)	95% LR BOUNDS	SCORE TEST P-VALUE
5-15 years	8.5	0.0 - #	0.1096
15 years or more	3.9	0.3 - 10.0	0.0015

NOTE: # - No feasible likelihood ratio upper bound found.
All findings base on a linear additive excess relative risk model (model3)

Results presented in table 5.5.4 shows that the ERR of lung cancer mortality per WML was twice as high (8.5%) for exposures sustained between 5-15 years previously when compared with the ERR (3.9%) for exposures experienced 15 years or more in the past. Though recent exposures indicated greater risk the estimate of ERR/WLM in the recent TSE window was subjected to much greater variability (SE = 0.1137) than the estimate in the more distant window of TSE (SE = 0.0283). A LRT for exposure effect in TSE windows resulted in a LR statistic of 6.842 on 2 degrees of freedom which implied a highly significant effect of exposure expressed under TSE windows of 5-15 years and 15 years or more ($p = 0.0327$).

Exposures measured in the two TSE windows 5-15 years and 15 years or more, were further examined in a log-linear model (model 1). This model had a slightly improved fit over the linear model (log-linear multiplicative model deviance = 187.106; linear additive model deviance 188.086; DF for both models 1,137). Estimates of RR/WLM obtained from log-linear modelling were 1.033 (95% CI: 0.9778 - 1.092) and 1.025 (95% CI: 1.008 - 1.042) respectively in the two TSE windows, which corresponded to estimates of ERR/WLM of 3.3% and 2.5%. Once again, estimates of risk in the recent TSE window, though higher, considerably lacked precision and were not found to be statistically significant.

It can be concluded from both these models that there was a statistically significant increase in ERR per WML due to exposures experienced in the TSE window of 15 years or more amongst the RH cohort.

It must also be noted in concluding, that, though the model fitted above uses the BEIR IV classification of TSE windows, it is not the complete BEIR IV model and comprised only of one part of it. Exploration of the complete BEIR IV model which additionally allows for the modifying effect of attained age is delayed till chapter 6 and chapter 7.

5.6 External Adjustment for Confounders 1: Indirect Standardization

The primary aim of indirect standardization performed in this chapter is to compare patterns of lung cancer mortality in the RH cohort with those in the Australian National Population. Comparative estimates of Rn related lung cancer mortality rates and risks were obtained for the RH cohort after adjustment to the lung cancer mortality rate in the Australian National Population. Estimates thus computed provide answers to the question: is there an excess in lung cancer mortality in the RH cohort if the lung cancer mortality rates of the Australian National Population were applied?

Australian National Population lung cancer mortality rates were used to compute age and calendar year specific expected numbers of lung cancer deaths in the RH cohort. Two methods of computing expected numbers of deaths were explored; the first using reference population rates specific to each calendar year of observation, and the second using five-year grouped reference population rates. The reason for exploring these alternative methods was to enable comparisons with other studies that had used five-year pooled reference populations and to examine the efficacy of such an approach. Reference population rates are generally more readily available in five year periods rather than on an annual basis and analyses based on five year periods are generally thought to be computationally less intensive than the derivation of annual expected numbers of deaths. In this study however, annual reference rates were available and five year pooled rates were constructed from these annual rates to enable the comparison between annual and pooled rates.

Indirect standardization was then performed using Poisson regression modelling in EPICURE where the expected number of deaths was modelled as the outcome variable with the confounders being declared as stratification variables.

Results of indirect standardization showed that age and calendar year specific SMRs were homogeneous (tables 5.6a and 5.6b). Summary SMR estimates were therefore considered representative of these stratum specific estimates.

Table 5.6a: Age Specific SMRs for Lung Cancer Mortality

Age Group	Events (Obs)	Annual Weights			5 Year Grouped Weights		
		Exp	SMR	95% CI	Exp	SMR	95% CI
>15	0	0.00	-	-	0.00	-	-
15-<20	0	0.00	-	-	0.00	-	-
20-<25	0	0.01	-	-	0.01	-	-
25-<30	0	0.00	-	-	0.01	-	-
30-<35	0	0.08	-	-	0.08	-	-
35-<40	0	0.22	-	-	0.22	-	-
40-<45	1	0.72	1.39	0.20 - 9.87	0.75	1.34	0.19 - 9.50
45-<50	6	2.20	2.73	1.23 - 6.07	2.13	2.81	1.26 - 6.26
50-<55	7	4.19	1.67	0.80 - 3.51	4.16	1.68	0.80 - 3.53
55-<60	6	6.14	0.98	0.44 - 2.17	5.95	1.01	0.45 - 2.24
60-<65	15	5.93	2.53	1.53 - 4.20	5.88	2.55	1.54 - 4.23
65-<70	10	4.48	2.31	1.20 - 4.15	4.52	2.21	1.19 - 4.12
70-<75	8	2.99	2.68	1.34 - 5.35	2.96	2.71	1.35 - 5.41
75-<80	1	1.56	0.64	0.09 - 4.56	1.56	0.64	0.09 - 4.55
80-<85	0	0.54	-	-	0.53	-	-
>85	0	0.11	-	-	0.11	-	-
Total	54	29.17	1.85	1.42 - 2.42	28.86	1.87	1.43 - 2.44

Note: All figures rounded to nearest second decimal.
Homogeneity Test for Age Specific SMRs:-
Based on Annual Weights: LRT Statistic = 11.88 DF = 14 $p = 0.6158$
Based on 5-year Grouped Weights: LRT Statistic = 11.75 DF = 14 $p = 0.6264$

Table 5.6b: Calendar Year Specific SMRs for Lung Cancer Mortality

CALENDAR YEAR	EVENTS (OBS)	ANNUAL WEIGHTS			5 YEAR GROUPED WEIGHTS		
		EXP	SMR	95% CI	EXP	SMR	95% CI
1948-1952	0	0.05	-	-	0.05	-	-
1953-1957	0	0.49	-	-	0.49	-	-
1958-1962	2	1.07	1.87	0.47 - 7.47	1.12	1.79	0.45 - 7.15
1963-1967	2	2.14	0.93	0.23 - 3.73	2.17	0.92	0.23 - 3.68
1968-1972	6	3.68	1.63	0.73 - 3.63	3.66	1.64	0.74 - 3.65
1973-1977	13	5.44	2.39	1.39 - 4.12	5.35	2.43	1.41 - 4.19
1978-1982	14	7.37	1.90	1.12 - 3.21	7.38	1.90	1.12 - 3.21
1983-1987	17	8.93	1.90	1.18 - 3.06	8.65	1.97	1.22 - 3.16
Overall	54		1.85	1.42 - 2.42		1.87	1.43 - 2.44

Note: All figures rounded to nearest second decimal.
Homogeneity Test for Age Specific SMRs:-
Based on annual weights: LRT Statistic 4.084 DF = 7 $p = 0.7701$
Based on 5 year grouped weights: LRT Statistic 4.268 DF = 7 $p = 0.7484$

Indirectly standardized rates were further examined using models with additional adjustment for age and calendar year to examine the existence of any residual effects of age and calendar year. Results of these analyses confirmed a lack of residual effects in age and calendar year after adjustment for background rates based on an external reference population ($p > .05$).

Estimates of lung cancer mortality rates and risk obtained through indirect standardization are summarized in table 5.6c, which included results obtained using both, annual weights and five-year grouped weights. These results show very little difference between the two systems of weighting. Hence, further reporting will be based on results obtained using annual weights, because detailed annual reference population data are readily available for this study and analyses based on annual rates may be more accurate than analyses based on pooled rates.

Results in table 5.6c show that compared to the Australian National Population, an excess of 85% was observed in the overall rate of lung cancer mortality amongst the RH cohort (SMR = 1.85); this excess was statistically significant at the 5% level (95% CI: 1.42 - 2.4).

The indirectly standardized lung cancer mortality rate was lower among those unexposed (SMR = 1.59) than those exposed (SMR = 2.09); however, though the risk was 32% greater in those exposed this increase was not statistically significant (RR = 1.32; 95% CI: 0.08-2.3).

Estimates of SMRs and RR obtained by more detailed exposure categories (table 5.6c) showed statistically significant heterogeneity in the stratum specific estimates of RR with an increasing trend with exposure. However, only those exposed to over 40 cumulative WML of Rn were seen to be at significantly higher risk.

Table 5.6c: Indirectly Standardized Estimates of Lung Cancer Mortality (Age and Calendar Year Adjusted)

EXPOSURE	ANNUAL WEIGHTS				5 YEAR GROUPED WEIGHTS			
GROUP	SMR	95% CI	RR	95% CI	SMR	95% CI	RR	95% CI
Overall	1.85	1.4-2.4	-	-	1.87	1.4-2.4	-	-
Unexposed	1.59	1.0-2.4	1	Fixed	1.60	1.2-2.4	1	Fixed
Exposed	2.09	1.5-3.0	1.32	0.8-2.3	2.12	1.5-3.0	1.33	0.8-2.3
0	1.59	1.0-2.4	1	Fixed	1.60	1.1-2.4	1	Fixed
0<-1	3.04	1.5-6.1	1.92	0.9 - 4.3	3.10	1.5-6.2	1.9	0.9 - 4.4
1<-10	1.13	0.6-2.1	0.71	0.3 - 1.5	1.14	0.6-2.1	0.7	0.3 - 1.5
10<-20	2.46	0.9-6.6	1.55	0.5 - 4.5	2.50	0.9-6.7	1.6	0.5 - 4.6
20<-40	3.70	1.7-8.2	2.33	0.9 - 5.8	3.73	1.7-8.3	2.3	0.9 - 5.8
>40	7.37	2.8-19.6	4.65	1.6 - 13.5	7.30	2.7-19.5	4.6	1.6 - 13.3

Results of Testing for Homogeneity of Exposure-Specific Risk Estimates:

Based on Annual Weights:- Dichotomous Categories: LRT χ^2 = 1.01; DF = 1; p = 0.3130
Multiple Categories: LRT χ^2 = 12.83; DF = 5; p = 0.0250

Based on 5-Year Weights:- Dichotomous Categories: LRT χ^2 = 1.08; DF = 1; p = 0.2997
Multiple Categories: LRT χ^2 = 12.74; DF = 5; p = 0.0260

5.7 External Adjustment for Confounders 2: Direct Standardization

The primary objective of direct standardization performed in this chapter is to obtain risk estimates of Rn related lung cancer mortality adjusted to the age distribution of the Australian National Population.

Direct standardization aims at eliminating any differences in mortality rates that may arise simply by virtue of the differences in age distributions between the study and reference cohorts. Directly standardized rates (DSRs) therefore answer the question: *'what would the mortality rate be if the study population had the same age distribution as the reference population?'*

Direct Standardized Rates (DSR) were obtained by weighting age and calendar year specific crude mortality rates in the study population by the reference population weights. Australian national census data obtained from the most recent census within the study follow-up period - the 1986 census data - were used as the reference population for direct standardization. Age-specific relative frequencies for the Australian national population, obtained from the 1986 census were used as reference population weights; weight numerators comprised the number of individuals in each age group and weight denominators comprised the total number of individuals in the reference population in 1986. Since the study cohort was an occupational cohort the use of a truncated reference population which excluded those below the age of 15 in the reference population as the weight denominator was also explored. Computation of DSRs then proceeded using the formulae given in section 5.7.1.

5.7.1 Computation of Direct Standardized Measures

I. Nomenclature

Symbols	Legend
~	Distributed
E	Expectation
Var	Variance
ln	Natural Logarithm

Notation:

i = i th age-group.

Variable	Target Population	Reference Population
Number of stratum specific deaths:	r_i	R_i
Number of stratum specific PYRS	n_i	N_i
Stratum specific death rate	$p_i = r_i / n_i$	$P_i = R_i / N_i$
	$[\, q_i = (1 - p_i) \,]$	$[\, Q_i = (1 - P_i) \,]$
Total number of deaths	$r = \sum_i r_i$	$r = \sum_i r_i$
Total number of PYRS	$n = \sum_i n_i$	$N = \sum_i N_i$
Overall crude death rate	$p_C = r / n$	$P_C = R / N$
Stratum Specific Weights		$W_i = N_i / N$

II. Direct Standardization Measures

The Direct Standardized Rate:

$$DSR = P_D = \sum_i W_i p_i$$

$$Var(P_D) = \sum_i W_i^2 (p_i q_i / n_j)$$ If N_i is known without error;

$$SE(P_D) = \sqrt{Var(P_D)}$$

The Comparative Mortality Figure:

$$CMF = P_D / P_C = \sum_i W_j p_i \Big/ \sum_i W_i P_i$$

If the reference population is very large relative to the target population, then the sampling errors in the standard rates may be relatively small and therefore, ignored.

Under this assumption the standard error of the CMF may be obtained directly from the standard error of its numerator.

Thus,

$$SE(\text{CMF}) = SE(P_D) / P_C$$

Statistical Significance:

Once a DSR has been computed for a target population it is necessary to confirm its significance i.e., to test that any difference observed between the target and the standard populations may not be attributed to chance alone. The statistical significance of the DSR can be tested under the assumption that the DSR is distributed normally, with an Z-test.

95% Confidence Interval for DSR: $P_D \pm 1.96\ SE(P_D)$

SE(CMF) may be regarded as a measure of statistical precision of the CMF. However, since the distribution of the CMF is skewed, it is preferable that confidence intervals and tests of significance requiring an assumption of normality, be based on a *log transformation* of the CMF. The *log transformation* helps correct the skewness of the statistical distribution of the CMF, thereby improving the normal approximation to the distribution of the test statistic based on it (Breslow and Day 1987).

The standard error of the transformed CMF is given by:

$$\begin{aligned} SE(ln\text{CMF}) &= SE(\text{CMF}) / \text{CMF} \\ &= [SE(P_D) / P_C] / (P_D / P_C) \\ &= SE(P_D) / P_D \end{aligned}$$

Confidence intervals and test of significance should then be based on:

Test statistic for CMF: $ln\text{CMF} / SE(ln\text{CMF}) \sim N(0,1)$

Direct Standardized Rate Ratios:

The ratio of two similarly standardized DSR*s* provides an estimate of relative risk between the target populations compared. The statistical significance of the ratio of DSR*s* can be tested with the use of the CMF, since, the ratio of CMF*s* is equal to the ratio of DSR*s*.

Once again, a log transformation of the CMF is used to correct for the skewness in the distribution of the CMF, in testing the ratio of two CMF*s*.

$$ln(\text{CMF}_A/\text{CMF}_B) = ln\text{CMF}_A - ln\text{CMF}_B$$

$$\begin{aligned} SEln(\text{CMF}_A/\text{CMF}_B) &= SE(ln\text{CMF}_A - ln\text{CMF}_B) \\ &= [Var(ln\text{CMF}_A) + Var(ln\text{CMF}_B)]^{1/2} \end{aligned}$$

Test statistic: $[ln(\text{CMF}_A/\text{CMF}_B) / SEln(\text{CMF}_A/\text{CMF}_B)] \sim N(0,1)$

5.7.2 Directly Standardized Estimates of Lung Cancer Mortality

Results of DSR computations are given in table 5.7.2. The choice of the truncated reference population had an impact on the DSR computed; the weight denominator being smaller, these rates were higher than those computed using denominators based on the total reference population. However, RR estimates obtained from the ratio of DSRs were not affected by the choice of reference population.

Table 5.7.2: Directly Standardized Estimates of Lung Cancer Mortality

EXPOSURE	DSR* (UNTRUNCATED REFERENCE POPN.)		DSR* (TRUNCATED REFERENCE POPN.)		RR ESTIMATE (BASED ON CMF RATIOS)	
CATEGORY	RATE	95% CI	RATE	95% CI	RR	95% CI
Overall	955	661-1249	1252	866-1638	-	-
Unexposed	786	423-1149	1031	556-1506	1	Fixed
Exposed	1160	655-1665	1521	859-2183	1.48	0.78-2.78
0 WML	786	423-1149	1031	556-1506	1	Fixed
0<-1 WML	1177	273-2081	1544	359-2729	1.50	0.61-3.67
1<-10 WML	790	149-1432	1036	195-1879	1.01	0.40-2.56
10<-20 WML	798	-30-1635	1046	-50-2143	1.01	0.32-3.19
20<-40 WML	2656	246-5066	3481	322-6640	3.38	1.22-9.34
>40 WML	3935	127-7743	5158	167-10149	5.00	1.71-14.61

NOTE: * Rates per 10^6 PYRS

In view of the differences between DSR obtained using untruncated and truncated reference populations, use of a truncated reference population appears more appropriate in examining the mortality in an occupational cohort study. The overall DSR of lung cancer mortality in the RH cohort thus estimated was 1,252 per 10^6 PYRS (95% CI: 866 - 1638).

The DSR amongst those exposed to Rn progeny at RH was higher than in those unexposed; the DSR being 1,521 (95% CI: 859 - 2,183) per 10^6 PYRS for those exposed, and 1,031 (95% CI: 556 - 1,506) per 10^6 PYRS for those unexposed. DSRs were also obtained for various exposure categories and ranged from 1,036 per 10^6 PYRS to 5,158 per 10^6 PYRS.

Estimates of rate ratios obtained using the Comparative Mortality Figure (CMF) showed a 48% greater risk amongst those exposed compared to those unexposed to Rn progeny at RH (RR = 1.48); however, this excess was not statistically significant at the 5% level (95% CI: 0.78 - 2.78).

Examination of directly standardized rates by various categories of exposure showed an overall increasing trend in DSR*s* with increasing exposure. Estimates of RR obtained from rate ratios and comparative mortality figures showed a significantly elevated risk amongst those exposed to over 20 WML; compared to those unexposed, those in the cumulative exposure category of 20-40 WML experienced a RR of greater than three-fold and those exposed to over 40 cumulative WML were at five times greater risk.

5.8 Conclusions

5.8.1 Comparison of Methods of Controlling for Confounders

The lung cancer mortality rate amongst former workers at the RH uranium mine was significantly higher than that in the Australian National Population (among those unexposed at RH: SMR = 1.59, 95% CI: 1.0-2.4). Thus, the Australian National Population reference does not appear to be representative of the mortality amongst those unexposed and cannot therefore be considered as a suitable reference population (Thomas *et al.* 1992).

Results of various methods of controlling for confounders are summarized in table 5.8.1 for further examination. Though the standard errors of estimates obtained by using external indirectly standardized rates are slightly smaller than those obtained by internal adjustment, this slight improvement in precision is gained at the cost of the accuracy of these estimates. Estimates obtained from external standardization tend to be systematically higher than those obtained from internal adjustment. With the appropriateness of the chosen external reference population being in doubt, any upward bias caused by this source is avoided in all further analyses in this work which will only be based on internal adjustment for confounders.

Table 5.8.1: Methods of Controlling for Confounders - Comparative Findings

Exposure	Internally Adj		Indirect STD		Direct STD	
Category	RR	95% CI	RR	95% CI	RR	95% CI
Unexposed	1	Fixed	1	Fixed	1	Fixed
Exposed	1.3	0.7 - 2.2			1.5	0.8 - 2.8
0 WML	1	Fixed	1	Fixed	1	Fixed
0<-1 WML	1.8	0.8 - 4.0	1.9	0.9 - 4.3	1.5	0.6 - 3.7
1<-10 WML	0.7	0.3 - 1.4	0.7	0.3 - 1.5	1.0	0.4 - 2.6
10<-20 WML	1.6	0.5 - 4.6	1.6	0.5 - 4.5	1.0	0.3 - 3.2
20<-40 WML	2.1	0.8 - 5.2	2.3	0.9 - 5.8	3.4	1.2 - 9.3
>40 WML	4.4	1.5 - 13.2	4.7	1.6 - 13.5	5.0	1.7 - 14.6

5.8.2 Comparison of Models of Exposure-Response Relationship

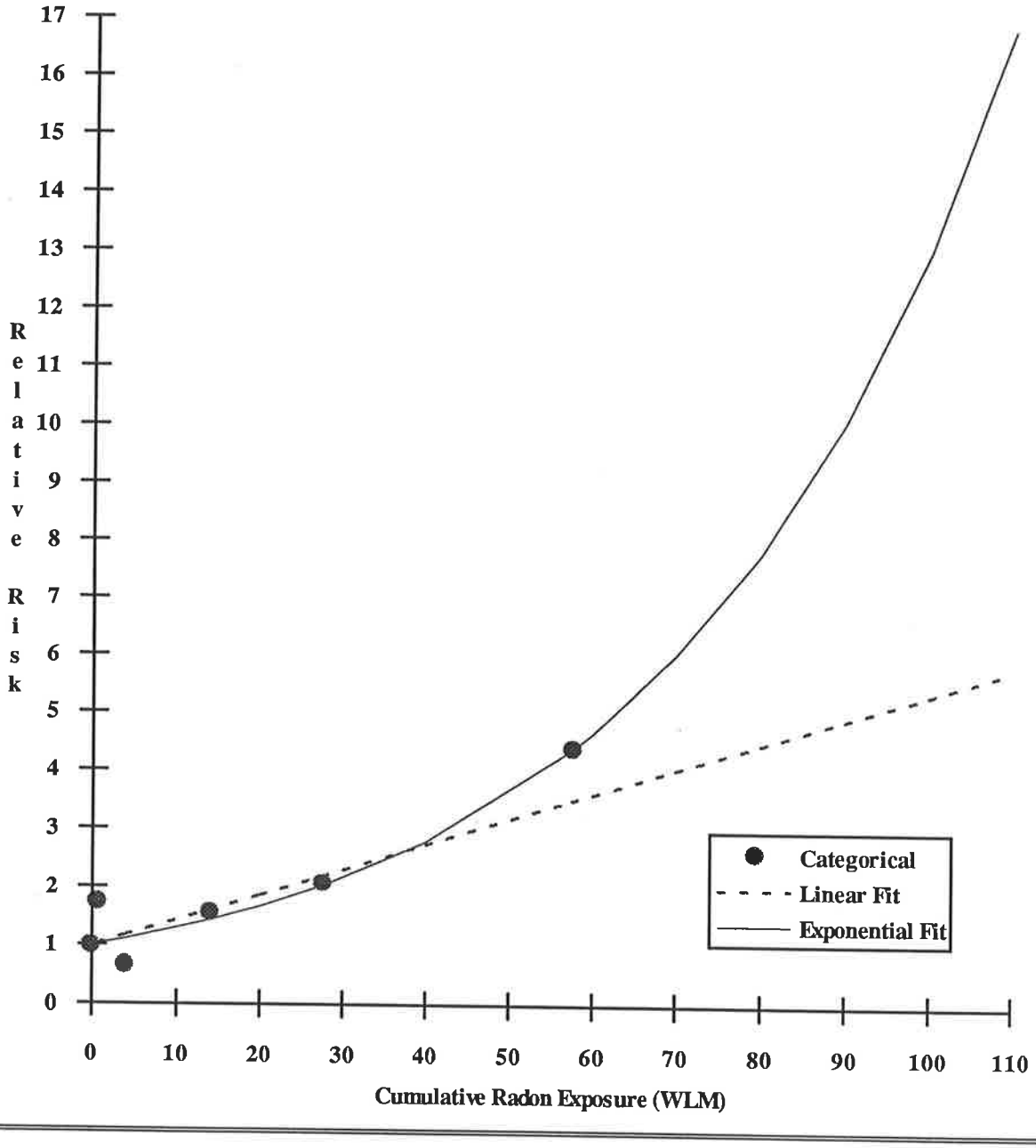
Table 5.8.2 summarizes results of Rn related risk of lung cancer mortality after internal adjustment for confounders by fitting linear and log-linear models to Rn progeny exposure expressed categorically and continuously.

Table 5.8.2: Summary of Results from Exposure-Response Modelling With Internal Adjustment for the Confounders: Attained Age and Calendar Year

Model	Risk Estimate	Goodness of Fit		
		Deviance	DF	χ^2
Null Model: RR = e^{β}		194.93	1139	883
Constant Slope Models (Continuous Exposure):				
Linear: RR = $1 + \beta w$	ERR/WLM = 4.3%	188.32	1138	603
Linear TSE Model:	ERR/WLM β_1 = 8.5%	188.09	1137	569
RR = $1 + \beta_1 w_{5\text{-}15} + \beta_2 w_{15\text{-}}$	ERR/WLM β_2 = 3.9%			
Log-Linear: RR = $e^{\beta w}$	ERR/WLM = 2.6%	187.05	1138	611
Separate Slopes Models (Categorical Exposures):				
Dichotomous Categories:	$RR_{(unexposed)}$ = 1	194.32	1138	811
RR = $e^{\beta w_C}$	$RR_{(exposed)}$ = 1.25			
Multiple Categories:	$RR_{(0\ WLM)}$ = 1	182.84	1134	462
RR = $e^{\beta w_C}$	$RR_{(0<\text{-}1\ WLM)}$ = 1.75			
	$RR_{(1<\text{-}10\ WLM)}$ = 0.66			
	$RR_{(10<\text{-}20\ WLM)}$ = 1.56			
	$RR_{(20<\text{-}40 WLM)}$ = 2.09			
	$RR_{(>40 WLM)}$ = 4.38			

Estimates of RR obtained in the strictly categorical analyses demonstrate a drop in RR in the second and third lowest categories of exposure. This suggests that the exposure-response relationship may not be strictly linear. A linear fit of Rn progeny exposure showed an ERR of 4.6% per WML and is plotted as the straight line in figure 5.8.2. When fitted as a log-linear function the coefficient of ERR was 2.6% per WML.

Figure 5.8.2: Exposure-Response Models
(All Male Workers)



Since models fitted in table 5.8.2 are not nested with each other, their goodness-of-fit cannot be compared using likelihood ratio tests. Therefore, the magnitude of the deviance of each model is used as an indicator of its goodness-of-fit with the interpretation being that smaller deviances indicate better fit. Another means of comparing the goodness-of-fit of these models is through visual examination; figure 5.8.2 shows that the log-linear model appears to fit the categorical risk estimates better than the linear model, particularly at higher levels of cumulative Rn progeny exposures; the log-linear model also has a smaller deviance than the linear model. However, it appears that there is no great advantage in the log-linear fit over the linear fit because of the small difference in deviances and its, closeness in fit to the linear model at lower levels of cumulative Rn progeny exposure (below 40 WML). Therefore, the linear model which is more easily interpretable may be regarded as the preferred model for risk assessment when considering exposures at low cumulative levels of Rn progeny.

Examination of the exposure-response relationship under time-since-exposure windows showed that risk of exposures experienced 5-15 years prior to observation was much higher than the risk from exposures experienced before that time. Estimates of risk showed an 8.5% increase in ERR/WLM for exposures the 5-15 year TSE window and a 3.9% increase in ERR/WLM for exposures experienced more than fifteen years previously.

Findings summarized in table 5.8.1 and figure 5.8.1 also provide an example of the influence of the artificial constraint of the reference exposure category being assigned a RR of 1 - in this case the lowest exposure category of 0 WML is used as the reference category - on relative risk estimates obtained from categorical evaluation, where a change in the reference category could cause considerable change to the pattern of the RR estimates. This phenomenon evades perfunctory eye examination of plots such as diagram 5.8.1 has been described in detail by Thomas *et. al* (Thomas *et al.* 1992) who ascribe it to the uncertainty that stems from the base-line reference group (Breslow and Day 1980). Fits based on continuous models are not thus constrained and therefore, not strictly comparable with patterns of categorical RR estimates.

5.8.3 Conclusions

Based on the finding from the analyses presented in this chapter, it is concluded that:

- attained age and calendar year were confounders of the relationship between cumulative Rn progeny exposure and lung cancer mortality.

- the lung cancer mortality rate amongst former workers at the RH uranium mine was significantly higher than that in the Australian National Population.

- the elevated SMR from lung cancer amongst those unexposed to Rn progeny at RH, shows that the Australian National Population cannot be considered a suitable reference population, representative of the mortality amongst RH miners who were not exposed to Rn at RH; all further risk evaluation contained in this work will therefore be based on internal references.

- categorical evaluation after controlling for the confounding effect of attained age and calendar time of observation, showed that lung cancer mortality rates tended to increase with increasing categories of cumulative Rn progeny exposure; those exposed to over 40 cumulative WML of Rn progeny at RH were at significantly greater risk (over 4 times) of dying from lung cancer, than those unexposed to Rn progeny at RH.

- risk evaluation based on exposure treated as a continuous variable after controlling for confounders - attained age and calendar time of observation - showed that cumulative exposure to Rn progeny had a significant effect on lung cancer mortality and that the ERR increased linearly, by 4.3% per WML, or exponentially, by 2.6% per WML; both the linear and log-linear trends in the exposure-response relationship were statistically significant at the 5% level of significance.

- risk evaluation based on time-since-exposure models showed that the excess risk of lung cancer mortality associated with exposures experienced 5-15 years previously was higher (over two-fold) than that associated with exposures experienced in the more distant past; estimates of ERR/WLM were 8.5% and 3.9%, respectively for these windows of time-since-exposure.

Analyses presented in this chapter concentrated on identifying and demonstrating the role of confounders of the effect of Rn progeny exposure on lung cancer mortality amongst the Radium Hill cohort and obtaining estimates of risk after adjustment for confounders alone. Risk evaluation continues in chapter 6 where the roles of several risk modifiers will be examined.

CHAPTER 6

6. Temporal Effect Modifiers and Surrogate Measures of Exposure: Person-Years Based Analyses

Page

Tables and Figures

Page

Page

6. Temporal Effect Modifiers and Surrogate Measures of Exposure: Person-Years Based Analyses

6.1 Aims

Time has been identified as an important modifier of exposure-response relationships with respect to Rn related lung cancer mortality. '*Time*' in this context includes such factors as age at exposure, latency or time since exposure, duration of exposure, pattern of exposure intensity over time and calendar time (Thomas 1988). Assessment of exposure-response relationships are studied in conjunction with other confounding or interacting factors.

Confounding effects have been addressed in chapter 5. This chapter examines the effects of interactions and effect modification, and the roles of surrogate measures of exposure using the PYRS based *contingency table* approach and Poisson regression techniques for grouped cohort data.

The specific aims of the analyses presented in this chapter are summarized as follow:

1. Examine effects of age at first exposure, time since last exposure, cumulative duration of exposure and intensity of exposure on the exposure-response relationship of Rn related lung cancer mortality.
2. Obtain estimates of Rn related lung cancer mortality risk in the presence of effect modifiers.
3. Examine modifiers of effect under time-since-exposure (TSE) windows. Modifiers studied comprise cumulative duration of exposure and intensity of exposure.
4. Examine the role of surrogate measures of Rn exposure on lung cancer mortality amongst the RH cohort. The surrogate roles of cumulative duration of exposure and intensity of exposure are examined in this chapter.

6.2 Methodology and Preliminary Analyses

This section commences with a description of the data preparation and an introduction to the nomenclature and methodology adopted in the analyses presented in this chapter. It is concluded with analyses that form the basis of reference for examining the role of potential modifiers of the exposure-response relationship in lung cancer mortality related to Rn progeny exposure and the efficacy of surrogate measures of exposure.

6.2.1 Data Preparation and Nomenclature

All analyses presented in this chapter are based on cohort data compiled into tables of PYRS and events viz., lung cancer deaths, cross-classified by each of the potential confounders, effect modifiers, other relevant covariates and exposure variables. Potential effect modifiers and other temporal covariates examined in this chapter comprise age at first exposure, time since last exposure, duration of exposure and intensity of exposure. The only exposure variable considered in the examination of the role of modifying effects in this chapter is a measure of cumulative Rn exposure (hereafter, referred to as cumulative exposure or CE) which was derived from a five year lagged exposure as described in chapter 5.

The purpose of the work presented in this chapter is to examine the modifying role of various factors on the exposure-response relationship, for individuals who experienced some exposure to Rn progeny at RH, where the exposure response relationship is characterized by the effect of increasing exposure given some exposure. Therefore, only those *exposed* to some *relevant* Rn progeny exposure are included in these analyses. Being PYRS based, these analyses also exclude contributions of PYRS and events made to the *unexposed* categories, by those who were subsequently exposed. The basic dataset used for the analyses presented in this chapter is therefore, only a subset of the dataset used in chapter 5 which comprised the complete

spectrum of Rn progeny exposures - from zero to the highest WLM - experienced by individuals at RH, including exposures of *zero* WLM.

For convenience of use, in the remainder of this work all variables will be referred to by variable names ascribed to them in the specific context of their use, i.e. whether continuous or categorical. These variable names are summarized in table 6.2.1a. Categorization of variables was based on the same principles described in chapter 5. All potential effect modifiers considered have been categorized into three categories. Exposure categorization adopted in chapter 5 yielded PYRS and event tables that were somewhat sparse when tabulated across categories of effect modifiers; cumulative exposure classification was therefore compressed into four categories of *relevant* exposure, with a combination of exposures in the ranges of 10-20 WLM and 20-40 WLM. Details of categorizations are summarized in table 6.2.1b.

Table 6.2.1a: Summary of Variable Names
(Abbreviations given in brackets)

Factors Studied and Abbreviations	Units	Variable Names	
		Continuous	Grouped
Exposure Variable:			
Cumulative Relevant Exposure (CE)	WLM	WLML05 or WLM	WLMG
Potential Effect Modifiers:			
Age at First Exposure (AFE)	years	AFE	AFEG
Time Since Last Exposure (TSLE)	years	TSLE	TSLEG
Cumulative Duration of Exposure (CDE)	WM	WML05	WMG
Intensity of Exposure (IE)	WL	WL05	WLG

Table 6.2.1b: Summary of Categorical Variable Groupings

Variable	Unit	Categorization			
WLMG	WLM	0<1	1-<10	10-<40	40->
AFEG	Years	<30	30-<40	40->	
TSLEG	Years	<20	20-<25	25->	
WMG	WM	0<3	3-<24	24->	
WLG	WL	0<0.4	0.4-<1	1->	

6.2.2 Methodological Overview

The roles of potential modifiers in the exposure-response relationship of Rn exposure and lung cancer mortality were examined by fitting relative risk models using Poisson regression techniques. Exposure was measured by cumulative Rn exposure levels based on a five year lag period (w) and the response was characterized by the lung cancer mortality rate $r(x,z,w)$. Relative risk models were based on the general assumption that the lung cancer mortality rate depended on the estimated exposure (w), the background disease rate $r_0(x)$, and factors - potential effect modifiers - which affected the exposure-response relationship $RR(z,w)$. Factors which contributed to the background disease rate comprised attained age and calendar year of observation, represented by the vector (x). Potential effect modifiers and covariates considered in this work - represented by the vector (z) - include age at first exposure, time since last exposure, duration of exposure and intensity of exposure. The modifying effect of smoking and other occupational exposures will not be examined using PYRS based analyses due to the sparseness of the data and will therefore not be considered in this chapter. These issues will be addressed in chapter 7 using nested case-control analyses.

Using the above notation, the general form of the RR model adopted in this chapter, can be represented as:

General Form of *RR* Models: $$r(x,z,w) = r_0(x)\, RR(z,w) \qquad (1)$$

The relative risk component of this model viz., $RR(z,w)$, may assume various functional forms. Several functional forms for $RR(z,w)$ will be examined in this chapter. These include the multiplicative or log-linear relative risk model where w and z are modelled as exponential functions, a simple linear function and the resulting linear additive excess relative risk model, the log excess relative risk model where z is modelled as an exponential function, and a power model in which z is modelled as a power function. These models are described in section 6.2.3 - 6.2.5.

Examination of the role of each covariate or potential modifier of the exposure-response effect commenced with the tabulation of events and PYRS, and the derivation of crude *RR* estimates, cross-classified by categories of exposure and potential effect modifiers. Crude risk estimates obtained at this stage were unadjusted for the background effects of the potential confounders attained age and calendar year of observation. Adjustment for potential confounders was then made through stratified analyses as described in chapter 5 and these findings were used to evaluate the confounding effects of attained age and calendar year of observation.

Categorical evaluation using simple multiplicative models also enabled examination of the specific nature of the effect of each covariate or potential modifier on the exposure-response relationship. After accounting for the presence of potential confounders, main effects and interactions were examined using simple relative risk models for categorical data; main effects were examined for their roles as component effects and the residual effect of interactions after removing main effects were used to examine the extent of effect modification or interaction as suggested by Pearce (Pearce 1989).

One of the problems arising from methods of analysis based on categorical data where continuous variables are 'discretized' is that the choice of boundaries for each category can influence the results and thereby, the assessment of interaction (Siemiatycki and Thomas 1981; Rothman and Keller 1972). This particular problem of 'discretisation' was avoided in the next stage of analysis which was based on continuous data and variations in ERR *per unit exposure* were examined using various parametric representations of the exposure-modifier-response and exposure-time-response relationships. Variation between estimates derived from different well-fitting statistical models are likely to be minimal compared to variation due to random and systematic errors (Siemiatycki and Thomas 1981). However, this approach can lead to incorrect inferences about interactions if assumptions about exposure-response relationships are incorrect (Thomas 1981). These approaches of statistical modelling are introduced in sections 6.2.3 - 6.2.5 and are systematically applied in the evaluation of the role of each potential effect modifier and covariate in sections 6.3 - 6.5.

6.2.3 Methods of Describing the Nature of the Modifying Effect Using Log-Linear Models

The specific nature of interactions was characterized using log-linear models. Tables of *RR*'s of interactions between categories of exposure and potential effect modifiers were then constructed with the lowest of these classification or cross-classification categories being the basis of reference. *RR* estimates were obtained by fitting log-linear models defined below.

Log-Linear *RR* Models: $$RR = e^{\beta_i w_c} e^{\gamma_j z_c} e^{\tau_k w_c z_c} \quad (2)$$

Marginal and interactive estimates of risk were obtained by including only the effects of interest in model (2) and are defined in models 2a-2c below:

Individual Main Effects Models: $RR = e^{\beta_i w_c}$; $RR = e^{\gamma_j z_c}$ (2a)

Joint Main Effects Model: $RR = e^{\beta_i w_c} e^{\gamma_j z_c}$ (2b)

Interaction Effects Model: $RR = e^{\tau_k w_c z_c}$ (2c)

Log-linear models were used at this stage of the analysis due to the convenience of their application and the ease with which parameter estimates and confidence bounds could be obtained. Based on the assumption that risks are multiplicative in nature, *RR* estimates could be conveniently obtained from these models, by simply dividing the risk estimate in each category by the risk in the appropriate baseline category. Confidence intervals for all effects were obtained from Wald's bounds (EPICURE 1992).

The effect of each factor on the exposure-response relationship was examined through the sequential inclusion of main effects and interactions in the log-linear model, and likelihood ratio tests for significance of the newly introduced effect. This corresponded to a test of heterogeneity in categorical estimates. The goodness of fit of log-linear models was examined through comparison of model deviances.

6.2.4 Methods of Examining Effect Modification by Modelling Variations in ERR/WLM Using Linear Additive Excess Relative Risk Models

Mathematical modelling then continued with fitting the simplest form for the *RR* function - a linear *RR* function - also referred to as the linear excess relative risk (ERR) model (model 3a).

Linear ERR Model: $$RR = 1 + \beta w \qquad (3a)$$

The linear ERR model assumes that the *RR* varies linearly with exposure and that the variation is constant over the range of exposures studied. The parameter β is therefore, an estimator of the constant increase in *RR* per unit increase in exposure (ERR/WLM); it is the common slope coefficient fitted in the linear exposure-response model. Confidence intervals for β and all parameters of linear effects were derived from likelihood ratio bounds (*LRB*), when attainable. The linear additive ERR model also provided the basis of reference for further tests of goodness-of-fit and significance. Where the parameter was known to be positive and *LRB* were difficult to obtain, the log excess relative risk model (model 3b) was fitted.

Log ERR Model: $$RR = 1 + w\, e^{\beta} \qquad (3b)$$

The next stage of the analysis involved examination of whether a single straight line provided an adequate expression of the exposure-response relationship in the presence of a potential effect modifier, or whether it was significantly influenced or modified by *z*. Whilst still assuming a linear relationship, independent slopes were fitted for each category of the effect modifier. Each potential effect modifier *z*, was treated as a categorical variable with J categories. Model (3a) was then re-fitted to examine the exposure-response slopes within each level of z, resulting in the following model:

Separate Slopes Model: $$RR = 1 + \beta_j\, wz_c \qquad (3b)$$

The parameters β_j provided estimates of the ERR/WLM within category j of the potential effect modifier, denoted by z_j; heterogeneity of the β_j's was considered an indication of z being an effect modifier. Any significant heterogeneity between these parameters (β_j's) was used as an evidence of effect modification. Heterogeneity of β_j's was tested by comparing the deviances from models (3) and (2). Under the null hypothesis of no effect modification (homogeneity of β_j's) the differences in model deviances would have a χ^2 distribution with j-1 degrees of freedom. A 'significant' p-value from this test of hypothesis - likelihood ratio test (LRT) - was considered as confirmation of *z* being an effect modifier (i.e. of the effect of Rn exposure on lung cancer mortality not being homogeneous across levels of *z*). Appropriate levels of significance are individually addressed with relevance to specific situations. Confirmation of heterogeneity in the stratum-specific slope estimates indicated that the exposure response relationship varied between the categories of *z*, and could therefore, not be adequately summarized by a single slope coefficient.

An estimate of the overall ERR per unit of *w* per unit of *z* was obtained by modelling the interaction of *w* and *z* as continuous variables in the linear ERR model as follows:

Continuous Effect Modification Model: $RR = 1 + \beta_1 w + \beta_2 wz$ (3c)

Another test for effect modification was performed using a LRT of the change in model deviance by including the interaction term in the model after including the main exposure effect term. The ERR parameter (β_2) thus obtained provided an estimate of the ERR/WLM/*unit of z*, after accounting for the main effect of exposure.

6.2.5 Methods of Examining Effect Modification by Modelling Variations in ERR/WLM Using Smooth Nonlinear Parametric Functions

Yet another approach of evaluating effect modification was to use an exponential effect modification model (model 4), where z was fitted as an exponential interaction term. More specifically, this model was fitted by introducing an interaction term with w as a linear function and z as a log-linear function in EPICURE (EPICURE 1992).

Exponential Modification Model: $$RR = 1 + \beta w \ e^{\gamma z}, \qquad (4)$$

Tests for effect modification based on LRTs, were performed in a similar manner to those outlined for the Linear Additive ERR model. In the presence of two potentially interactive effects, the ERR of each increasing unit of z for a fixed level of exposure was deemed a more appropriate and interpretable estimate. An estimate of ERR/*unit of z for a fixed level of exposure* was obtained from g after constraining the parameter of the exposure effect to 1 (i.e. by fixing $\beta = 1$), in the log ERR model.

When the potential effect modifiers were used as continuous variables, parameter γ yielded an estimate of the ERR per Unit increase in z for a given exposure. When used categorically, independent estimates of ERR were obtained for each category of z. Additional estimates were also obtained for proportional change in ERR's and *RR*'s were obtained between levels of the potential effect modifiers. The ratio of parameter estimates provided estimates of proportional change in ERR/WLM; the differences in parameter estimates provided estimates of ERR between categories and the ratio of $1+\gamma_j$'s provided estimates of the *RR* of exposure between categories of the potential effect modifiers.

Departures from linearity and variations of ERR/WLM were further examined using another smooth parametric function of z - a power function - which resulted in the *power model* (model 5) defined as follows:

Power Modification Model: $$RR = 1 + \beta w \; z^{\gamma} \qquad (5)$$

The significance of departures from linearity were tested using LRTs. Overall assessment of model fit was based on a direct comparison of the deviances of the various models fitted. Though no tests of goodness of fit exist for this purpose, the model with the smallest deviance was accepted as the *'best fit'*.

The exponential modification model and the power model also facilitated the examination of departures from the linear model, with the exponential and power functions representing departures from linearity and γ being the *non-linearity* parameter.

The roles of each effect modifier studied are examined independently and findings summarized in sections 6.3 - 6.6. At the end of each of these sections, results from the various models fitted are summarized in a single table to enable quantitative comparison; thereafter, to enable visual comparison of the various models, modifier-response relationships for a given level of Rn progeny exposure are also graphically represented as simple plots.

6.2.6 Preliminary Analyses: The Effect of Radon Progeny Exposure on Lung Cancer Mortality

The role of each potential effect modifier is examined by comparison against the unmodified effect of Rn exposure on lung cancer mortality. Therefore, results from examining the unmodified effect of Rn exposure on lung cancer mortality, are a pre-requisite for all effect modification analyses. These results are reported in this section and will form a general basis of reference for all further analyses undertaken in this chapter.

The methods of analyses presented in this section are similar to those presented in chapter 5, section 5.5. The important question in this chapter, lies in the increase in risk given that one is exposed to Rn progeny and the modifying effect of time-related factors on this exposure-response relationship. Therefore it is important to base risk evaluation, particularly in the presence of effect modifiers, only on PYRS and events experienced in categories exposed to some relevant Rn exposure. It must therefore be noted that slope estimates obtained here will differ from those obtained using the entire dataset which included the zero WLM exposure category - in fact in this case, the estimates are higher than those obtained in chapter 5 - since they are not influenced by the unexposed category, the analyses presented in this chapter being restricted only to exposure categories of *relevant* exposure > 0 WLM.

The effect of cumulative Rn exposure on lung cancer mortality was first examined with a log-linear stratified exposure effect model (2a). The results of this modelling are given in table 6.2.6a. The *RR* of lung cancer mortality was seen to increase among levels of cumulative exposure (CE) with the exception of the second lowest category viz., 1-10 WLM. Stratum specific *RR* estimates with reference to the base-line reference category - the lowest exposure category who experienced a CE of 0-1 WLM - were 0.37, 1.10 and 3.39 respectively, for each increasing category of exposure. The statistically significant heterogeneity observed in these estimates confirmed a significant exposure-response relationship (LRT $p = 0.0035$).

Confidence intervals for the individual *RR*s showed that whilst none of these estimates was significantly higher than the *RR* of 1 at the 5% level of significance, those exposed in the highest exposure category (>40 WLM) were at significantly higher risk than those with CE of 1-10 WLM. The significance of this comparison would have been the main contributor to the overall significance in the heterogeneity observed among the stratum specific estimates of risk.

Table 6.2.6a: Output From Log-Linear Stratified Exposure Effect Model

#	Name	Parameter Estimates: Estimate	Std. Error	95% Confidence Bounds: Lower	Upper
Log-linear term 0					
1	%CON	0.4571	Aliased		
3	WLMG4_2	-0.9829	0.4753	-1.915	-0.5132E-01
	EXP(estimate)	0.3742	1.609	0.1474	0.9500
4	WLMG4_3	0.9570E-01	0.4804	-0.8458	1.037
	EXP(estimate)	1.100	1.617	0.4292	2.821
5	WLMG4_4	1.219	0.6275	-0.1065E-01	2.449
	EXP(estimate)	3.385	1.873	0.9894	11.58

Deviance = 175.987 df = 1907
Pearson Chi2 = 1016.18

LR statistic = 13.63 df = 3
p = 0.0035

Table 6.2.6b: Output From Log-Linear Exposure-Response Model

#	Name	Parameter Estimates: Estimate	Std. Error	95% Confidence Bounds: Lower	Upper
Log-linear term 0					
1	%CON	-0.2385	Aliased		
2	WLML05	0.2880E-01	0.8311E-02	0.1251E-01	0.4509E-01
	EXP(estimate)	1.029	1.008	1.013	1.046

Deviance = 181.021 df = 1909
Pearson Chi2 = 1212.83

Patterns of variation in ERR/WLM were further examined using the linear additive ERR and exponential modification exposure-response models. Results of fitting these models are given in tables 6.2.3c and 6.2.3d. Both these models showed a 5.4% increase in ERR/WLM, after adjustment for attained age and calendar year, implying that the *RR* of Rn related lung cancer mortality increased linearly by 5.4% for each additional WLM of exposure. This increase was statistically significant at the 5% level of significance.

Table 6.2.6c: Output From Linear Additive Excess Relative Risk Model

Parameter Summary Table

#	Name	Estimate	Std.Err.	Score	Status
	Log-linear term 0				
1	%CON	-0.4541	0	0	Aliased
	Linear term 1				
2	WLML05	0.05435	0.04121	0.00414	Free
	Confidence Bounds:	97.5% Lower Bound	0.00807		
		97.5% Upper Bound	0.17427		

Deviance	=	182.814	df =	1909
Pearson Chi2	=	1191.79		
LR Statistic	=	6.806	df =	1
p	=	0.0091		

Table 6.2.6d: Output From Log Excess Relative Risk Model

#	Name	Parameter Estimates Estimate	Std.Err.	95% Confidence Bounds Lower	Upper
	Log-linear term 0				
1	%CON	8.335	Aliased		
	Linear term 1				
2	WLML05	1.000	Fixed		
	Log-linear term 1				
3	%CON	-2.911	0.7579	-4.397	-1.426
	EXP(estimate)	0.05441	2.134	0.01232	0.2403

Deviance	=	182.814	df =	1909
Pearson Chi2	=	1191.80		
LR Statistic	=	6.806	df =	1
p	=	0.0091		

6.3 Examining the Effect of Age at First Exposure

6.3.1 Describing The Crude Effect of Age at First Exposure

Members of the study cohort who died of lung cancer following *relevant* exposure to Rn progeny at RH were all aged 20 years or more at the time of their initial exposure at RH. Those who commenced their exposure before the age of 20 years contributed only 6.6% (914 PYRS) of the PYRS in the first AFEG category (<30 years). Isolation of this portion of PYRS by using finer age at first exposure groups (AFEG) categorization made no difference to the analytical findings. Hence, this broad categorization was maintained for all further analyses.

Table 6.3.1a shows the mean ages at first exposure and follow-up (attained age) and mean CE by AFEG; those who commenced their exposure after the age of 40 were observed to have had a higher mean CE. Table 6.3.1b gives the distribution of lung cancer deaths and PYRS by CE and AFE categories. The majority of the study cohort were aged between 20 - 30 years at first exposure and experienced relevant exposures of 1 - 10 cumulative WLM. Most lung cancer deaths occurred amongst those who were aged between 30 - 40 years at initial exposure.

Results of the crude risk of lung cancer mortality by CE and AFE categories are presented in table 6.3.1c. The AFE specific crude mortality rates and crude *RR*s increased with AFE and demonstrated a statistically significant heterogeneity (LRT p = 0.0004). Statistically significant heterogeneity was also seen in crude estimates of risk among levels of interaction (LRT p = 0.0004), indicating that AFE was a significant modifier of the crude exposure-response relationship. The crude risk of mortality tended to increase with AFE and exposure. However, once adjusted for potential confounders, these trends were considerably changed, indicating that these findings were merely an artifact of confounding.

Table 6.3.1a-c: Cohort Characteristics by Cumulative Exposure and Age at First Exposure Categories

Table 6.3.1a: Mean Age and Cumulative Exposure

Characteristic	Age at First Exposure (Years)			
Mean:	**<30**	**30-<40**	**40->**	**Overall**
Age at First Exposure (Years)	24	34	45	29
Age at Follow-up (Years)	42	51	61	47
Cumulative Exposure (WLM)	7.8	7.2	8.2	7.7

Table 6.3.1b: Lung Cancer Deaths and PYRS

(Number of PYRS rounded to the nearest integer are shown in brackets)

Exposure (WLM)	Age at First Exposure (Years)			
	<30	**30-<40**	**40->**	**Overall**
0<1	2 (2749)	5 (1167)	1 (619)	8 (4536)
1-<10	2 (7975)	4 (4315)	4 (1629)	10 (13920)
10-<40	3 (2674)	5 (1464)	2 (403)	10 (4542)
40->	1 (399)	1 (77)	2 (105)	4 (582)
Overall	8 (13799)	15 (7025)	9 (2758)	32 (23583)

Table 6.3.1c: Crude Estimates of the *RR* of Lung Cancer Mortality

* Reference Category for *RR*)

Exposure (WLM)	Age at First Exposure (Years)			
	<30	**30-<40**	**40->**	**Overall**
0<1	1*	5.9	2.2	1*
1-<10	0.3	1.3	3.4	0.4
10-<40	1.5	4.7	6.8	1.2
40->	3.4	17.8	26.0	3.9
Overall	1*	3.7	5.6	

6.3.2 Examining the Effect of Age at First Exposure Using Multiplicative Models

I. Examining The Effect of Confounding

The confounding effects of potential confounders, in particular, attained age and calendar year of observation (which are equivalent to birth cohort) were examined. Attained age and calendar year of observation were found to be independent confounders of the relationship between AFE and lung cancer mortality. However, after adjusting for the confounding effect of either one of these factors, the remaining factor was not found to have a statistically significant confounding effect. Though the joint effect of confounders was not significantly higher than their individual confounding effects, in keeping with accepted standards for the analyses of cohort studies in epidemiology and the reasons outlined in chapter 5, both these factors will be controlled for in all further analyses of the modifying effect of AFE; this examination of confounding effects is only undertaken to demonstrate the role of confounding.

When compared with patterns observed in the crude estimates of *RR*, after controlling for confounding, the heterogeneity among risk estimates for AFE categories was no longer statistically significant (LRT p = 0.3286). Furthermore, contrary to the increasing trend observed in AFEG specific crude estimates of risk, estimates of risk declined with AFE in all other than the highest exposure category. Based on this categorical evaluation therefore, AFE was not seen as a significant modifier of the exposure-response relationship after controlling for confounders.

II. Examining the Effect of Age at First Exposure

Based on a log-linear model, after removing the effect of cumulative relevant Rn exposure, age at first exposure to Rn did not have a statistically significant effect on lung cancer mortality (p = 0.6717). After simultaneous adjustment for both main effects, the main effect of age at first exposure became more pronounced and the exposure effect less so. The *RR* in the highest interaction category (40 or older at first exposure and exposed to over 40 WLM) also dropped considerably. This could be a reflection of the fact that those exposed to the highest levels of radiation were older than others when they commenced their exposure at RH. This corresponds to the experience of professional underground miners who were known to have been older than other categories of underground workers. Their increased risk is therefore contributed to by their increased age at first exposure. This may also contribute to the increasing trend in risk estimates with increasing AFE categories, seen among those in the highest category of exposure. In all other categories of exposure, risk seems to decline with AFE. Overall too, risk declines with AFE categories. However, this pattern of declining risk with increasing AFE may not necessarily be reflected when continuous AFE is modelled as a smooth parameter. These findings are all summarized in tables 6.3.2a and 6.3.2b.

Table 6.3.2a: Unconfounded Marginal and Interactive Relative Risks by Cumulative Exposure and Age at First Exposure Categories

Exposure (WLM)	Age at First Exposure (Years)			Overall	
	<30	30-<40	40->	Unadjusted	Adjusted [a]
0<1	1[d] (Fixed)	0.5 (0.7 - 3.2)	0.1 (0.0 - 1.1)	1[b] (Fixed)	1[b] (Fixed)
1-<10	0.2 (0.0 - 1.4)	0.1 (0.0 - 0.6)	0.1 (0.0 - 1.1)	0.4 (0.1 - 1.0)	0.4 (0.1 - 0.9)
10-<40	0.6 (0.1 - 3.9)	0.2 (0.0 - 1.7)	0.3 (0.0 - 3.9)	1.1 (0.4 - 2.8)	1.0 (0.4 - 2.6)
40->	0.5 (0.0 - 6.6)	0.7 (0.1 - 9.8)	6.0 (0.2 - 162.0)	3.4 (1.0 - 11.6)	2.7 (0.7 - 10.2)
Overall Unadjusted	1[c] (Fixed)	0.5 (0.1 - 1.5)	0.3 (0.1 - 1.7)		
Overall Adjusted [a]	1[c] (Fixed)	0.7 (0.2 - 2.6)	0.4 (0.1 - 2.9)		

Note: [a] Simultaneous adjustment for both main effects WLMG and AFEG
[b] Baseline category for marginal *RR* estimates among WLMG categories
[c] Baseline category for marginal *RR* estimates among AFEG categories
[d] Baseline category for obtaining interaction *RR*s

III. Describing the Nature of Interactions and Marginal Effects

Estimates of *RR* obtained by modelling the interaction term as categorical variables are presented in table 6.3.2a to provide more detailed description of the pattern of interaction in each cross-classification category. Estimates of overall effect were obtained by modelling WLMG and AFEG individually and jointly. Estimates of individual main effects are referred to as 'unadjusted' and presented in the column and row titled 'overall unadjusted' estimates in this table. Estimates obtained after simultaneous adjustment for both main effects WLMG and AFEG are referred to as 'adjusted' estimates and presented under this title in the final column and row of this table.

The marginal effects of cumulative Rn exposure after accounting for the effect of age at first exposure as a categorical variable showed a steady increase in *RR* with increasing exposure categories of 10 WLM or more; both these *RR*s were statistically significant at the 5% level. Those exposed from 1 upto 10 WLM of cumulative Rn exposure at RH were not found to be at any significantly greater risk than those exposed to 1 WLM or less. After accounting for the main effect of age at first exposure, there was still an increasing trend in the exposure-response relationship.

Compared to the baseline AFEG category (<30 years), before accounting for the main effect of exposure, the marginal *RR* of Rn related lung cancer mortality was halved amongst those aged between 30-40 at initial exposure, and a third amongst those who were aged 40 years or more at initial exposure. After accounting for the main effect exposure however, these risk estimates changed to declines of 30% and 60%, respectively.

IV. Comparison of Effect Modification Characterizations

First order interaction terms between cumulative Rn exposure and age at first exposure after removing the main effects of WLMG and AFEG, were examined in three stages. Results of fitting these three interaction models are provided in table 6.3.2b. Tests of significance for nested models are all based on the improvement of model deviance due to the added effect of the interaction term after removing the main effects. Nesting of models is represented by indentations in table 6.3.2b.

In the first interaction model, a common exposure-response slope was fitted to these data based on PYRS weighted WLM and AFE. The deviance corresponding to this model was 173.59 and LRT results showed that the contribution of this single interaction term (common exposure-response slope) based on 1 df was not statistically significant ($p = 0.2062$).

Separate slopes were then fitted for each AFEG after removing the main effects of AFEG and WLMG and the first order linear interaction term AFE*WLM. The LRT result of fitting separate slopes for each AFEG was not statistically significant ($p = 0.1860$) and the single slope coefficient appeared to provide an adequate fit at this level.

In the second stage of fitting interaction models, separate slopes were first fitted for each exposure category after removing the main effects and the first order interaction term based on categorical exposure and continuous age at first exposure. The purpose of fitting this effect was to examine the modifying effect of AFE for a given level of exposure. This was done by including the categorical interaction term in the existing model. Results of this fit were not statistically significant ($p = 0.1524$) and failed to show any further evidence of effect modification.

In the third stage of modelling interactions, the categorical interaction effect was found to be highly significant (p = 0.0151), when it was included in the model after fitting separate slopes for each AFEG category with exposure treated as a continuous variable. This finding implied that within a given category of age at first exposure, the risk was found to be significantly heterogeneous - increased with exposure - and could therefore, not be summarized by a single parameter. There remained a significant exposure-response effect even after accounting for the main effects of exposure and age at first exposure, and the first order interaction.

Finally, an interaction model with separate slopes for each AFEG, WLMG interaction was fitted after removing the main effects WLMG and AFEG. Though the statistical significance of this interaction effect was marginal (p = 0.1096), it showed some indication of effect modification after accounting for main effects.

It may be concluded that the above findings based on multiplicative log-linear models fail to confirm age at first exposure as a significant effect modifier after accounting for the main effects; the interaction effect seen here could have been due to chance alone and is further examined in sections 6.3.3-6.3.6 using parametric models of variations in ERR/WLM with age at first exposure.

Table 6.3.2b: Results of Fitting Log-Linear Models to Examine the Modifying Effect of Age at First Exposure

Source of Variation	Deviance	DF	Pearson χ^2	Likelihood Ratio Test		
				Statistic	DF	p - Value
NULL	189.62	1910	1392.64			
WLMG	175.99	1907	1016.17	13.6300	3	
AFEG	175.19	1905	897.21	0.7959	2	0.6717^{w}
WLM*AFE	173.59	1904	984.58	1.5980	1	0.2062^{wa}
WLMG*AFEG	164.81	1898	1042.10	8.786	6	0.1860^{wal}
WLMG*AFE	169.06	1900	938.53	4.5330	4	0.3386^{wal}
WLMG*AFEG	159.67	1894	925.41	9.3970	6	0.1524^{wga}
WLM*AFEG	168.68	1901	857.45	4.9270	3	0.1772^{wal}
WLMG*AFEG	152.90	1895	749.39	15.7600	6	0.0151^{wag}
WLMG*AFEG	164.81	1899	1046.75	10.3800	6	0.1096^{wa}

NOTE: a LRT based on model with WLMG main effect removed.
wa LRT based on model with WLMG and AFEG main effects removed.
wal LRT based on model with WLMG, AFEG and WLM*AFE removed.
wag LRT based on model with WLMG, AFEG and WLM*AFEG removed.

Parameter estimates of ERR/WLM were also obtained by fitting common and separate slopes using log-linear models both before and after removal of the main effects (tables 6.3.2c - 6.3.2f). Results of fitting a common slope to the interaction term (tables 6.3.2c - 6.3.2d) showed little difference in slope estimates before and after removal of main effects. The slope coefficient of 1.001 indicated that the ERR increased by 0.1% per WLM per year of first exposure. When separate slopes were fitted without main effects being removed, estimates of ERR/WLM obtained were 1.026, 1.008 and 1.038 respectively for each increasing category of AFEG (table 6.3.2e). These estimates were not significantly heterogeneous (p = 0.3856) and did not provide a significantly better fit than the common slopes model. However, these estimates of ERR/WLM changed considerably after the removal of main effects; the resulting AFEG specific ERR/WLM estimates were 0.9674, 0.9507 and 1.012 respectively (table 6.3.2f). These estimates were significantly heterogeneous at the 10% level of significance (p = 0.0852). It was also seen that after accounting for the overall effect of exposure and AFE, the direction of the ERR/WLM changed in the first two categories of AFEG, with their magnitude becoming less than 1.0. Confidence intervals for each of these risk estimates show however that these estimates were not significantly lower than the constant *RR* of 1.

It could therefore be concluded that a mild degree of effect modification existed due to AFE, with the exposure-response relationship declining for those who commenced exposure before the age of 40, and increasing among those who commenced their exposure at later ages. Furthermore, after adjusting for the overall effect of exposure, the risk of lung cancer mortality was seen to decrease with AFE i.e. for a fixed level of exposure, risk decreases with AFE.

Table 6.3.2c-d: Output From Log-Linear Relative Risk Modelling to Examine the Modifying Effect of Age at First Exposure Common Slope Models Before and After Removing Main Effects

Table 6.3.2c: Before Removing Main Effects

#	Name	Estimate	Std. Error	95% Confidence Bounds Lower	Upper
Log-linear term 0					
1	%CON	-0.2230	Aliased		
2	WLML05	-0.4343E-02	0.4431E-01	-0.9120E-01	0.8251E-01
	EXP(estimate)	0.9957	1.045	0.9128	1.086
3	WLML05 * AFE	0.9333E-03	0.1202E-02	-0.1423E-02	0.3289E-02
	EXP(estimate)	1.001	1.001	0.9986	1.003

Deviance	=	180.376	df = 1908
Pearson Chi2	=	1242.42	
LR statistic	=	0.6443	df = 1
p	=	0.4222	

Table 6.3.2d: After Removing Main Effects

#	Name	Parameter Estimates Estimate	Std. Error	95% Confidence Bounds Lower	Upper
Log-linear term 0					
1	%CON	1.147	Aliased		
3	WLMG4_2	-1.141	0.4883	-2.099	-0.1843
	EXP(estimate)	0.3194	1.630	0.1226	0.8317
4	WLMG4_3	-0.5247	0.6294	-1.758	0.7089
	EXP(estimate)	0.5917	1.877	0.1723	2.032
5	WLMG4_4	-0.3182	1.242	-2.753	2.117
	EXP(estimate)	0.7275	3.464	0.6372E-01	8.306
7	AFEG3_2	-0.5483	0.6257	-1.775	0.6781
	EXP(estimate)	0.5780	1.870	0.1696	1.970
8	AFEG3_3	-0.9883	0.9611	-2.872	0.8954
	EXP(estimate)	0.3722	2.614	0.5659E-01	2.448
9	WLML05 * AFE	0.6329E-03	0.4598E-03	-0.2684E-03	0.1534E-02
	EXP(estimate)	1.001	1.000	0.9997	1.002

Deviance	=	173.594	df = 1904
Pearson Chi2	=	984.279	
LR statistic	=	1.598	df = 1
p	=	0.2062	

Table 6.3.2e-f: Output From Log-Linear Relative Risk Modelling to Examine the Modifying Effect of Age at First Exposure Separate Slopes Models Before and After Removing Main Effects

Table 6.3.2e: Before Removing Main Effects

#	Name	Parameter Estimates Estimate	Std. Error	95% Confidence Bounds Lower	Upper
	Log-linear term 0				
1	%CON	-0.1952	Aliased		
2	AFEG3_1 * WLML05	0.2543E-01	0.1327E-01	-0.5799E-03	0.5143E-01
	EXP(estimate)	1.026	1.013	0.9994	1.053
3	AFEG3_2 * WLML05	0.7625E-02	0.2101E-01	-0.3356E-01	0.4881E-01
	EXP(estimate)	1.008	1.021	0.9670	1.050
4	AFEG3_3 * WLML05	0.3747E-01	0.1102E-01	0.1587E-01	0.5908E-01
	EXP(estimate)	1.038	1.011	1.016	1.061

Deviance	= 179.115	df =	1907
Pearson Chi2	= 1364.22		
LR statistic	= 1.906	df =	2
p	= 0.3856		

Table 6.3.2f: After Removing Main Effects

#	Name	Parameter Estimates Estimate	Std. Error	95% Confidence Bounds Lower	Upper
	Log-linear term 0				
1	%CON	1.297	Aliased		
3	WLMG4_2	-1.038	0.4875	-1.993	-0.8206E-01
	EXP(estimate)	0.3543	1.628	0.1363	0.9212
4	WLMG4_3	0.4623	0.7050	-0.9195	1.844
	EXP(estimate)	1.588	2.024	0.3987	6.322
5	WLMG4_4	2.344	1.741	-1.069	5.757
	EXP(estimate)	10.43	5.705	0.3435	316.5
7	AFEG3_2	-0.4517	0.7451	-1.912	1.009
	EXP(estimate)	0.6365	2.107	0.1478	2.742
8	AFEG3_3	-1.456	1.013	-3.440	0.5293
	EXP(estimate)	0.2333	2.753	0.3205E-01	1.698
9	AFEG3_1 * WLML05	-0.3318E-01	0.3373E-01	-0.9928E-01	0.3293E-01
	EXP(estimate)	0.9674	1.034	0.9055	1.033
10	AFEG3_2 * WLML05	-0.5056E-01	0.3676E-01	-0.1226	0.2149E-01
	EXP(estimate)	0.9507	1.037	0.8846	1.022
11	AFEG3_3 * WLML05	0.1232E-01	0.1903E-01	-0.2499E-01	0.4962E-01
	EXP(estimate)	1.012	1.019	0.9753	1.051

Deviance	= 168.668	df =	1902
Pearson Chi2	= 858.777		
LR statistic	= 4.926	df =	2
p	= 0.0852		

6.3.3 Examining The Modifying Effect of Age at First Exposure Using Linear Additive Excess Relative Risk Models

Examination of the modifying effect of AFE commenced by fitting the linear ERR model with a common slope model, where the exposure-response was assumed to be constant across the range of AFE values. Results of fitting this model are presented in table 6.3.3a. The ERR of lung cancer mortality increased by 0.15 % per WLM and increasing year of AFE (i.e. per WLM*AFE), after accounting for the effect of exposure alone. However, this modifying effect was not statistically significant (p = 0.7255). Therefore, based on this linear ERR model, AFE did not have a significant continuous modifying effect on the exposure-response relationship.

Independent exposure covariates were then fitted for each age at AFE category (AFEG), using the separate slopes model (3b). Results of this fit are given in table 6.3.3b. Estimates of ERR/WLM thus obtained were 7.0%, 1.3% and 10.1% respectively for each AFEG (table 6.3.3c). Likelihood Ratio Bounds of these estimates showed a marginally significant increase in the ERR/WLM amongst those aged between 30-40 years at first exposure; those in the highest AFEG (>40 years at first exposure) were at significantly higher risk with each increasing unit of exposure. The ERR of lung cancer mortality was lowest amongst those who commenced their exposure between the ages of 30-40 years; compared to them, those aged below 30 years at first exposure were at 5.6% greater risk, and those aged 40 years or more at first exposure were at 8.9% greater risk. The proportional change in ERR per WLM was 5.2% and 7.7% respectively for these groups. The *RR* of these categories were 1.06 and 1.09. The heterogeneity of these estimates was not statistically significant at the 5% level of significance. Age at first exposure considered as a categorical variable could therefore not be identified as a statistically significant modifier (LRT: p = 0.3593) of the effect of Rn related lung cancer mortality based on this linear additive ERR model.

Table 6.3.3a-c: Results of Linear Additive Excess Relative Risk Modelling to Examine the Modifying Effect of Age at First Exposure

Table 6.3.3a: Common Slope Model

Parameter Summary Table

#	Name	Estimate	Std.Err.	Score	Status
	Log-linear term 0				
1	%CON	-0.9239	0	0	Aliased
	Linear term 1				
2	WLML05	0.002090	0.1866	-0.000892	Free
3	AFE * WLML05	0.001532	0.005777	-0.00282	Free

Deviance	=	182.691	df = 1908
Pearson Chi2	=	1204.41	
LR statistic	=	0.1233	df = 1
p	=	0.7255	

Table 6.3.3b: Separate Slopes Model

Parameter Summary Table

#	Name	Estimate	Std.Err.	Score	Status
	Log-linear term 0				
1	%CON	-0.4136	0	0	Aliased
	Linear term 1				
2	AFEG3_1 * WLML05	0.06974	0.06513	0.00258	Free
3	AFEG3_2 * WLML05	0.01340	0.03182	0.000319	Free
4	AFEG3_3 * WLML05	0.1011	0.1069	-0.00258	Free

Deviance	=	180.767	df = 1907
Pearson Chi2	=	1386.71	
LR statistic	=	2.047	df = 2
p	=	0.3593	

Table 6.3.3c: Parameter Estimates and LR Bounds

ERR Function	Parameter	Parameter Estimates ERR	95% CB
$b\,w$	b	5.4%	(1 - 17)%
$b_1 w + b_2 wz$	b_1	0.21%	
	b_2	0.15%	(0 - 0.5)%
$b_i\, wz_C$	b_1	7.0%	(0 - 32)%
	b_2	1.3%	(0 - 12)%
	b_3	10.1%	(1 - 46)%

Departure from linearity was examined by adding an additional interaction term with a quadratic AFE component to this model. This linear-quadratic model in AFE had a model deviance of 181.99. The LRT contribution of the quadratic term in AFE failed to detect a significant departure from linearity at the 5% level of significance (p = 0.6969). The goodness of fit results of this linear-quadratic modelling are presented in table 6.3.3d.

Table 6.3.3d: Results of Fitting Models to Examine Departures from Linearity to Examine the Modifying Effect of Age at First Exposure

Source of Variation	Deviance	DF	Pearson c^2	Likelihood Ratio Test		
				Statistic	DF	p - Value
NULL	189.62	1910	1392.64			
WLM	182.81	1909	1191.79			
WLM*AFE	182.69	1908	1204.24	0.1234	1	0.7254[w]
WLM*AFE^2	181.99	1907	1286.12	0.6969	1	0.4038[wl]

NOTE: w LRT based on model with WLM main effect removed.

wl LRT based on model with WLM, and WLM*AFE removed.

6.3.4 Examining the Modifying Effect of Age at First Exposure Using a Log Excess Relative Risk Model

When AFE was modelled as an exponential function in the linear ERR model, the model was found to have a better fit than the previous model. Results of this fit, given in table 6.3.4a, show that for a fixed level of exposure, the ERR was found to multiply by 1.04 with each increasing year of age at first exposure. However, this multiplicative ERR was not significantly different from 1 and AFE modelled as a continuous exponential variable was not found to be a significant modifier of the exposure-response relationship ($p = 0.6764$).

Once again, as seen in the previous linear additive models, fitting separate exponential slopes for each AFEG category resulted in marked changes to the patterns of ERR (table 6.3.4b). Estimates of ERR parameter estimates were almost identical to those obtained with linear additive models; estimates of ERR/WLM were 7.0%, 1.3% and 10.1% respectively in each of the AFE categories. The goodness of fit characteristics of the log ERR model were also extremely similar to those in the linear additive ERR model. Therefore, it was concluded that the log ERR model did not offer any great advantage over the simple linear model.

Table 6.3.4a-b: Output From Log Excess Relative Risk Modelling to Examine the Modifying Effect of Age at First Exposure

Table 6.3.4a: Common Slope Model

#	Name	Parameter Estimates: Estimate	Std. Error	95% Confidence Bounds: Lower	Upper
	Log-linear term 0				
1	%CON	8.340		Aliased	
	Linear term 1				
2	WLML05	1.000		Fixed	
	Log-linear term 1				
3	%CON	-4.300	3.844	-11.83	3.235
	EXP(estimate)	0.1357E-01	46.72	0.7253E-05	25.40
4	AFE	0.3962E-01	0.1091	-0.1742	0.2535
	EXP(estimate)	1.040	1.115	0.8401	1.289

Deviance	=	182.640	df = 1908
Pearson Chi2	=	1213.04	
LR statistic	=	0.1742	df = 1
p	=	0.6764	

Table 6.3.4b: Separate Slopes Model

#	Name	Parameter Estimates: Estimate	Std. Error	95% Confidence Bounds: Lower	Upper
	Log-linear term 0				
1	%CON	16.57		Aliased	
	Linear term 1				
2	WLML05	1.000		Fixed	
	Log-linear term 1				
3	AFEG3_1	-2.662	0.9337	-4.492	-0.8324
	EXP(estimate)	0.6979E-01	2.544	0.1119E-01	0.4350
4	AFEG3_2	-4.312	2.374	-8.965	0.3409
	EXP(estimate)	0.1341E-01	10.74	0.1278E-03	1.406
5	AFEG3_3	-2.292	1.057	-4.363	-0.2201
	EXP(estimate)	0.1011	2.878	0.1274E-01	0.8025

Deviance	=	180.767	df = 1907
Pearson Chi2	=	1386.61	
LR statistic	=	2.047	df = 2
p	=	0.3593	

6.3.5 Examining Variations in ERR/WLM With Age at First Exposure Using Smooth Parametric Functions

Finally, in the analyses of AFE as a modifier of the exposure-response effect, variations of the ERR/WLM were modelled using two smooth parametric functions viz., the exponential function and the power function. As expected, findings from the exponential function fit were similar to the model fitted in section 6.3.4; the only difference being that the exposure parameter was now freed and estimated by the model. The results of fitting this model are presented in table 6.3.5a and correspond to findings in table 6.3.4a.

The exponential model was also fitted after centering AFE at 30 years (table 6.3.5b). This changed the exposure coefficient to 0.0456, but did not change the model fit or the exponential parameter. With this fit, above 30 years of age at first exposure the *RR* is multiplied by a factor of 1.04 for each increasing AFE, and is reduced below the age of 30.

The power model was fitted by fitting the log of the AFE as an exponential term in EPICURE (EPICURE 1992) and using its untransformed parameter estimate as the maximum likelihood estimate of the power parameter. Results of fitting this model presented in table 6.3.5c show a maximum likelihood estimate of a power of 0.6168 with extremely wide confidence intervals (-6.5 - 7.8). The exposure coefficient from this model was 0.0058. The deviance of this model was 182.725 which was slightly higher than that of all other models examined thus far. Examination of the role of AFE as an effect modifier was therefore, concluded at this point.

Table 6.3.5a: Output From Modelling the ERR/WLM as an Exponential Function to Examine the Modifying Effect of Age at First Exposure

#	Name	Estimate	Std. Error	95% Confidence Bounds Lower	Upper
Log-linear term 0					
1	%CON	-0.5112		Aliased	
Linear term 1					
2	WLML05	0.1350E-01	0.5197E-01	-0.8836E-01	0.1154
Log-linear term 1					
3	AFE	0.3970E-01	0.1092	-0.1744	0.2538
	EXP(estimate)	1.040	1.115	0.8400	1.289

Deviance	=	182.640	df = 1908
Pearson Chi2	=	1213.06	
LR statistic	=	0.1742	df = 1
p	=	0.6764	

Table 6.3.5b: Modelling the ERR/WLM as an Exponential Function With a Change of Location Transformation for the AFE (AFE30 = AFE - 30)

#	Name	Estimate	Std. Error	95% Confidence Bounds Lower	Upper
Log-linear term 0					
1	%CON	-0.4513		Aliased	
Linear term 1					
2	WLML05	0.4456E-01	0.4113E-01	-0.3606E-01	0.1252
Log-linear term 1					
3	AFE30	0.3963E-01	0.1091	-0.1742	0.2535
	EXP(estimate)	1.040	1.115	0.8401	1.289

Deviance	=	182.640	df = 1908
Pearson Chi2	=	1213.04	
LR statistic	=	0.1742	df = 1
p	=	0.6764	

Table 6.3.5c: Output From Modelling the ERR/WLM as a Power Function to Examine the Modifying Effect of Age at First Exposure

#	Name	Parameter Estimates Estimate	Std. Error	95% Confidence Bounds Lower	Upper
	Log-linear term 0				
1	%CON	-0.4689	Aliased		
	Linear term 1				
2	WLML05	0.5798E-02	0.7413E-01	-0.1395	0.1511
	Log-linear term 1				
3	LAFE	0.6168	3.642	-6.522	7.755

Deviance	=	182.725	df = 1908
Pearson Chi2	=	1198.79	
LR statistic	=	0.8904E-01	df = 1
p	=	0.7654	

6.3.6 Summary of Findings on the Role of Age at First Exposure

The crude risk of lung cancer mortality was seen to increase with AFE. However, this trend was found to be an artifact of confounding and disappeared when controlled for age at observation and calendar time of observation.

Of all the exposure-modifier-response models examined, multiplicative log-linear models provided the best fit; AFE could not be identified as a statistically significant effect modifier which was possibly due to the limited variation observed in AFE amongst the RH cohort. However, there was a indication of increasing ERR/WLM with age at first exposure.

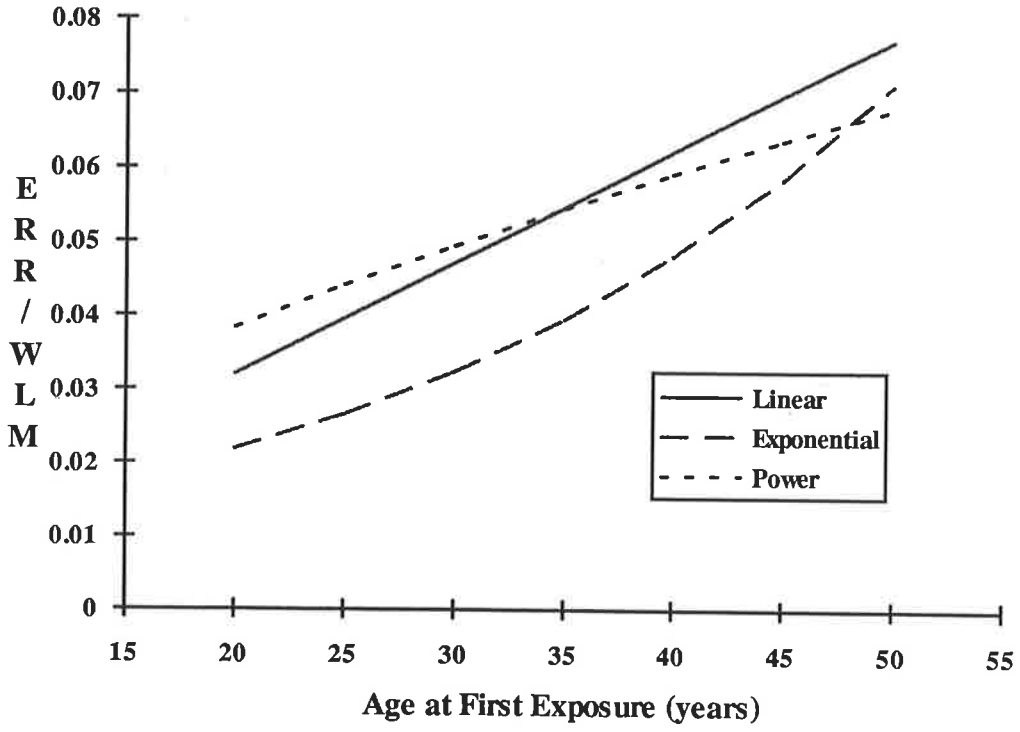
From multiplicative log-linear modelling a common overall slope coefficient of 1.001 was found to provide an adequate explanation of the interaction effect between continuous measures of AFE and exposure when compared with fitting separate slopes for each AFEG and WLMG cross-classification categories. However, it was found that within each AFEG category, the separate slope coefficients for each WLMG category could not be summarized into a single slope coefficient viz., 0.0.9674, 0.9507 and 1.012. The main contributor to this heterogeneity was identified as the exceptionally high *RR* coefficient of 6.0 in the highest cross-classification category. This heterogeneity was however not strong enough to indicate any effect modification overall.

Findings from mathematical modelling to examine variations in ERR/WLM with AFE are summarized in tables 6.3.6. Modifier-response relationships for a given level of exposure obtained from the various models fitted, are plotted in figure 6.3.6. All models fitted showed a clear pattern of increasing ERR with AFE, but no significant heterogeneity could be detected between the AFEG specific estimates of ERR/WLM. The ERR/WLM was lowest amongst those aged between 30-40 years at first exposure. No significant departures from linearity could be detected. For a given level of exposure, ERR/WLM was highest amongst those in the highest AFE category (i.e., those commencing exposure after 40 years of age).

Table 6.3.6: Summary of Results From Fitting Models to Examine the Modifying Effect of Age at First Exposure

ERR Function	Parameter Estimates			Model Fit		LR Test		
	Par	ERR	95% CB	Dev	DF	χ^2	DF	p
Linear:								
βw	β	5.4%	(1 - 17)%	182.8	1909			
$\beta_1 w + \beta_2 wz$	β_1	0.21%		182.7	1908	0.123	1	0.7255
	β_2	0.15%						
$\beta_i wz_c$	β_1	6.97%	(0 - 32)%	180.8	1907	2.047	2	0.3593
	β_2	1.34%	(0 - 12)%					
	β_3	10.1%	(1 - 46)%					
Exponential:								
$\beta_1 w\, e^{\beta_2+\gamma z}$	β_1	1	Fixed					
	e^{β_2}	0.01	(0.0 - 25.4)					
	e^{γ}	1.04	(0.8 - 1.3)					
$\beta_1 w\, e^{\gamma_i z_c}$	β_1	1	Fixed	180.8	1907	2.047	2	0.3593
	$e^{\gamma 1}$	0.07	(0.0 - 0.4)					
	$e^{\gamma 2}$	0.01	(0.0 - 1.4)					
	$e^{\gamma 3}$	0.10	(0.0 - 0.8)					
Exponential:								
$\beta w\, e^{\gamma z}$	β	0.01		182.6	1908	0.174	1	0.6764
	e^{γ}	1.04	(0.8 - 1.3)					
$\beta w\, e^{\gamma(z-30)}$	β	0.04		182.6	1908	0.174	1	0.6764
	e^{γ}	1.04	(0.8 - 1.3)					
Power:								
$\beta w\, z^{\gamma}$	β	0.006		182.7	1908	0.089	1	0.7654
	γ	0.62	(-6.5 - 7.8)					

Figure 6.3.6: Modelling Variations in ERR/WLM With Age at First Exposure



Note:

Risk Functions Fitted: Linear: $ERR/WLM = 0.0021 + 0.0015\ z$

Exponential: $ERR/WLM = 0.01 * 1.04^{z}$

Power: $ERR/WLM = 0.006 * z^{0.62}$

6.4 Examining the Effect of Time Since Last Exposure

6.4.1 Describing The Crude Effect of Time Since Last Exposure

Characteristics of the study cohort identified by categories of cumulative exposure (CE) and time since last exposure (TSLE) are presented in tables 6.4.1a - 6.4.1c. As expected, the mean age at follow-up increased with TSLE category (TSLEG). The lowest TSLEG had the highest level of mean CE and mean CE levels were seen to decline steadily with TSLE categories. The crude risk of lung cancer mortality increased steadily with TSLE, overall and in each CE category with the exception of the 10-40 WLM category. However, this pattern in lung cancer mortality risks was found to be an artifact of confounding, which disappeared after adjustment for confounding effects (table 6.4.2).

6.4.2 Describing the Effect of Time Since Last Exposure Using Multiplicative Models

Lung cancer mortality risk estimates obtained using log-linear models after controlling for the confounding effects of attained age and calendar year of observation are summarized in table 6.4.2. These estimates showed a steady decline in the *RR* of lung cancer mortality with TSLE among those exposed to 10 or more WLM of Rn progeny. A weaker pattern was apparent in the overall estimates; the *RR* estimate in the highest TSLE category (25 years or more) was very slightly higher than that in the middle group (20-25 years) but, had a much wider confidence interval. Simultaneous adjustment for the main effect of TSLE made little difference to the marginal estimates of risk in each exposure category. Overall, *RR* were found to be reasonably homogeneous across TSLE categories.

Table 6.4.1a-c: Cohort Characteristics by Cumulative Exposure and Time Since Last Exposure Categories

Table 6.4.1a: Mean Age and Cumulative Exposure

Characteristic	Time Since Last Exposure (Years)			
Mean:	**<20**	**20-<25**	**25->**	**Overall**
Time Since Last Exposure (Years)	11	22	28	16
Age at Follow-up (Years)	42	52	58	47
Cumulative Exposure (WLM)	8.0	7.6	6.6	7.7

Table 6.4.1b: Lung Cancer Deaths and PYRS

(Number of PYRS rounded to the nearest integer are shown in brackets)

Exposure (WLM)	Time Since Last Exposure (Years)			
	<20	**20-<25**	**24->**	**Overall**
0<1	3 (2946)	2 (847)	3 (742)	8 (4536)
1-<10	1 (8982)	2 (2645)	7 (2293)	10 (13920)
10-<40	4 (3093)	4 (857)	2 (592)	10 (4542)
40->	2 (401)	1 (103)	1 (77)	4 (582)
Overall	10 (15423)	9 (4454)	13 (3706)	32 (23583)

Table 6.4.1c: Crude Estimates of the *RR* of Lung Cancer Mortality

(* Reference Category for *RR*)

Exposure (WLM)	Time Since Last Exposure (Years)			
	<20	**20-<25**	**25->**	**Overall**
0<1	1*	2.3	4.0	1*
1-<10	0.1	0.7	3.0	0.4
10-<40	1.3	4.6	3.3	1.2
40->	4.9	9.5	12.7	3.9
Overall	1*	3.1	5.4	

Table 6.4.2: Estimates of Unconfounded Marginal and Interactive Relative Risk by Cumulative Exposure and Time Since Last Exposure Categories

Exposure	Time Since Last Exposure (Years)			Overall	
(WLM)	<20	20-<25	25->	Unadjusted	Adjusted [a]
0<1	1[d] (Fixed)	0.67 (0.1 - 4.6)	0.89 (0.1 - 6.6)	1[b] (Fixed)	1[b] (Fixed)
1-<10	0.13 (0.0 - 1.2)	0.19 (0.0 - 1.3)	0.54 (0.1 -3.3)	0.37 (0.1 - 1.0)	0.38 (0.1 - 1.0)
10-<40	1.25 (0.3 - 5.8)	1.11 (0.2 - 5.9)	0.51 (0.1 -4.3)	1.1 (0.4 - 2.8)	1.11 (0.4 - 2.8)
40->	3.6 (0.6 - 22.0)	2.47 (0.2 - 26.1)	2.30 (0.2 -30.7)	3.39 (1.0 - 11.6)	3.38 (1.0 -11.6)
Overall Unadjusted	1[c] (Fixed)	0.66 (0.21 - 2.05)	0.69 (0.16 - 2.95)		
Overall Adjusted [a]	1[c] (Fixed)	0.80 (0.3 - 2.5)	0.95 (0.2 - 4.1)		

Note: [a] Simultaneous adjustment for both main effects WLMG and TSLEG
[b] Baseline category for marginal *RR* estimates among WLMG categories
[c] Baseline category for marginal *RR* estimates among TSLEG categories
[d] Baseline category for obtaining interaction *RR*s

Based on this categorical evaluation using a log-linear *RR* model, the interaction between Rn progeny exposure and TSLE was not found to be statistically significant (LRT: $p = 0.5753$) with respect to the outcome of lung cancer mortality after removing the main effects of CE and TSLE. Next, the influence effect of TSLE on the exposure-response relationship (exposure to Rn progeny and the outcome, lung cancer mortality) is further examined using continuous data and smooth parametric representations of the exposure-response relationship.

6.4.3 Examining Variations ERR/WLM With Time Since Last Exposure Using Smooth Parametric Functions

The effect of TSLE on the exposure-response relationship was further examined using other mathematical models, results of which are summarized in table 6.4.3 and figure 6.4.3. All models showed that the ERR/WLM declined with increasing TSLE. However, neither the log-linear *RR* model discussed in section 6.4.2 nor the linear additive ERR model identified any statistically significant effect due to TSLE.

However, when the effect of TSLE was represented as exponential or power functions into ERR models that were linear in exposure, the models showed evidence of highly significant variations in ERR/WLM with TSLE. Exponential modelling resulted in a parameter estimate of 0.72 for TSLE modelled as a continuous variable; this estimate was significantly (p = 0.0375) different from 0 (95% CI: 0.5-1.03). This implied that for a given level of exposure the ERR declined by a factor of 72% for each increasing year since last exposure. A graph of this function showed that the ERR dropped steeply till 10 years after last exposure, and then tapered off to almost zero after the next five years. In interpreting this result, it must be remembered that all analyses were based on 5 year lagged exposure and therefore, these results should strictly be interpreted with the 5 year lag subtracted from the time since last exposure. In this sense, the ERR will be seen to continue to drop steeply till about 5 years after relevant exposure and taper off to zero from about 10 years after relevant exposure. The flatness of the curve beyond this point could account for the lack of heterogeneity observed in ERR estimates for the TSLE categorization chosen. Evaluation of the effect of TSLE as a categorical variable using linear additive and log ERR functions also showed that ERR/WLM declined steadily with increasing TSLE categories. These estimates however, were not sufficiently heterogeneous to establish a statistically significant interactive effect. The lack of evidence supportive of an interaction effect from these categorical analyses may be due to the choice of the specific categorizations, as explained above.

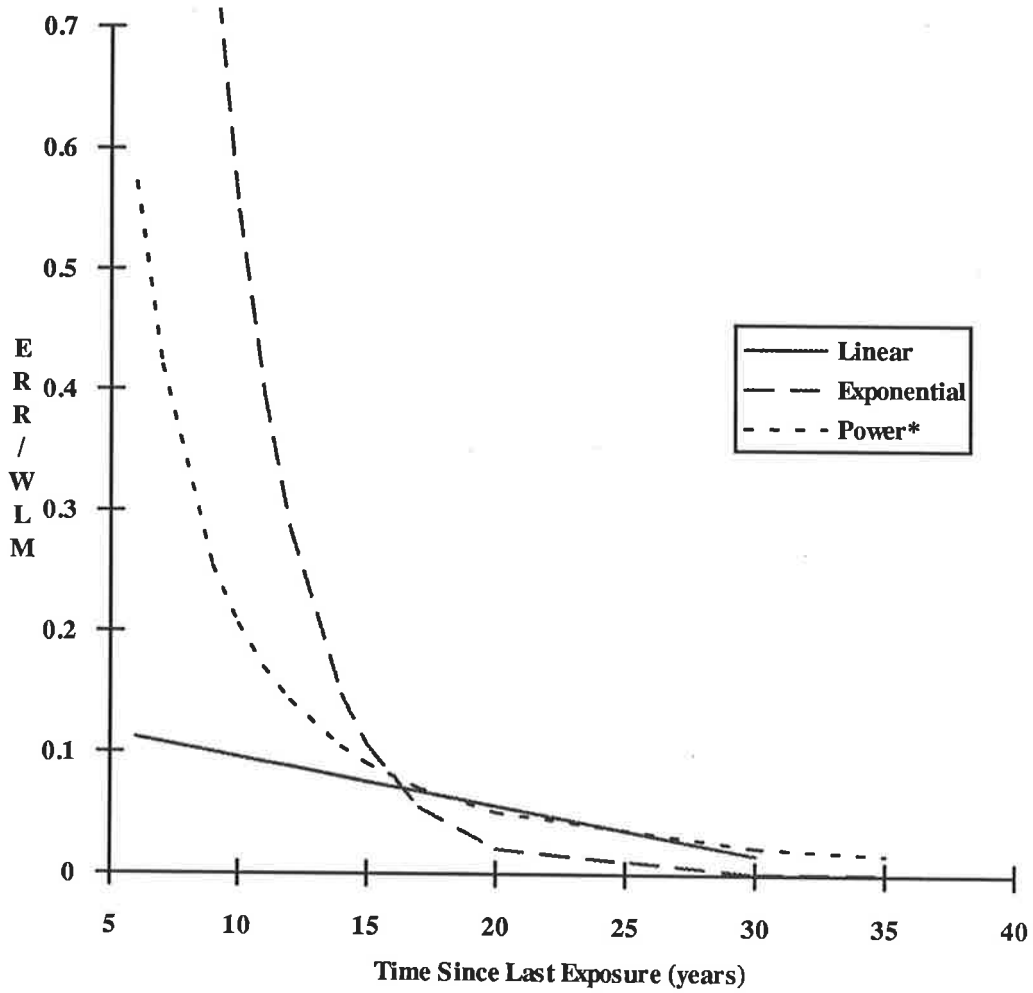
The power model provided an estimate of -2.03; meaning that ERR declined inverse quadratically each increasing year since last relevant exposure. Since a log transformation of TSLE was used in fitting the power model, a further change of location transformation was made for each observation with 0.1 being added to TSLE when modelled continuously. This power model provided the *best fit* in the examination of variations of ERR/WLM and the effect of TSLE. The power parameter estimated from this model was significantly different from 1 (95% CI: -4.7 to -0.6).

Plots of the functions of variations in ERR/WLM with TSLE (figure 6.4.3) show that the exponential function tends to overestimate the ERR/WLM for recent exposures (experienced 5-15 years prior to observation).

Table 6.4.3: Summary of Results From Fitting Models to Examine the Effect of Time Since Last Exposure

ERR	Parameter Estimates			Model Fit		LR Test		
Function	Par	ERR	95% CB	Dev	DF	χ^2	DF	p
Linear:								
βw	β	5.4%	(1 - 17)%	182.8	1909			
$\beta_1 w + \beta_2 wz$	β_1	13.6%		181.8	1908	1.032	1	0.3097
	β_2	-0.4%						
$\beta_i wz_c$	β_1	8.6%		182.1	1907	1.700	2	0.7048
	β_2	5.7%						
	β_3	2.7%						
Exponential:								
$\beta_1 w\, e^{\beta_2+\gamma z}$	β_1	1	Fixed	178.5	1908	4.326	1	**0.0375**
	e^{β_2}	14.63						
	e^{γ}	0.72	(0.5 - 1.03)					
$\beta w\, e^{\gamma z}$	β	14.56		178.5	1908	4.326	1	**0.0375**
	e^{γ}	0.72	(0.5 - 1.03)					
$\beta_1 w\, e^{\gamma_i z_c}$	β_1	1		182.1	1907	0.700	2	0.7048
	$e^{\gamma 1}$	8.6%	(0.0 - 0.5)%					
	$e^{\gamma 2}$	5.7%	(0.0 - 0.4)%					
	$e^{\gamma 3}$	2.7%	(0.0 - 0.7)%					
Power:								
$\beta w\,(z+0.1)^{\gamma}$	β	22.44		172.5	1908	10.34	1	**0.0013**
	γ	-2.03	(-4.7)-(-0.6)					
$\beta w\,(z+0.5)^{\gamma}$	β	137.9	-824 - 1110	172.7	1908	10.14	1	**0.0015**
	γ	-2.63	(-5.0)-(-0.2)					

Figure 6.4.3: Modelling Variations in ERR/WLM With Time Since Last Exposure



Note: * Best Fitting Model

Risk Functions Fitted: Linear: $ERR/WLM = 0.136 - 0.004\ z$

Exponential: $ERR/WLM = 14.63 * 0.72^{z}$

Power: $ERR/WLM = 22.44\ (z+0.1)^{-2.03}$

6.4.4 The Role of Calendar Time in the Association Between Time Since Last Exposure and the Exposure-Response Relationship

Previous analyses of the RH cohort in which only attained age was controlled for as a confounder, showed no evidence of TSLE being a modifier of the Rn exposure-response relationship. Similar analyses performed in this present work, demonstrated a marked difference in findings between models that controlled for attained age alone, and those that jointly controlled for attained age and calendar time (year) of observation. For example using the log ERR model, the LRT statistic for the effect of TSLE on CE after controlling for age alone was not statistically significant (LRT statistic = 0.0135; df = 1; $p = 0.9075$). The results reported in table 6.4.3 show highly significant effect modification after controlling for attained age and calendar year of observation. Likewise, the LRT for the effect of TSLE based on a power model after controlling for attained age alone showed no significant evidence of effect modification ($p = 0.1335$). The relationship between TSLE and calendar time (synonymous with birth cohort) is obvious; TSLE increases with calendar time once an individual completes his exposure; for short exposure durations, TSLE acts as a surrogate for calendar time. Levels of exposure were also strongly associated with calendar time. Therefore, it is understandable that unless exposure durations were short, calendar time could influence the association between TSLE and the exposure-response relationship.

6.5 Examining the Effect of Duration of Exposure

6.5.1 Describing The Crude Effect of Duration of Exposure

Cohort characteristics cross-tabulated by categories of CE and duration of exposure are presented in tables 6.5.1a - 6.5.1c. Mean age at follow-up and mean CE increased with duration of exposure category, implying that those who worked underground longest, were older and experienced higher CE. The strong association between duration of exposure and CE is reflected in table 6.5.1b. As described in chapter 3, cumulative exposures were computed by summing individual exposures which were based on calendar time, estimated Rn daughter levels at that time, the nature of job and the duration of exposure; individual annual job-specific exposure were computed as the multiple of exposure level and duration of exposure. There are few PYRS amongst those experiencing low CE over long durations of exposure, and vice versa.

Crude lung cancer mortality rates showed an increasing overall trend with duration of exposure. This pattern was sustained in all but the highest exposure category. Those exposed to CE of 40 WLM or more were seen to be at twice the crude risk if they acquired their exposure within 24 WM, than over a longer duration.

Table 6.5.1a-c: Cohort Characteristics by Cumulative Exposure and Duration of Exposure Categories

Table 6.5.1a: Mean Age and Cumulative Exposure

Characteristic	Duration of Exposure (WM)			
Mean:	0<3	3-<24	24->	Overall
Duration of Exposure (WM)	1.7	8.5	43.9	10.7
Age at Follow-up (Years)	47	46	48	47
Cumulative Exposure (WLM)	1.7	6.9	27.2	7.7

Table 6.5.1b: Lung Cancer Deaths and PYRS

(Number of PYRS rounded to the nearest integer are shown in brackets)

Exposure	Duration of Exposure (Working Months)			
(WLM)	0<3	3-<24	24->	Overall
0<1	7 (4381)	1 (155)	0 (0)	8 (4536)
1-<10	2 (4464)	8 (9243)	0 (212)	10 (13920)
10-<40	0 (0)	2 (2081)	8 (2461)	10 (4542)
40->	0 (0)	1 (81)	3 (501)	4 (582)
Overall	9 (8846)	12 (11561)	11 (3175)	32 (23583)

Table 6.5.1c: Crude Estimates of the *RR* of Lung Cancer Mortality

(* Reference Category for *RR*)

Exposure	Duration of Exposure (Years)			
(WLM)	0<3	3-<24	24->	Overall
0<1	1*	4.03	-	1*
1-<10	0.28	0.54	-	0.41
10-<40	-	0.60	2.03	1.25
40->	-	7.67	3.74	3.89
Overall	1*	1.02	3.4	

6.5.2 Describing the Effect of Duration of Exposure Using Multiplicative Models

After controlling for confounding effects of attained age and calendar time, the overall risk of lung cancer mortality was found to increase steadily with duration of exposure and stratum specific estimates of risk were found to be significantly heterogeneous; those exposed to durations of over 24 WM had over a three-fold increase in *RR* compared with those in the lowest duration of exposure category (<3 WM). After accounting for the effect of CE however, though the duration-specific estimates of *RR* were found to increase by two-fold, they were no longer heterogeneous enough to be statistically significant.

Patterns of risk among CE categories changed considerably after simultaneously accounting for the main effect of exposure duration. Though CE still had a significant effect on lung cancer mortality after accounting for duration of exposure, those in higher categories of CE were now seen to be at lower risk than those exposed to less than 1 WLM of CE.

Examination of interaction effects in this categorical evaluation based on log-linear models showed no significant effect modification arising from duration of exposure.

Table 6.5.2: Unconfounded Marginal and Interactive Relative Risks by Cumulative Exposure and Duration of Exposure Categories

Exposure	Duration of Exposure (Years)			Overall	
(WLM)	0<3	3-<24	24->	Unadjusted *	Adjusted [a]
0<1	1[d] (Fixed)	4.36 (0.5 - 37.2)	-	1[b] (Fixed)	1[b] (Fixed)
1-<10	0.23 (0.0 - 1.1)	0.53 (0.2 - 1.5)	-	0.37 (0.1 - 1.0)	0.18 (0.0 - 0.7)
10-<40	-	0.47 (0.1 - 2.3)	1.96 (0.7 - 5.5)	1.1 (0.4 - 2.8)	0.26 (0.0 - 1.5)
40->	-	4.73 (0.6 - 40.0)	3.53 (0.9 - 14.2)	3.39 (1.0 - 11.6)	0.65 (0.1 - 5.0)
Overall * Unadjusted	1[c] (Fixed)	1.07 (0.5 - 2.5)	3.59 (1.5 - 8.8)		
Overall Adjusted [a]	1[c] (Fixed)	2.74 (0.7 - 11.0)	6.24 (1.0 - 37.3)		

Note:
[a] Simultaneous adjustment for both main effects WLMG and *WMG*
[b] Baseline category for marginal *RR* estimates among WLMG categories
[c] Baseline category for marginal *RR* estimates among *WMG* categories
[d] Baseline category for obtaining interaction *RR*s
* Significant Heterogeneity ($p = 0.0106$).

6.5.3 Examining Variations in ERR/WLM With Duration of Exposure Using Smooth Parametric Functions

Results of further examination of the role of DE by modelling variations in ERR/WLM are summarized in table 6.5.3 and figure 6.5.3. Assessment based on a linear ERR model with exposure and duration of exposure modelled as continuous variables showed that duration of exposure had a significant modifying effect on the exposure-response relationship (p = 0.0350). The ERR/WLM was seen to increase with duration of exposure. For a fixed level of exposure, each increasing WM added to the ERR by 0.27%. Though the linear additive ERR model provided the 'best fit' of all models fitted, this model indicated a negative ERR/WLM from exposures experienced in less than 18 WM (figure 6.5.3), which implied a possible 'protective effect' in this range.

Evaluation based on a power model showed a reasonably significant modifying effect (p = 0.06) due to duration of exposure. The power parameter estimated (0.946) implied a very nearly linear modifying effect and a reasonably close fit to the linear additive ERR model.

Though exponential models based on continuous variables failed to detect any significant effect modification, when duration of exposure was evaluated as a categorical variable, the model showed significant effect modification (p = 0.0496), with the ERR/WLM showing a ten-fold increase amongst those exposed for 24 WM or more compared to those exposed to between 3-24 WM. Those exposed to less than 3 WM were seen to have a negligible (5.1×10^{-52}) ERR/WLM.

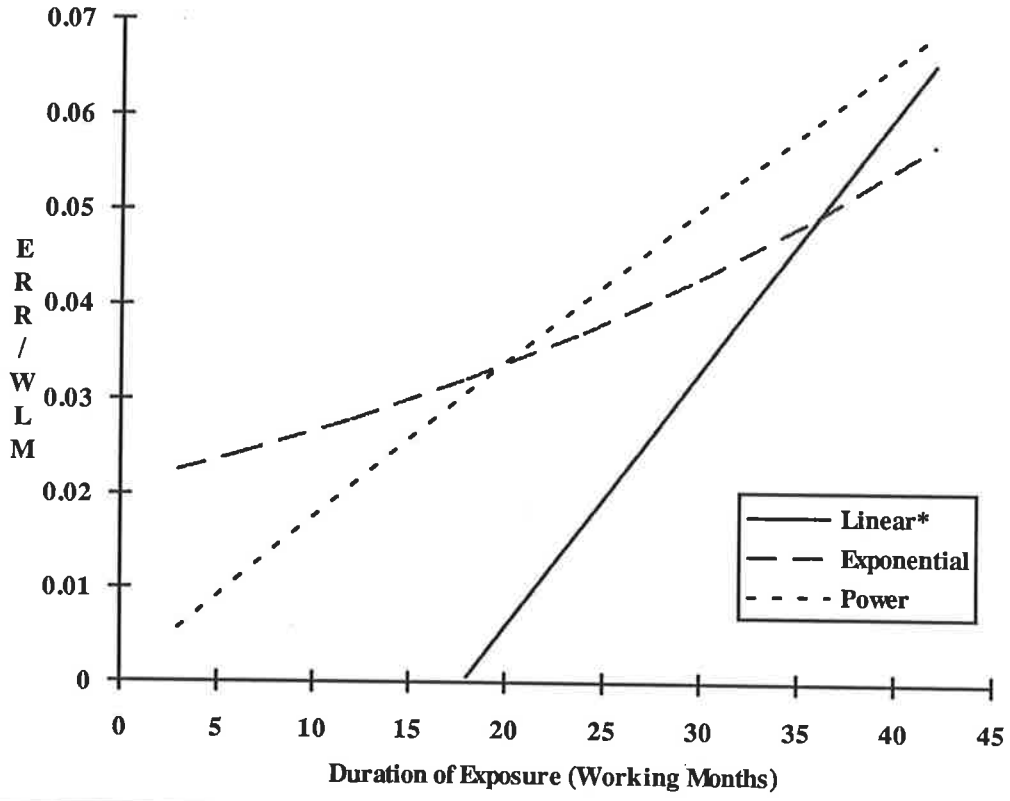
It may be concluded that duration of exposure had a significant linear modifying effect on the exposure-response relationship. Even though duration of exposure is closely associated with CE, the significance of this result cannot be attributed to multicollinearity; because of the complex multi-stage nature of CE computations, there was no evidence to suggest a 1:1 relationship or any specific linear association between individual duration of exposure and CE .

Table 6.5.3: Summary of Results From Fitting Models to Examine the Modifying Effect of Duration of Exposure

ERR	Parameter Estimates			Model Fit		Likelihood Ratio		
Function	Par	ERR	95% CB	Dev	DF	χ^2	DF	p
Linear:								
$\beta_1 w + \beta_2 wz$	β_1	-4.8%		178.1	1908	4.681	1	**0.0305**
	β_2	0.27%						
Exponential:								
βw e $^{\gamma z}$	β	0.23		180.8	1908	2.02	1	0.1553
	e $^{\gamma}$	1.02	(0.98 - 1.07)					
$\beta_1 w$ e $^{\beta 2+\gamma z}$	β_1	1	Fixed	180.8	1908	2.032	1	0.1540
	e $^{\beta 2}$	0.021	(0.00 - 0.47)					
	e $^{\gamma}$	1.024	(0.98 - 1.07)					
$\beta_1 w$ e $^{\gamma_i z c}$	β_1	1	Fixed	179.0	1908	3.855	1	**0.0496**
	e $^{\gamma 1}$	0.000$^+$						
	e $^{\gamma 2}$	0.009	(0.0 - 13.8)					
	e $^{\gamma 3}$	0.090	(0.0 - 0.3)					
Power:								
βw (z) $^{\gamma}$	β	0.002		179.3	1908	3.491	1	**0.0617**
	γ	0.946	(-1.0) - (2.9)					

Note: $^+$ Greater than 0 but very nearly zero.

Figure 6.5.3: Modelling Variations in ERR/WLM With Duration of Exposure



Note: * Best Fitting Model

Risk Functions Fitted: Linear: $ERR/WLM = -0.048 + 0.0027\ z$

Exponential: $ERR/WLM = 0.0 * 0.72^{z}$

Power: $ERR/WLM = 0.002\ z^{0.946}$

6.6 Examining the Effect of Intensity of Exposure

6.6.1 Describing The Crude Effect of Intensity of Exposure

Cohort characteristics summarized in table 6.6.1a-c show that those exposed to the highest intensity of exposure (IE) category - of 1.0 WL or more - were older than others, and experienced a higher level of cumulative exposure (CE). Mean CE increased with IE categories. The distribution of lung cancer deaths (table 6.6.1b) showed that all lung cancer deaths that occurred amongst those exposed to 40 or more WLM of CE were restricted to those in the highest IE category. This category comprised the highest crude mortality rate (table 6.6.1c), which was five-fold greater than the risk in the baseline reference category, who were exposed to the lowest CE and IE. Overall estimates of exposure intensity specific crude risk showed a decline in risk with increasing intensity.

6.6.2 Examining the Effect of Intensity of Exposure Using Multiplicative Models

These overall patterns of declining risk with increasing intensity of exposure were sustained after controlling for confounders and also after simultaneous adjustment for CE. After simultaneous adjustment for both - the main effects of IE and CE - patterns of risk became more pronounced amongst categories of CE and less pronounced with intensity of exposure (table 6.6.2a). Estimates of *RR* showed that those in the highest intensity group faced only half the risk of others. After adjusting for CE the risk amongst those in the highest intensity group dropped a further 10%. These categorical analyses based on multiplicative RR models revealed a significant ($p < 0.05$) modifying effect due to intensity of exposure (the interaction effect was statistically significant after removing main effects).

Table 6.6.1a-c: Cohort Characteristics by Cumulative Exposure and Intensity of Exposure Categories

Table 6.6.1a: Mean Age and Cumulative Exposure

Characteristic	Intensity of Exposure (WL)			
Mean:	**0<0.4**	**0.4-<1**	**1.0->**	**Overall**
Intensity of Exposure (WL)	0.03	0.53	1.72	0.63
Age at Follow-up (Years)	46	45	49	47
Cumulative Exposure (WLM)	2.5	8.7	10.1	7.7

Table 6.6.1b: Lung Cancer Deaths and PYRS

(Number of PYRS rounded to the nearest integer are shown in brackets)

Exposure	Intensity of Exposure (Working Levels)			
(WLM)	**0<0.4**	**0.4-<1.0**	**1.0->**	**Overall**
0<1	5 (2612)	2 (1478)	1 (445)	8 (4536)
1-<10	5 (2823)	2 (5698)	3 (5399)	10 (13920)
10-<40	1 (310)	8 (2480)	1 (1751)	10 (4542)
40->	0 (0)	0 (211)	4 (371)	4 (582)
Overall	11 (5747)	12 (9869)	9 (7967)	32 (23583)

Table 6.6.1c: Crude Estimates of the *RR* of Lung Cancer Mortality

(* Reference Category for *RR*)

Exposure	Intensity of Exposure (WL)			
(WLM)	**0<0.4**	**0.4-<1**	**1.0->**	**Overall**
0<1	1*	0.71	1.17	1*
1-<10	0.93	0.18	0.29	0.41
10-<40	1.68	1.69	0.30	1.25
40->	-	-	5.63	3.89
Overall	1*	0.64	0.59	

Table 6.6.2a: Unconfounded Marginal and Interactive Relative Risks by Cumulative Exposure and Intensity of Exposure Categories

Exposure	Intensity of Exposure (WL)			Overall	
(WLM)	0<0.4	0.4-<1.0	1.0->	Unadjusted	Adjusted [a]
0<1	1[d] (Fixed)	1.28 (0.2 - 6.7)	1.0 (0.1 - 8.6)	1[b] (Fixed)	1[b] (Fixed)
1-<10	0.81 (0.2 - 2.8)	0.35 (0.1 - 1.8)	0.22 (0.1 - 0.9)	0.37 (0.1 - 1.0)	0.51 (0.2 - 1.4)
10-<40	1.80 (0.2 - 15.7)	2.09 (0.7 - 6.6)	0.24 (0.0 - 2.0)	1.1 (0.4 - 2.8)	1.60 (0.5 - 4.8)
40->	- -	- -	6.68 (1.7 - 26.6)	3.39 (1.0 - 11.6)	5.32 (1.3 - 21.5)
Overall Unadjusted	1[c] (Fixed)	1.06 (0.4 - 2.5)	0.50 (0.2 - 1.2)		
Overall Adjusted [a]	1[c] (Fixed)	0.74 (0.3 - 2.0)	0.41 (0.1 - 1.2)		

Note:
[a] Simultaneous adjustment for both main effects WLMG and *WLG*
[b] Baseline category for marginal *RR* estimates among WLMG categories
[c] Baseline category for marginal *RR* estimates among *WLG* categories
[d] Baseline category for obtaining interaction *RR*s

LRT Results: Effect of Interaction After Removing Main Effects:-
LRT Statistic $\chi^2 = 11.59$; DF = 5; $p = \mathbf{0.0409}$.

The role of intensity of exposure was examined further using Log-linear modelling based on continuous exposure variables (table 6.6.2b).

Table 6.6.2b: Summary of Results From Fitting Log-Linear Models to Examine the Effect of Intensity of Exposure

Risk Function	Parameter Estimates			Model Fit		LR Test		
RR Function	Par	*RR*	95% CB	Dev	DF	χ^2	DF	p
$e^{\beta w}$	e^{β}	1.029	(1.01 - 1.05)	181.0	1909			
$e^{\beta w}\ e^{\gamma z}$	e^{β}	1.035	(1.02 - 1.05)	174.1	1908	6.911	1	**0.0086**
	e^{γ}	0.440	(0.23 - 0.85)					
$e^{\beta w}\ e^{\gamma_i z_c}$	e^{β}	1.039	(1.02 - 1.06)	174.7	1907	6.304	2	**0.0428**
	$e^{\gamma 1}$	1	Fixed					
	$e^{\gamma 2}$	0.652	(0.26 - 1.63)					
	$e^{\gamma 3}$	0.308	(0.12 - 0.82)					

Analyses based on both categorical and continuous evaluation identified the intensity of exposure as a significant multiplicative modifier of the exposure-response relationship. In both cases, for a given level of CE, the risk of lung cancer mortality was seen to decline with increasing intensity of exposure. Categorical estimates showed that compared to those in the lowest exposure intensity category, the *RR* dropped to approximately two-thirds in among those exposed to medium levels of exposure intensity (0.4-<1.0 WL), and to a third among those exposed to higher levels (table 6.6.2b). Estimates obtained from continuous evaluation showed that for a fixed level of exposure, the *RR* declined by a factor of 0.44 for each additional WL (table 6.6.2b).

The role of intensity of exposure was further examined through variations in ERR/WLM with IE in the next section.

6.6.3 Examining Variations in ERR/WLM With Intensity of Exposure Using Smooth Parametric Functions

Variations in ERR/WLM with intensity of exposure were further examined using smooth parametric models. Results summarized in table 6.6.3 showed that an ERR model based on a power function provided the 'best fit' in describing variations in ERR/WLM with IE.

When variations in ERR/WLM were examined using a linear model, parameter estimates indicated that the ERR/WLM became negative beyond 2.4WL; thus, implying a protective effect due to exposures at Rn progeny concentrations beyond 2.4WL. In the absence of any demonstrated mechanism for such a protective effect, this finding should be interpreted with caution. It may be a manifestation of other contributing causes such as smoking.

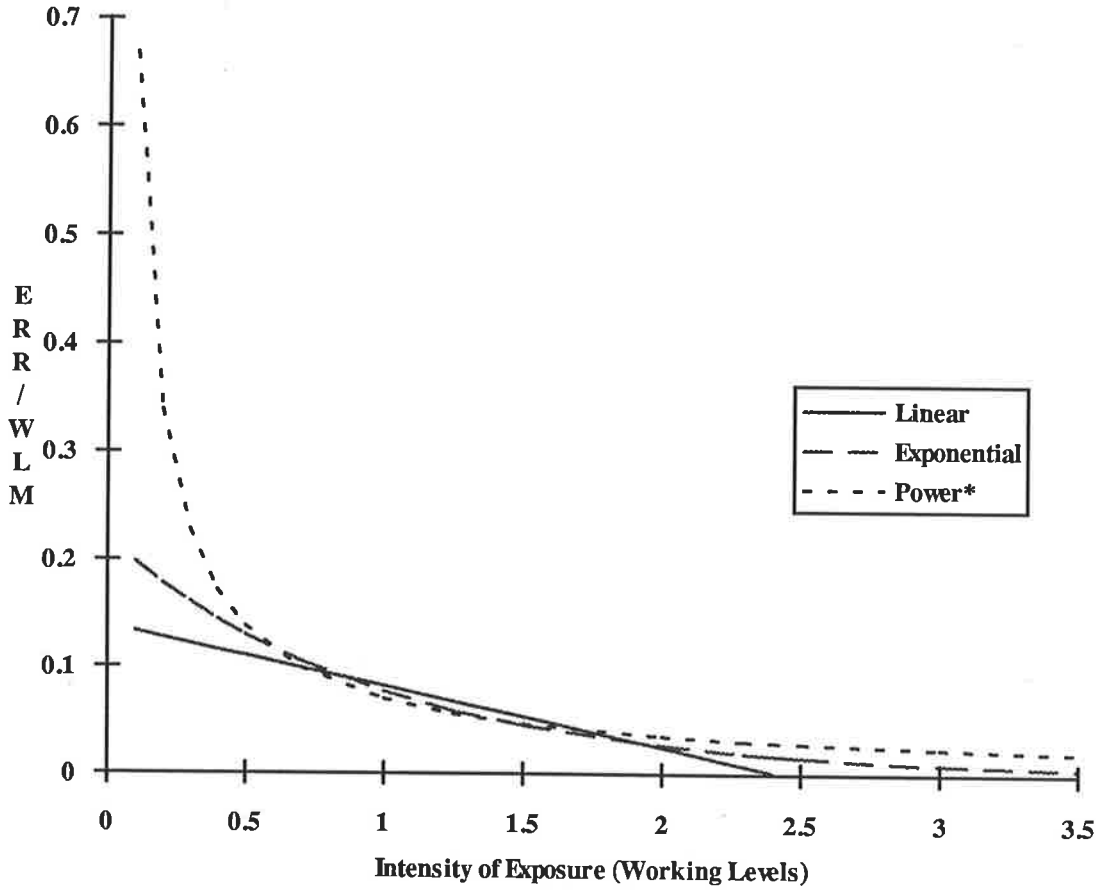
Results summarized in tables 6.6.2b and 6.6.3 show that log-linear multiplicative *RR* models provided the *best fit* in describing the association between intensity of exposure and the exposure-response relationship. Log-linear modelling based on both categorical and continuous variables identified the intensity of exposure as a significant multiplicative modifier of the exposure-response relationship. Though evaluation based on multiplicative RR models showed a highly significant modifying effect due to IE, evaluation based on models of smooth parametric functions of ERR/WLM could not confirm as significant a modifying effect. Despite its lack of statistical significance, results of fitting a power function showed substantial decrease in ERR/WLM with increasing IE (figure 6.6.3) - a pattern that cannot be ignored in view of the findings from categorical examination.

Hence, it is concluded that the relationship between cumulative Rn progeny exposure and lung cancer mortality was significantly modified by intensity of exposure. The role of IE on the exposure-response relationship was best described by categorical evaluation based on a multiplicative RR model and the best description of variations in ERR/WLM with IE was given by a power model.

Table 6.6.3: Summary of Results From Fitting Models to Examine Variations in ERR/WLM With Intensity of Exposure

Risk Function	Parameter Estimates			Model Fit		LR Test		
ERR Function	Par	*RR*	95% CB	Dev	DF	χ^2	DF	p
Linear:								
$\beta_1 w + \beta_2 w*z$	β_1	13.9%		180.9	1908	1.897	1	0.1684
	β_2	-5.7%						
$\beta_i w*z_c$	β_1	29.7%		181.4	1907	1.449	2	0.4846
	β_2	9.7%						
	β_3	5.1%						
Exponential:								
$\beta_1 w\, e^{\beta_2+\gamma z}$	β_1	1		180.7	1908	2.114	1	0.1460
	e^{β_2}	0.22	(0.4 - 1.2)					
	e^{γ}	0.35	(0.1 - 1.9)					
$\beta w\, e^{\gamma z}$	β	0.22		180.7	1908	2.114	1	0.1460
	e^{γ}	0.35	(0.1 - 1.9)					
Power:								
$\beta w\,(z)^{\gamma}$	β	0.07		180.6	1908	2.221	1	**0.1361**
	γ	-0.98	(-2.3) - (0.3)					
$\beta w\,(z+1)^{\gamma}$	β	0.33		180.6	1908	2.207	1	0.1374
	γ	-2.20	(-5.4) - (1.0)					

Figure 6.6.3: Modelling Variations in ERR/WLM With Intensity of Exposure



Note: * Best Fitting Model

Risk Functions Fitted: Linear: $ERR/WLM = 0.139 - 0.057\,z$

Exponential: $ERR/WLM = 0.22 * 0.35^{z}$

Power: $ERR/WLM = 0.07\,z^{-0.98}$

6.7 Modifiers of Effect Under Time Since Exposure Windows

Risk evaluation under time since exposure (TSE) windows showed once again that the risk of Rn progeny related lung cancer mortality declined with increasing TSE and that the major component of risk arose from more recent exposures than more distant exposures (table 6.8.1).

The modifying effects of cumulative duration of exposure (CDE) and intensity of exposure (IE) on TSE were then examined using log ERR models (4); tests of significance were based on LR tests for improvement in model fit. Results of these analyses which are summarized in table 6.7.1 show that neither CDE nor IE had a significant modifying effect on the exposure-response relationship when examined under TSE windows. No distinctive patterns could be observed in the modification parameters. Deviances for both TSE windows effect modification models fitted were very slightly smaller than their non-windows corollaries (sections 6.6.2, 6.6.3); the only major difference in these findings was the disappearance of the significant modifying effect seen in section 6.6.2.

Despite this lack of statistical significance under TSE windows, variations of ERR/WLM should not be ignored and deserve further examination (see chapter 2). Though the current data are too sparse to draw firm conclusions, with further follow-up, distant windows of TSE will have more data and may provide greater power for the examination of effect modifiers under TSE windows.

It is also noted that the choice of three TSE windows may have resulted in sparse distributions of PYRS and events in each window. Evaluation based on two windows of TSE viz., 5-15 years and 15 years or more did make any difference to the overall findings on effect modification. Therefore, the three window classification of TSE which provided greater elaboration of the variations in risk was adopted for the presentation of results in this section.

Table 6.7.1a-b: Modifiers of Effect Under Time Since Exposure Windows

Table 6.7.1a: Examining the Modifying Effect of Duration of Exposure

Cumulative Exposure				Duration of Exposure				Goodness of Fit			
ERR/ WLM *	*RR* Under TSE Windows (years)			ERR/ WM	ERR/WM Under TSE Windows (years)			Dev-iance	Likelihood Ratio		
	5_15	15_25	25_		5_15	15_25	25_		χ^2	DF	p
86%	1	0.06	0.04					179	4.06	2	0.13
23%	1	0.07	0.08	1.02				177	1.29	1	0.27
42%	1	0.08	0.02		1.02	1.01	1.04	177	0.72	2	0.70

Note: * ERR/WLM Coefficient in Reference Category: 5-15 years TSE Window

Table 6.7.1b: Examining the Modifying Effect of Intensity of Exposure

Cumulative Exposure				Intensity of Exposure				Goodness of Fit			
ERR/ WLM *	*RR* Under TSE Windows (years)			ERR/ WL	ERR/WL Under TSE Windows (years)			Dev-iance	LR Test		
	5_15	15_25	25_		5_15	15_25	25_		χ^2	DF	p
86%	1	0.06	0.04					179	4.06	2	0.13
157%	1	0.08	0.08	0.41				178	1.16	1	0.28
174%	1	0.08	0.06		0.42	0.29	0.44	177	0.11	2	0.95

Note: * ERR/WLM Coefficient in Reference Category: 5-15 years TSE Window

6.8 Surrogate Measures of Exposure

The relationship between other indicators of exposure and lung cancer mortality were examined as alternative means of evaluating the risk of Rn related lung cancer mortality at the RH mine. Surrogate measures of exposure considered comprised cumulative duration of exposure (CDE) and intensity of exposure (IE). Some subtly different additional roles of surrogate measures were also examined. The association between IE and lung cancer mortality is examined mainly for the purpose of establishing how good a predictor IE is, compared to CE. Use of CDE as a surrogate measure of exposure, does not require specific estimates of exposure levels obtained from the job exposure matrix (chapter 3), but depended nonetheless on the starting and stopping dates that were recorded. Examination of the surrogate role of CDE therefore enabled not only the evaluation of the predictive role of CDE on lung cancer mortality, but also provided a means of measuring the closeness between the risk estimates provided by CDE and CE, which may be regarded as an indication of the efficacy of measured exposures. The role of surrogate measures was examined using log-linear multiplicative *RR* regression models; their efficacy was judged from surrogate model *goodness-of-fit* characteristics relative to the CE exposure-response model - the baseline *reference model.*

Individual exposure estimates were based on job category, duration worked in that job category and calendar time - an indicator of the geographic level of mining and exposure levels in the mine. Therefore, a more complete examination of surrogate measures of exposure would ideally involve these factors being modelled as multiplicative risk factors in a multiple regression model. A simple application of this approach to longitudinal individual data is presented in chapter 7. Since such analyses are essentially exploratory techniques, description of more detailed methods are delayed till the final chapter of this work.

6.8.1 The Surrogate Role of Cumulative Duration of Exposure

Findings on the surrogate role of CDE are summarized in table 6.8.1. The effect of CDE was first examined by fitting log-linear (multiplicative *RR*) regression models (2) based on categorical variables. Cumulative duration of exposure was found to be a significant explanatory factor of lung cancer mortality, with the *RR* increasing by over three fold amongst those exposed over periods of 24 WM or more. The model deviance with only the main effect of CDE fitted in the model was 180.52. When CE was entered into this model, also as a categorical variable, the model showed a significantly improved fit (p = 0.0311) with the deviance dropping to 171.66. The joint effect of CE increased the *RR* estimates among the categories of CDE, with almost a doubling of the effect seen before. Next, the model was refitted with CE entering the model first; this resulted in a model deviance of 175.99 which was considerably smaller than the deviance when CDE was fitted by itself. It was therefore concluded from this categorical evaluation that even though CDE was seen to be an effective predictor of lung cancer mortality, CE was a slightly better predictor. However, it was noted that the percentage increase in *RR* between the higher two categories of exposure were very nearly the same for CDE and CE.

When this analysis was repeated with both CDE and CE used as continuous variables, CDE alone provided a slightly better fit than CE alone. Neither factor was seen to have any significant multiplicative modifying effect on the other. Furthermore, ERR estimates obtained were also very similar; the ERR/WLM for a given CDE was 1.4%, whilst the ERR/WM for a given level of exposure was 1.7%. It can therefore be concluded that when considered as continuous variables, exponential functions of CDE and CE were found to be equally good predictors of lung cancer mortality. Cumulative duration may therefore be considered as a reasonable surrogate measure of exposure, for the purpose of estimating Rn progeny related risk of lung cancer mortality in the RH cohort.

Table 6.8.1: Examining the Surrogate Role of Duration of Exposure

Risk Function	Parameter Estimates			Model Fit		Likelihood Ratio		
	Par	*RR*	95% CB	Dev	DF	χ^2	DF	p-value
Constant				189.6	1910			
$e^{\gamma_i s_c}$	$e^{\gamma 1}$	1	Fixed	180.5	1908	9.097	2	**0.0106**
	$e^{\gamma 2}$	1.071	(0.45 - 2.55)					
	$e^{\gamma 3}$	3.59	(1.47 - 8.78)					
$e^{\beta_i w_c}$	$e^{\beta 1}$	1	Fixed	176.0	1907	4.536	1	**0.0332**
	$e^{\beta 2}$	0.37	(0.15 - 0.95)					
	$e^{\beta 3}$	1.1	(0.43 - 2.82)					
	$e^{\beta 4}$	3.39	(0.99 - 11.6)					
$e^{\beta_i w_c}$ $e^{\gamma_i s_c}$	$e^{\beta 1}$	1	Fixed	171.7	1905	8.869	3	**0.0311**^{s}
	$e^{\beta 2}$	0.18	(0.04 - 0.74)					
	$e^{\beta 3}$	0.26	(0.04 - 1.51)					
	$e^{\beta 4}$	0.65	(0.08 - 5.0)					
	$e^{\gamma 1}$	1	Fixed					
	$e^{\gamma 2}$	2.74	(0.68 - 11.0)					
	$e^{\gamma 3}$	6.24	(1.05 - 37.3)					
$e^{\gamma z}$	e^{γ}	1.027	(1.01 - 1.04)	180.6	1909	9.069	1	**0.0026**
$e^{\beta w}$	e^{β}	1.029	(1.01 - 1.05)	181.0	1909	8.600	1	**0.0034**
$e^{\beta w}$ $e^{\gamma z}$	e^{β}	1.017	(0.99 - 1.05)	179.8	1908	0.778	1	0.3779^s
	e^{γ}	1.014	(0.98 - 1.05)					

Note: s Surrogate Measure - Cumulative duration of exposure
w Cumulative relevant Rn progeny exposure
s Residual effect of cumulative exposure after accounting for duration of exposure.
Unless otherwise specified, all LR Tests are made against the null model.

Though no further examination of the surrogate role of CDE is undertaken here, it should be noted at this stage that CDE was identified as having a significant modifying effect on CE, as detailed in section 6.6. Therefore, the importance of CE cannot be discounted merely on the grounds of the relative predictive role of CDE.

6.8.2 The Surrogate Role of Intensity of Exposure

The role of intensity of exposure as a surrogate measure for cumulative exposure was examined using similar methods to those described above in the examination of the surrogate role of duration of exposure. Findings from this analysis are summarized in table 6.8.2 below. Categorical analyses showed that intensity of exposure alone was a poor predictor of lung cancer mortality.

When intensity of exposure was considered as a continuous variable, the risk of lung cancer mortality was seen to decline significantly with a halving of *RR* for each increasing WL of exposure. However, there was still a substantial amount of residual variation that was explained significantly by the introduction of CE ($p = 0.0006$). Therefore, it can be concluded that the intensity of exposure alone did not provide an adequate strong predictor of the risk of Rn related lung cancer mortality. However, the importance of intensity of exposure cannot be discounted on these grounds alone, since it is a significant modifier of the exposure response relationship (section 6.6.3) and therefore, an important factor in describing the exposure response relationship.

Table 6.8.2: Examining the Surrogate Role of Intensity of Exposure

Risk Function	Parameter Estimates			Model Fit		Likelihood Ratio		
	Par	*RR*	95% CB	Dev	DF	χ^2	DF	p-value
Constant				189.6	1910			
$e^{\gamma_i s_c}$	$e^{\gamma 1}$	1	Fixed	186.0	1908	3.609	2	0.1645
	$e^{\gamma 2}$	1.063	(0.45 - 2.52)					
	$e^{\gamma 3}$	0.498	(0.21 - 1.21)					
$e^{\beta_i w_c}$	$e^{\beta 1}$	1	Fixed	176.0	1907	13.63	3	**0.0035**
	$e^{\beta 2}$	0.37	(0.15 - 0.95)					
	$e^{\beta 3}$	1.1	(0.43 - 2.82)					
	$e^{\beta 4}$	3.39	(0.99 - 11.6)					
$e^{\beta_i w_c}$ $e^{\gamma_i s_c}$	$e^{\beta 1}$	1	Fixed	172.7	1905	13.31	3	**0.0040**^{*s}
	$e^{\beta 2}$	0.51	(0.19 - 1.39)					
	$e^{\beta 3}$	1.60	(0.53 - 4.83)					
	$e^{\beta 4}$	5.32	(1.32 - 21.5)					
	$e^{\gamma 1}$	1	Fixed					
	$e^{\gamma 2}$	0.74	(0.27 - 2.00)					
	$e^{\gamma 3}$	0.41	(0.15 - 1.15)					
$e^{\gamma z}$	e^{γ}	0.57	(0.32 - 1.03)	185.8	1909	3.782	1	**0.0518**
$e^{\beta w}$	e^{β}	1.029	(1.01 - 1.05)	181.0	1909	8.600	1	**0.0034**
$e^{\beta w}$ $e^{\gamma z}$	e^{β}	1.035	(1.02 - 1.05)	174.1	1908	11.73	1	**0.0006**^{*s}
	e^{γ}	0.440	(0.23 - 0.85)					

Note: s Surrogate Measure - Intensity of exposure

w Cumulative relevant Rn progeny exposure

^{*s} Residual effect of cumulative exposure after accounting for intensity of exposure.

Unless otherwise specified, all LR Tests are made against the null model.

6.8.3 Conclusions on Surrogate Measures of Exposure

Cumulative duration of exposure was found to be a better predictor of lung cancer mortality than cumulative exposure when modelled as continuous variables. However, estimates of ERR obtained per unit of exposure (i.e. per WM or per WLM) were very close. Note that in previous analyses CDE was identified as having a significant (linear additive) modifying effect on CE (section 6.6). Therefore, the importance of CE cannot be undermined by the predictive power of CDE on lung cancer mortality. Implications of this finding in terms of measurement error in Rn exposure levels are discussed in chapter 8.

Intensity of exposure alone could not be identified as a stronger predictor of lung cancer mortality than cumulative exposure. Despite its poor singular predictive power, IE appears to act as a significant modifier (section 6.6.3).

The examination of surrogate measures of exposure based on PRYS data is concluded at this stage. Further examination of surrogate measures of exposure is undertaken in chapter 7 using proportionate hazards modelling under nested case control analyses. Interpretation of the findings on surrogate measures of exposure and the implications of these findings are discussed in the final chapter of this work - chapter 8.

CHAPTER 7

7. Further Issues of Risk Evaluation Using Exposure-Time-Modifier-Response Models: Nested Case-Control Analyses

Page

Tables and Figures

Page

7. Further Issues of Risk Evaluation Using Exposure-Time-Modifier-Response Models: Nested Case-Control Analyses

7.1 Aims

The purpose of the work presented in this chapter is to use the nested case-control (NCC) approach to cohort analyses to examine the following issues:

1. The risk of radon progeny related lung cancer mortality amongst the Radium Hill cohort.

2. The effect of other occupational exposures to radiation and asbestos on radon related lung cancer mortality amongst the Radium Hill cohort.

3. The of smoking on the risk of lung cancer mortality associated with radon progeny exposure at Radium Hill

4. The effect of exposure extrapolation assumptions using sensitivity analyses.

5. The use of surrogate measures of radon exposure in the suited of lung cancer mortality amongst the RH cohort. Surrogate measures examined include duration of employment, duration of exposure and job category.

6. The effect of protracted exposure on the risk of lung cancer mortality related to Rn progeny exposure at RH.

7.2 Methodology

7.2.1 Overview on Nested Case-Control Studies

Poisson regression techniques used so far assumed that the rate of lung cancer mortality was constant throughout the follow-up described by each cell. To justify this assumption it was necessary to partition the PYRS into bands that were sufficiently small, resulting in large data sets which restricted more complex multivariate analyses. These restrictions may be avoided if analyses are based on algebraic expressions for the time effect, and a profile likelihood is calculated for the other parameters in the model. Cox's method for the analysis of follow-up studies provides such a method.

Cox's method is based on a multiplicative model that represents the rate for exposed subjects as a constant multiple of the rate for unexposed subjects across all time bands; rates among unexposed are regarded as *baseline rates* and for most purposes, treated as *nuisance parameters*. The general form of this model introduced by Cox in 1972 (Cox 1972) was named the *proportional hazards model*. With this approach to analysis, time need no longer be divided into intervals and can be analysed as a continuous variable allowing rates to vary continuously over time. This is achieved by a generalization of the profile log-likelihood, referred to as the *partial likelihood* for historical reasons when applied to Cox's approach.

The adaptation of Cox's method to deal with confounders based on further stratification of confounding factors closely resembles a matched case-control analysis in which controls are matched to cases with respect to the confounder. The small number of subjects in each *risk set* - matched case-control set - resulting from such stratification reduces the computational intensity of the analysis; thus stratification results in simplifying the analysis rather than complicating it. The extent of stratification is limited by the fact that over stratification can leave few or no controls for each case. Risk sets with cases and no controls do not contribute to the analysis and are therefore lost events. Hence, compromise has to be made.

Study methods based on risk sets extracted from the cohort for all cases arising during follow-up are known as *nested case-control* (NCC) designs. The computational burden of nested case-control analyses is further reduced by sampling the risk set. In sampling the risk set, the risk set is replaced by a set containing the case and a random sample of all the remaining subjects in the original risk set. Analytical techniques are the same as conditional logistic regression analysis of individually matched case-control studies. Provided the sampling of different risk sets are independent and have a sufficient number of controls, NCC analyses yield similar estimates of model parameters and precision as analyses based on a complete enumeration of controls for each case. Independence of risk sets necessitate sampling with replacement for controls i.e., selection based on random samples of controls drawn for each case from all eligible non-cases at the time. Therefore, all cases have to be available for selection as controls until they become a case. In the analyses of temporal variables, the NCC approach avoids computational complexities of time-dependent covariate allocation. The NCC approach is also most useful in detailed analysis of extensive personal record data such as employment histories where it enables considerable economies. The precision of nested case-control analyses versus analyses based on complete controls is given by:

$$\sqrt{\frac{c}{c+1}}$$

where, c = the number of controls per case.

Cox's method can be extended to deal with temporal variables. Temporal variables can be accommodated either by using alternative time scales reflecting elapsed time from different origins or by addressing changing exposures with time.

This methodological overview was written after ideas were crystallized with the aid of works by Breslow, Day, Langholz and others (Breslow and Day 1980; Breslow *et. al.* 1983), and a manuscript on Cohort Analyses from David Clayton (Clayton and Hills Draft 1992) of the MRC Unit of Biostatistics, Cambridge.

7.2.2 Construction and Characteristics of Risk sets

Nested case-control analyses of the Radium Hill cohort was set up by matching cases with controls on age and year of birth. Lung cancer mortality being the outcome of interest, for convenience of reference with respect to NCC analyses, each lung cancer death amongst the RH cohort will be referred to as a *case*. For each case, controls were selected from amongst survivors by matching on age at the time of case occurrence - *reference age* - and year of birth. Exact matching on the integer component of age and year of birth resulted in few controls for those born prior to 1900; hence, broader bases of matching - within 5 years - were used in these instances.

Risk sets were constructed using two SAS macros specifically written for this purpose. The first program identified all *potential controls* - all possible controls - for each case, which formed the *complete* risk sets. The second program then selected a random sample of controls from each complete risk set by incorporating SAS uniform random number generation commands into a random sample selection algorithm. Being written as a macro, this program allowed for the number of random controls to be selected to vary by introducing the number of random controls desired as a macro call variable. In risk sets where the number of potential controls fell short of the number of random controls desired, the program allowed all potential controls to be selected.

The programs for constructing risk sets only required the identifiers, the matching variables and outcome of follow-up variables as input variables. The output from these programs comprised additional variables which included identifiers of risk sets and case or control status. Therefore, once the risk sets were constructed, the body of data needed for analyses was extracted from the main RH dataset using SAS extraction and merging commands.

7.2.3 Risk Evaluation Models Used in this Chapter

Variations in ERR/WLM - given the presence of some relevant exposure - were further examined in this chapter, using more complex models than those fitted previously, viz., *linear-exponential "cell-killing" model* (Thomas *et al.* 1985) (model 1) , the *non-linear model* (model 2), and a more general model incorporating exponential and non-linear components (model 3). Exposure-response models were also fitted under time-since-exposure windows using the model defined in chapter 5 (model 4) and a further extension incorporating the effect of attained age (model 5) which is generally referred to as the *BEIR IV TSE model* (BEIR IV, 1987) and described in chapter 2. These models are defined as follows:

Linear-Exponential "Cell Killing" Model: $RR = [1 + \beta w]\, e^{\lambda w}$ (1)

Non-Linear Model: $RR = [1 + \beta w^{\nu}]$ (2)

General Model: $RR = [1 + \beta w^{\nu}]\, e^{\lambda w}$ (3)

Time-Since-Exposure Model: $RR = [1 + \beta_1 w_{5\text{-}15} + \beta_2 w_{15\text{-}}]$ (4)

BEIR IV Time-Since-Exposure Model: $RR = [1 + \gamma(a)\beta_1 (w_{5\text{-}15} + \beta_2 w_{15\text{-}})]$ (5)

where,
- β - measures the ERR/WLM
- λ - the exponential modification (*cell-killing*) parameter
- ν - the *non-linear* parameter
- a - age at risk,
- $\gamma(a)$ - a function of age at risk categorized as: $a < 55$ years, $55 \leq a < 65$ years, and $a \geq 65$ years
- $w_{5\text{-}15}$ exposures accumulated 5-15 years previously
- $w_{15\text{-}}$ exposures accumulated prior to the last 15 years

7.3 Risk Evaluation Using Nested Case-Control Analyses

7.3.1 Estimating ERR/WLM Using Smooth Parametric Functions

The risk of lung cancer mortality associated with Rn progeny exposure at RH was further examined using smooth parametric functions defined by models 1, 2 and 3. These models were fitted in a nested case-control analysis utilising all matched controls within each risk-set. Since the purpose of this analysis was to examine patterns of variation in the risk of lung cancer mortality associated with Rn progeny exposure among those exposed to Rn progeny, only those who were exposed to Rn progeny at RH were included in this analysis. This analysis was therefore based on the 32 risk sets identified by the workers who died of lung cancer and were exposed to Rn progeny at RH.

Results of fitting the linear exponential model, the non-linear model and a general model incorporating both these components (models 1, 2 and 3) are summarized in table 7.3.1a. Based on the results of a likelihood ratio test for the effect of including the exponential modification component into the general model (model 3) after fitting all other effects in the model, a significant exponential modifying effect was identified in the association between Rn progeny exposure and lung cancer mortality amongst the RH cohort. The non-linear component when similarly examined showed a marginal effect. Overall, the best fitting model was provided by the following functional form:

$$\mathbf{RR} = (1 + 1.201\, w^{-0.88})\, 1.027^{w}$$

Risk estimates evaluated by this risk function were computed for various levels of Rn progeny exposure (table 7.3.1c) and graphically represented in figures 7.3.1a and 7.3.1b.

Table 7.3.1a: Results of Modelling Variations in Excess Relative Risk Using Smooth Parametric Functions

Risk Function	Parameter Estimates			Model Fit		LR Test		
	Par	Est	95% CB	Dev	DF	χ^2	DF	p
NULL:				181.45	32			
Linear Model:								
RR = $(1+\beta w)$	β	0.029		177.93	31	3.52^n	1	**0.06**
Exponential Model:								
RR = $e^{\beta w}$	e^{β}	1.018	1.00-1.04	177.61	31	3.84^n	1	**0.05**
Linear-Exponential Model:								
RR = $(1+\beta w)$	β	0.029		177.93	31	3.52^n	1	**0.06**
RR = $(1+\beta w)e^{\lambda w}$	e^{λ}	1.029	0.99-1.07	177.55	30	0.38^l	1	0.54
	β	-0.007						
Power Model:								
RR = $(1+\beta w^{\nu})$	ν	1.653		177.12	30	0.81^l	1	0.37
	β	0.003						
General Model:								
RR = $(1+\beta w^{\nu})e^{\lambda w}$	ν	-0.88	-2.93-1.12	174.66	29	6.79^n	3	**0.08**
	β	1.201	0.02-46.6			3.39^e	1	**0.07**
	e^{λ}	1.027	1.00-1.06			2.46^p	1	0.17

Note: w - Rn Progeny Exposure; λ - exponential modifying parameter; ν - non-linear parameter; β - ERR/WLM.

^{n} LRT: Improvement on Null Model.

^{e} LRT: Improvement on Linear-Exponential Model.

^{p} LRT Improvement on Power Model. ^{l} LRT Improvement on Linear Model.

Due to the difficulty of estimating confidence bounds for the linear parameter β in these models, these models were re-fit using a log excess relative risk term for which confidence bounds could be obtained more easily; these results are summarized in table 7.3.1b. Even though the risk function modelled included a constant of 1, the parameter β was initially allowed to be freely estimated. It was only after β was identified as positive, that it was re-fitted as an exponential term - thereby, implicitly constrained to being positive - in the analysis presented in table 7.3.1b where this alternative was used to obtain confidence bounds for β. Therefore, if a protective effect from Rn progeny exposure could have existed at any level of Rn progeny exposure studied, this model would have been capable of identifying it.

The risk function estimated by the general model showed a turning point at 6.3 WLM. Though it would have been more appropriate to plot the ERR rather than RR in figures 7.3.1a and 7.3.1b, RR were used to enable easier interpretation.

It is concluded from these findings that the risk of lung cancer mortality associated with Rn progeny exposure at RH increased with increasing exposure beyond 6.3 WLM, which was slightly lower than the mean exposure experienced at RH by underground workers. This turning point could have been estimated more accurately by differentiating the risk function with respect to *w* (exposure), but the level of accuracy gained from graphic estimation and tabulations appeared adequate for the purpose of this work. An approximate four-fold risk was associated with 50 WLM of cumulative exposure. The doubling dose estimated from this model incorporating non-linear and exponential components was approximately 25 WLM.

Examination of exposure-response relationships *per se,* is concluded at this stage; hereafter, analyses will focus mainly on exposure-time-modifier-response relationships.

Table 7.3.1b: Results of Modelling Variations in Excess Relative Risk Using Smooth Parametric Functions (Log-Excess Relative Risk Term)

Risk Function	Parameter Estimates			Model Fit		LR Test		
	Par	Est	95% CB	Dev	DF	χ^2	DF	p
NULL:				181.45	32			
Linear-Exponential Model:								
RR = $(1+\beta w)$	β	0.029		177.93	31	3.52[n]	1	**0.06**
RR = $(1+\beta w)e^{\lambda w}$	e^{λ}	1.029	0.99-1.07	177.55	30	0.38[l]	1	0.54
	β	-0.007						
Power Model:								
RR = $(1+w^{\nu}e^{\beta})$	ν	1.646	-6.21-3.54	177.12	30	0.81[l]		0.37
	e^{β}	0.003	0.00-95.80					
General Model:								
RR = $(1+w^{\nu}e^{\beta})e^{\lambda w}$	ν	-0.864		174.68	29	6.79[n]		0.08
	e^{β}	0.965	0.02-54.35			2.20[e]		0.14
	e^{λ}	1.027	1.00-1.06			2.46[p]		0.12

Note: w - Rn Progeny Exposure; λ - exponential modifying parameter; ν - non-linear parameter; β - ERR/WLM.

[n] LRT: Improvement on Null Model.

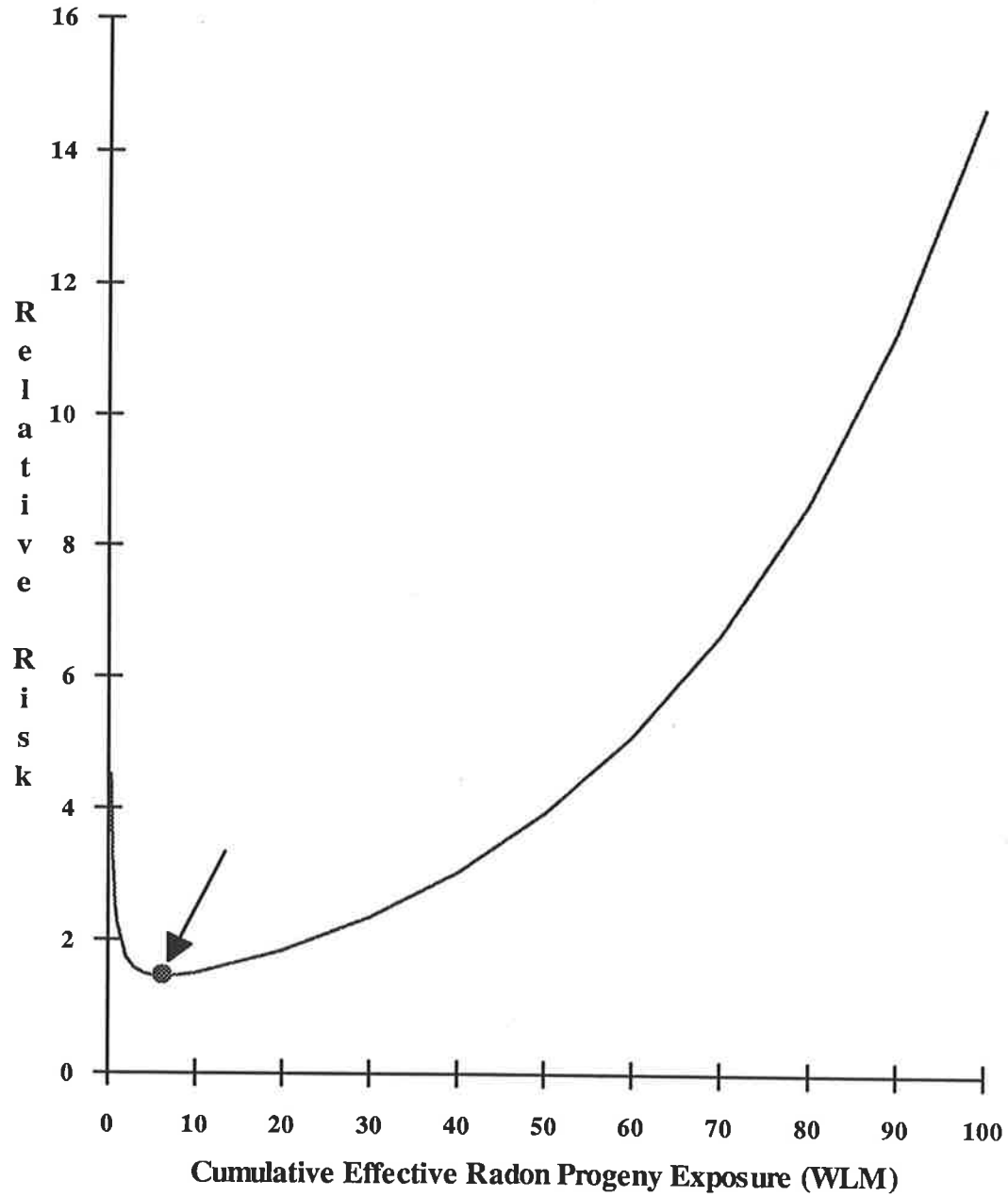
[e] LRT: Improvement on Linear-Exponential Model.

[p] LRT Improvement on Power Model.

Table 7.3.1c: Relative Risk of Lung Cancer Mortality Associated With Cumulative Relevant Radon Progeny Exposure at Radium Hill Estimated From Best Fitting Model : RR = $(1 + 1.201\ w^{-0.88})\ 1.027^{w}$ (Incorporating Non-Linear and Exponential Components)

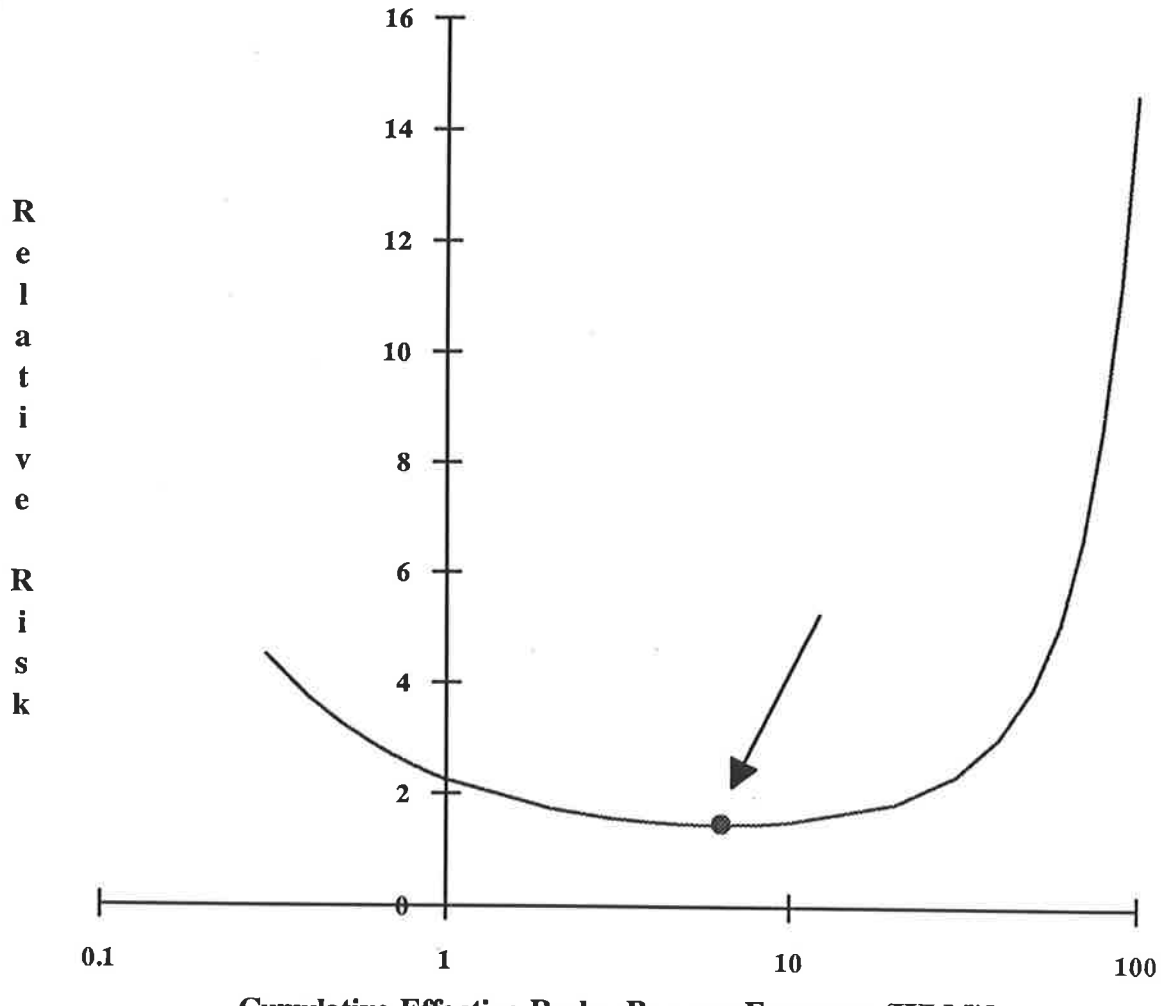
Cumulative Relevant Radon Progeny Exposure (WLM)	Relative Risk of Lung Cancer Mortality (RR)	
0.2	5.982148	
0.3	4.500613	
0.4	3.729396	
0.5	3.253341	
0.6	2.929101	
0.7	2.693598	
0.8	2.514612	
0.9	2.373923	
1.0	2.260427	
2.0	1.743029	
3.0	1.577959	
4.0	1.506921	
5.0	1.475380	
6.0	1.464538	
6.1	1.464228	
6.2	1.464035	
6.3	**1.463953**	• **Turning Point**
6.4	1.463978	
6.5	1.464106	
6.6	1.464331	
6.7	1.464651	
7.0	1.466143	
7.7	**1.472338**	• **Mean Exposure**
8.0	1.475997	
9.0	1.491738	
10.0	1.511938	
20.0	1.850332	
30.0	2.357793	
40.0	3.038494	
50.0	3.934515	
60.0	5.107494	
70.0	6.639929	
80.0	8.640295	
90.0	11.25052	
100.0	14.65599	

Figure 7.3.1a: Relative Risk of Lung Cancer Mortality With Cumulative Relevant Radon Progeny Exposure Amongst The Radium Hill Cohort (Exposed Workers Only)



$$RR = (1 + 1.201\, w^{-0.88})\, 1.027^{w}$$

Figure 7.3.1b: Relative Risk of Lung Cancer Mortality With Cumulative Relevant Radon Progeny Exposure Amongst The Radium Hill Cohort X-Axis Plotted on Logarithmic Scale (Exposed Workers Only)



$$RR = (1 + 1.201\, w^{-0.88})\, 1.027^{w}$$

7.3.2 Risk Evaluation Under Time-Since-Exposure Windows

Cox's proportionate hazards models with exposures modelled as linear excess relative risk functions were used to evaluate the risk of Rn progeny related lung cancer mortality in these NCC analyses. Several representations of exposure in TSE windows were examined across categories of age at death from lung cancer classified using the BEIR IV (BEIR IV 1988) categorization for reference age. The distribution of cases and controls across various exposure windows and reference age categories are summarized in table 7.3.2a.

Table 7.3.2a: Distribution of Cases and Controls Across Exposure Windows and Reference Age Categories

Exposure Window	Cases Reference Age				Controls Reference Age			
	<55	55-<65	65->	All	<55	55-<65	65->	All
5_ years:								
0 WLM	6	7	9	22	94	132	101	327
>0 WLM	8	14	10	32	116	182	101	399
Total	14	21	19	54	210	314	202	726
5_15 Years								
0 WLM	8	21	19	48	179	311	201	691
>0 WLM	6	0	0	6	31	3	1	35
Total	14	21	19	54	210	314	202	726
15_ Years								
0 WLM	10	7	9	26	119	134	101	354
>0 WLM	4	14	10	28	91	180	101	372
Total	14	21	19	54	210	314	202	726
5_20 Years								
0 WLM	7	20	17	44	137	295	192	624
>0 WLM	7	1	2	10	73	19	10	102
Total	14	21	19	10	210	314	202	726
20_ Years								
0 WLM	13	7	10	30	151	147	110	408
>0 WLM	1	14	9	24	59	167	92	318
Total	14	21	19	54	210	314	202	726

Analyses in this section commenced by fitting cumulative relevant exposure (hereafter referred to as cumulative exposure or CE) based on a five year lag, denoted by w5_, as defined in model Ia. The modifying effect of reference age was then examined by modelling reference age as an exponential modification term (model Ib) using notation defined previously in chapter 6.

Cumulative Relevant Exposure: $ERR = w5\_\, e^{\beta}$ (Ia)

Modifying Effect of Reference Age: $ERR = w5\_\, e^{\beta}\, e^{\gamma_1 z_c}$ (Ib)

Results of fitting these models are summarized in table 7.3.2. A statistically significant ERR/WLM of 2.7% was estimated in the simplest model with a 95% CI of (0.01 - 0.12) for this estimate. A score test for linear trend yielded a highly significant linear trend (score statistic = 5.68; p = 0.02) for ERR/WLM. Results of likelihood ratio tests show that this effect of Rn exposure on the risk of lung cancer mortality was not significantly modified by age at death from lung cancer when reference age was fitted as a dichotomous or trichotomous variable. However, the risk of lung cancer mortality was seen to increase with increasing categories of reference age; a three-fold increase in risk was observed in the reference age category of 65 years or more. The score test for log-linear trend in reference age was not significant at the 5% level (score statistic = 0.01; p = 0.93).

Table 7.3.2b: Results from Fitting Linear Excess RR Models for Cumulative Relevant Exposure with Reference Age Effect Modification

ERR		Parameter Estimates			Model Fit		Likelihood Ratio Test		
Ref.	Variable	Par	Est.	95% CB	Dev	FP	χ^2	DF	p
	NULL				283.98	0			
Ia	w5_		1	Fixed	279.99	1	9.99	1	0.0458
	Constant	e^{β}	0.027	(0.01 - 0.12)					
Ib	w5_		1	Fixed	279.83	2	0.16	1	0.6910
	Constant	e^{β}	0.032	(0.01 - 0.16)					
	<55 years	e^{γ_1}	0.487	(0.68 - 34.8)					
Ib	w5_		1	Fixed	279.40	2	0.59	1	0.4413
	Constant	e^{β}	0.017	(0.01 - 0.19)					
	65_ years	e^{γ_2}	3.058	(0.12 - 73.0)					
Ib	w5_		1	Fixed	279.40	3	0.60	2	0.7427
	Constant	e^{β}	0.017	(0.00 - 0.39)					
	<55 years	e^{γ_1}	0.901	(0.01 - 137)					
	55-65 years	e^{γ_0}	1	Fixed					
	65_ years	e^{γ_2}	2.925	(0.07 - 119)					

Note: Ref. - Reference to models given in section 7.3 and defined as:-

Cumulative Relevant Exposure (w5_): $ERR = w5\_\, e^{\beta}$ (Ia)

Modifying Effect of Reference Age (z): $ERR = w5\_\, e^{\beta}\, e^{\gamma_1 z_c}$ (Ib)

Separate models were then fitted for cumulative exposure measured in two TSE windows, using the BEIR IV (NAP 1988) exposure windows, 5-15 years and 15 years or more (model IIa). Once again risk estimates were also allowed to vary with age at death from lung cancer in examining the modifying effect of reference age (model IIb). Results of fitting these models based on TSE windows showed slight improvement in fit compared to models based on cumulative relevant exposures (table 7.3.2c).

Exposure in TSE Windows:	$ERR = (w5\_15 + \beta_1 w15\_)\, e^{\beta_0}$	(IIa)
Modifying Effect of Reference Age:	$ERR = (w5\_15 + \beta_1 w15\_)\, e^{\beta_0}\, e^{\gamma_i z_c}$	(IIb)
Continuous Effect of Reference Age:	$ERR = (w5\_15 + \beta_1 w15\_)\, e^{\beta_0}\, e^{\gamma z}$	(IIc)

Where, e^{β_0} - ERR/WLM parameter

β_1 - RR parameter

e^{γ} - *effect modification* parameter

The risk of Rn progeny related lung cancer mortality dropped in weightage with increasing TSE windows. The effect of exposures accumulated up to 15 years prior to the reference age was considerably lower than that sustained 5 - 15 years before the reference age; the RR of lung cancer mortality dropped to 0.055 for exposures sustained 15 years or more prior. A score test for trend in exposures sustained between 5 and 15 years prior to the reference age confirmed a highly significant linear trend (score statistic = 7.2998, p = 0.0069). After accounting for the effect of exposures in the 5_15 year window, a linear trend, though only of borderline statistical significance (score statistic = 3.3415, p = 0.0676), was also noticed in exposures sustained in the previous window of time.

Table 7.3.2c: Results from Fitting Linear Excess RR Models Using the BEIR IV Classification of TSE Windows and Reference Age

ERR		Parameter Estimates			Model Fit		Likelihood Ratio Test		
Ref	Variable	Par	Est.	95% CB	Dev	FP	χ^2	DF	p
	NULL				283.98	0			
IIa	w5-15		1	Fixed	277.29	2	6.69	2	0.0352
	w15_	β_1	0.055						
	Constant	e^{β_0}	0.334	(0.03 - 2.93)					
IIb	w5-15		1	Fixed	277.27	3	0.02	1	0.8848
	w15_	β_1	0.053						
	Constant	e^{β_0}	0.392	(0.01 - 19.8)					
	Age < 55	e^{γ_1}	0.796	(0.02 - 38.6)					
IIb	w5-15		1	Fixed	277.17	3	0.13	1	0.7237
	w15_	β_1	0.058						
	Constant	e^{β_0}	0.289	(0.03 - 2.85)					
	Age 65->	e^{γ_2}	1.884	(0.07 - 50.6)					
IIb	w5-15		1	Fixed	277.16	4	0.13	2	0.9371
	w15_	β_1	0.060						
	Constant	e^{β_0}	0.260	(0.00 - 42.3)					
	Age < 55	e^{γ_1}	1.132	(0.01 - 155)					
	55 - 65	e^{γ_0}	1	Fixed					
	Age 65->	e^{γ_2}	2.014	(0.03 - 129)					
IIc	w5-15		1	Fixed	273.96	3	3.33	1	0.0680
	w15_	β_1	-0.028	(-1.11 - 0.05)					
	Constant	e^{β_0}	$2x10^5$	$(0.00 - 2x10^{14})$					
	Ref. Age	e^{γ}	0.767	(0.5 - 1.16)					

Note: Ref. - Reference to models given in section 7.3 and defined as:-

Exposure (w) in TSE Windows: $ERR = (w5\_15 + \beta_1 w15\_)\, e^{\beta_0}$ (IIa)

Modifying Effect of Reference Age (z): $ERR = (w5\_15 + \beta_1 w15\_)\, e^{\beta_0}\, e^{\gamma_i z_c}$ (IIb)

Continuous Effect of Reference Age (z): $ERR = (w5\_15 + \beta_1 w15\_)\, e^{\beta_0}\, e^{\gamma z}$ (IIc)

The risk of lung cancer mortality varied with reference age but, reference age was not a statistically significant modifier of effect when modelled categorically; the risk of Rn progeny related lung cancer mortality was lowest in the reference age category of 55-65 years with a two-fold increase amongst those above 65 years. However, when reference age was modelled as a continuous variable it showed a borderline significant modifying effect (LRT $\chi^2 = 3.331$, $p = 0.0680$); exposures accumulated up to 15 years prior to the reference age had a smaller effect than exposures sustained thereafter; this effect though negative, was not statistically significant. No significant trend was seen in reference age (score statistic = 0.39229, $p = 0.5311$).

Further categorizations of TSE windows were also explored - three windows of 5-15, 15-25, and 25 years or more - but, the data were found to be too sparse to fit any worthwhile models. To circumvent problems arising from the sparseness of the BEIR IV TSE windows classification, another categorization of TSE windows - exposure windows spanning 5-20 years and 20 years or more - was also applied. Though the distribution of cases and controls across these categories were less sparse, the models fitted did not improve the goodness of fit (table 7.3.4). These models were therefore dismissed in favour of the more parsimonious models fitted under the BEIR IV categorization.

Table 7.3.2d: Results from Fitting Linear Excess RR Models Using and Alternative Classification of TSE Windows

ERR		Parameter Estimates			Model Fit		LR Test		
Ref.	Variable	Par	Est.	95% CB	Dev	FP	χ^2	DF	p
	NULL				283.98	0			
IIa	w5-20		1	Fixed	279.93	2	4.05	2	0.1319
	w20_	β_1	0.682						
	Constant	$e^{\beta 0}$	0.037	(0.00 - 0.58)					
IIb	w5-20		1	Fixed	279.84	3	0.09	1	0.7616
	w20_	β_1	0.933						
	Constant	$e^{\beta 0}$	0.034	(0.00 - 2.15)					
	Age < 55	$e^{\gamma 1}$	0.572	(0.01 - 38.9)					
IIb	w5-20		1	Fixed	279.07	3	0.86	1	0.3546
	w20_	β_1	20.50						
	Constant	$e^{\beta 0}$	0.000	(0.00 - 0.00)					
	Age 65->	$e^{\gamma 2}$	8.416	(0.01 - 6150)					
IIb	w5-20		1	Fixed	279.49	4	0.44	2	0.8013
	w20_	β_1	0.633						
	Constant	$e^{\beta 0}$	0.025	(0.00 - 2.78)					
	Age < 55	$e^{\gamma 1}$	1.027	(0.01 - 149)					
	Age 55-<65	$e^{\gamma 0}$	1	Fixed					
	Age 65->	$e^{\gamma 2}$	2.960	(0.06 - 135)					

Note: Ref. - Reference to models given in section 7.3 and defined as:-

Exposure (w) in TSE Windows: $ERR = (w5\_20 + \beta_1 w20\_)\, e^{\beta 0}$ (IIa)

Modifying Effect of Reference Age (z): $ERR = (w5\_20 + \beta_1 w20\_)\, e^{\beta 0}\, e^{\gamma i z_c}$ (IIb)

Continuous Effect of Reference Age (z): $ERR = (w5\_20 + \beta_1 w20\_)\, e^{\beta 0}\, e^{\gamma z}$ (IIc)

7.4 Sensitivity Analyses of Exposure Extrapolation Assumptions

Estimates of Rn progeny levels in the uranium mine at Radium Hill from 1948-1952 were obtained by direct extrapolation from the 1953 estimates, which were based on Rn gas measurements (chapter 3). These extrapolations were made on the assumption that work conditions were the same in both periods. Further examination of reports on mining activities at Radium Hill and anecdotal evidence obtained from members of the study cohort revealed however, that those employed at the mine from 1948 to 1952 were involved in securing old mine shafts that remained from previous mining attempts. The nature of their work may have exposed them to much higher levels of exposure than those estimated in the new mine in 1953. This section explores the use of a simple method of sensitivity analysis to examine this particular aspect of measurement error.

Sensitivity analyses were performed by weighting exposure measurements from 1948-1952 by various multipliers and gauging the effect of these weights from the goodness-of-fit of linear ERR models - similar to those used in the previous sections - fitted thereafter. Sensitivity analyses examined the effects of multiplying factors of 2, 3 and 4. It thereby reduced to a *'what if?'* analysis, examining the questions *'what if exposures experienced from 1948-1952 had been twice, thrice or four times those estimated in 1953?'*.

Finally, one further exposure extrapolation assumption was examined. This comprised exploring the assumption of any exposure extrapolation at all attributed to the periods 1948-1952 and for 1962. The reason for this exploratory analysis was that at the time of the initial construction of the JEM, no allowance was made for exposures experienced when the RH mine was not producing ore since no Rn gas measurements were made before 1953 or after 1961; ore production commenced in 1953 and concluded in 1961. The effect of this assumption was explored by using a multiplier of zero for exposures acquired prior to 1953 and after 1961, in a similar sensitivity analysis.

Before embarking on sensitivity analyses, study data were examined with a view to understanding how exposures experienced prior to 1953 and during 1962 contributed to the individual cumulative exposures.

The nested case-control study dataset comprised a total of 54 cases and 726 controls of whom 22 (41%) of the cases and 327 (45%) of the controls were not exposed to any relevant Rn exposure at RH. Of those effectively exposed at RH 5 cases (16%) and 69 controls (17%) experienced exposures prior to 1953; amongst them, only 1 case and 21 controls accumulated their total relevant Rn exposure before 1953. Exposures experienced prior to 1953 by nested case-control study participants are summarized in table 7.4.1.

Table 7.4a: Summary of Exposure Patterns Prior to 1953

Exposure	Calendar Period					
	1948	1949	1950	1951	1952	1948-52
Number (%):						
Cases	1 (1.9%)	1 (1.9%)	1 (1.9%)	3 (5.6%)	5 (9.3%)	5 (9.3%)
Controls	5 (0.7%)	2 (0.3%)	4 (0.6%)	14 (1.9%)	66 (9.1%)	69 (9.5%)
Total Exposure (WLM):						
Cases	13.68	13.64	15.76	32.63	71.37	147.07
Controls	38.22	27.28	35.19	180.01	505.82	786.53
Mean Exposure (WLM):						
Cases	13.68	13.64	15.76	10.88	14.27	29.4
Controls	7.64	13.64	3.94	0.78	7.66	11.40

In each of the calendar years prior to 1953, a higher proportion of cases appeared to be exposed than controls. Prior to 1953, cases experienced a substantially higher exposure than controls - the mean exposure for cases was over 2.5 times higher than the mean exposure for cases. Results of sensitivity analyses (table 7.4.2) showed that for exposures experienced prior to 1953, exposure multiplying factors 2, 3 and 4 made no improvement to the model fit; deviances for these models were higher than that of the reference model (multiplier of 1).

Table 7.4b. Summary of Results from Sensitivity Analyses

MODEL PARAMETERS			EXPOSURE MULTIPLYING FACTORS			
			1*	2	3	4
Cumulative Exposure:	w5_	$e^{\beta 0}$	0.02	0.01	0.07	0.05
Deviance			279.397	280.530	280.310	280.571
Cumulative Exposure:	w5_	$e^{\beta 0}$	0.02	0.01	0.07	0.05
Reference Age (years):	< 55	$e^{\gamma 1}$	0.90	0.89	0.69	0.25
	55-<65	$e^{\gamma 0}$	1	1	1	1
	65->	$e^{\gamma 2}$	2.92	3.45	4.01	4.52
Deviance			279.40	280.53	280.31	280.57
Exposure in Windows:	w5_15		1	1	1	1
	w15_	β_1	0.06	0.04	0.04	0.03
Deviance			277.16	277.15	277.22	277.28
ERR Multiplier		$e^{\beta 0}$	0.26	0.21	0.17	0.14
Exposure in Windows:	w5_15		1	1	1	1
	w15_	β_1	0.06	0.04	0.04	0.03
Reference Age (years):	< 55	$e^{\gamma 1}$	1.14	1.54	2.03	2.62
	55-<65	$e^{\gamma 0}$	1	1	1	1
	65->	$e^{\gamma 2}$	2.02	2.37	2.66	2.91
Deviance			277.16	277.15	277.22	277.28

Note: * Reference Model: Direct Extrapolation (Exposure Multiplier = 1).

Cumulative Relevant Exposure (w5_): $ERR = w5\_\ e^{\beta}$ (Ia)

Modifying Effect of Reference Age (z): $ERR = w5\_\ e^{\beta}\ e^{\gamma_t z c}$ (Ib)

Exposure in TSE Windows: $ERR = (w5\_15 + \beta_1 w15\_)\ e^{\beta 0}$ (IIa)

Modifying Effect of Reference Age: $ERR = (w5\_15 + \beta_1 w15\_)\ e^{\beta 0}\ e^{\gamma_i z c}$ (IIb)

The effect of increased exposure estimates prior to 1953 in these models was absorbed more by the coefficients of reference age categories. Risk increased with reference age. Estimates of ERR parameters under TSE exposure windows showed little variation in all other than the window of 25 years plus, perhaps reflecting that these early exposures had been experienced mainly 25 years prior to the reference age. Further examination revealed that 3 of the cases (60%) who experienced some exposure prior to 1953 died within 25 years of this time; 25 (6%) of the controls experienced exposures prior to 1953 and within 25 years of their reference ages. Of all those who experienced exposures prior to 1953 and within 25 years of their reference ages, only 3 controls accumulated their total exposure before 1953.

Exposure patterns were further examined to see whether the lack of effect due to variation in pre-1953 exposures could be attributed to a specific aspect of individual dosimetry. Among those ever exposed the mean proportion of exposure experienced prior to 1953 was 8.7% for cases and 10.8% for controls. Of those exposed prior to 1953, the proportion of exposure experienced before 1953 averaged 55.7% among the cases and 62.7% among the controls; 21 of the controls and one case accumulated all their exposure during this period. The lack of an effect in varying the early exposures may have been because the majority of those working prior to 1953 continued working at RH thereafter, and the fraction of their exposure experienced prior to 1953 was small.

Results of fitting models to CE with zero weightage ascribed in turn to the periods prior to 1953 and after 1962 are summarized in table 7.4.3. All models under the zero exposure assumption had better fits than the reference model (model based on direct extrapolation, *i.e.* multiplier of 1). The best fit was obtained when exposures prior to 1953 were estimated by 1953 exposures and no exposures were ascribed to 1962. However, the differences between the deviances of these models was very small. Estimates of ERR/WLM ranged from 1.75% to 3.08% under these assumptions, after accounting for the effect of reference age. Estimates of RR amongst reference age categories were much more precise when modelled under the zero exposure assumption prior to 1953.

Table 7.4c: Results from Sensitivity Analyses With Zero Exposure Multipliers

(95% Confidence Intervals for Parameter Estimates Given in Brackets)

MODEL PARAMETERS	REFERENCE NO EXPOSURE MULTIPLIERS	ZERO EXPOSURE MULTIPLIERS		
		1962 ONLY	1948-52 ONLY	1948-52 AND 1962
Exposure:				
w5_ e^{β}	0.0175 (0.001 - 0.388)	0.0175 (0.001- 0.387)	0.0307 (0.002 - 0.432)	0.0308 (0.002 - 0.432)
Reference Age:				
<55 years e^{γ_1}	0.9012 (0.006 - 136.7)	1.0050 (0.008 - 119.1)	0.8070 (0.013 - 50.84)	0.8099 (0.013 - 50.62)
55-<65 years e^{γ_0}	1 (Fixed)	1 (Fixed)	1 (Fixed)	1 (Fixed)
65-> years e^{γ_2}	2.926 (2.925 - 119.1)	2.9220 (0.072 - 118.6)	2.4100 (0.081 - 71.27)	2.4040 (0.081 - 70.94)
Deviance	279.397	279.257	279.350	279.341

Note: Cumulative Relevant Exposure (w5_): $ERR = w5\_\, e^{\beta}$ (Ia)

Modifying Effect of Reference Age (z): $ERR = w5\_\, e^{\beta}\, e^{\gamma_1 z_c}$ (Ib)

Sensitivity analyses carried out thus far comprised examination of fixed multipliers of Rn exposure estimates obtained through extrapolation. The weightage of exposures experienced during the periods in question are now estimated using maximum likelihood estimation.

Results of maximum likelihood estimation of individual exposure weights for years prior to 1953, showed the largest (though non-significant) contributor to be 1949. However, this may have been spurious due to the few controls exposed that year and the lack of variability between cases and controls in exposure levels. Closer examination of the data revealed that the cases and controls exposed in 1949 referred to the same individual who was randomly selected as a matched control in two risk sets before becoming a case.

Based on relevant exposures accumulated during the periods prior to and after 1953, MLE showed that exposure from 1953 to 1962 plus 0.3416 times the cumulative exposure prior to 1953 had the best fit (deviance 279.160). Maximum likelihood estimation of a exposure weight for exposures experienced in 1962 did not yield any worthwhile estimate.

After accounting for the effect of age at risk, the ERR/WLM was 1.76% using the exposures with simple extrapolation and 3.08% when no exposures were ascribed to the calendar periods in question. Given any greater weightage, these estimates fluctuate non- systematically with exposure and increase with reference age. Maximum Likelihood Estimation shows that the best fit is obtained with weights greater than zero but less than 1. Therefore we conclude that the ERR /WLM must lie between 1.76% - 3.08%.

7.5 Further Examination of Surrogate Measures of Exposure: The Role of Surrogate Measures in TSE Windows

The predictive power of surrogate measures of exposure was examined by fitting identical models to each surrogate measure and comparing their goodness-of-fit. Surrogate measures examined comprised cumulative duration of employment, cumulative duration of exposure and average intensity of exposure. Cumulative durations of employment and exposure were measured in working months and intensity of exposure in working levels. All modelling was based on continuous variables. In keeping with analyses presented in previous sections of this chapter, surrogate measures were modelled under two alternate sets of TSE windows models - the first consisting of just one window based on a 5 year lag, and the other comprising two windows of 5-15 years prior and 15 or more. Since models for the various surrogate measures are not nested models, their goodness-of-fit was examined by a direct comparison of the model deviances with the deviance of the reference models - models of Rn exposure measured in working level months.

Results of fitting these models (summarized in table 7.5a) showed that cumulative duration of exposure had the smallest deviances in both sets of models; thus, implying that cumulative duration of exposure may have been a better linear predictor of Rn progeny related lung cancer mortality after controlling for birth cohort effect, than estimated cumulative relevant Rn progeny exposure. Estimates obtained showed a 3.61% ERR/WM of cumulative duration exposed in the 5 year lagged period, as compared with 2.75% ERR/WLM in the reference model based on cumulative relevant Rn progeny exposure. Further stratification of the TSE showed that compared to exposures sustained in previous periods, the RR of exposures sustained in the 5-15 year TSE window was 18.0 based on WLMs and 11.7 based on WMs of exposure.

Table 7.5a: Results from Modelling Surrogate Measures of Exposure Excess Relative Risk per Unit Increase in Measure of Exposure

Exposure in Time Since Exposure Window		Reference Model Exposure WLM	Surrogate Measures: Duration Employed WM	Surrogate Measures: Duration Exposed WM	Surrogate Measures: Exposure Intensity WL
w5_	e^{β}	0.0275	0.0038	0.0361	-0.2225
	SE(e^{β})	(0.0208)	(0.0071)	(0.0227)	(0.1227)
Deviance		279.992	283.655	277.361	283.447
w5_15	e^{β_0}	0.3334	0.0075	0.2761	4.699
	SE(e^{β_0})	(0.3698)	(0.0245)	(0.2818)	(4.699)
w15_	β_1	0.0185	0.0035	0.0236	-0.2271
	SE(β_1)	(0.0189)	(0.0077)	(0.0198)	(0.1274)
Deviance		277.290	283.632	274.521	280.193

Note: Cumulative Relevant Exposure (w5_): $ERR = w5\_\ e^{\beta}$ (Ia)

Exposure in TSE Windows: $ERR = (w5\_15 + \beta_1 w15\_)\ e^{\beta_0}$ (IIa)

The role of duration of exposure as the best fitting surrogate measure was further examined in the presence of the modifying effect of reference age. Results of fitting these models are summarized in table 7.5b.

Compared to the reference models, models based on duration of exposure continued to have smaller deviances; they also had consistently smaller and more precise estimates of risk (smaller variances). The change in RR estimates across reference age categories indicated an increasing trend in risk with reference age. However, a score test for trend based on continuous reference age showed that this trend was not statistically significant.

The other surrogate measures examined proved to be no better than the reference measure. The negative coefficient of ERR/WL obtained in modelling exposure intensity indicated that higher intensities of exposure may have had somewhat of a protective effect from lung cancer mortality. The results were however, not strong enough to confirm this notion.

Table 7.5b: Further Comparison of Reference Model and Best Surrogate Model
(Standard Errors or 95% Confidence Bounds)

Effect Estimated and Goodness of Fit	Reference Model Exposure (WLM)	Surrogate Measure Duration Exposed (WM)
TSE Window:		
w5_ e^{β}	0.0175	0.0009
SE(e^{β})	(0.0276)	(0.0010)
Reference Age:		
<55 years $e^{\gamma 1}$	0.9012	0.5445
SE($e^{\gamma 1}$)	(0.01 - 136.7)	(0.01 - 21.29)
55-65 years $e^{\gamma 0}$	1	1
SE($e^{\gamma 0}$)	(Fixed)	(Fixed)
65-> years $e^{\gamma 2}$	2.926	2.130
SE($e^{\gamma 2}$)	(0.07 - 119.2)	(0.13 - 33.98)
Deviance	279.397	276.554
TSE Window:		
w5_15 $e^{\beta 0}$	0.2524	0.0104
SE($e^{\beta 0}$)	(0.6529)	(0.0206)
w15_ $\beta 1$	0.0157	0.0008
SE($\beta 1$)	(0.0266)	(0.0009)
Reference Age:		
<55 years $e^{\gamma 1}$	1.157	0.5413
SE($e^{\gamma 1}$)	(0.01 - 152.8)	(0.01 - 23.0)
55-65 years $e^{\gamma 0}$	1	1
SE($e^{\gamma 0}$)	(Fixed)	(Fixed)
65-> years $e^{\gamma 2}$	2.028	1.722
SE($e^{\gamma 2}$)	(0.03 - 128.4)	(0.09 - 34.35)
Deviance	277.161	273.867

Note: Modifying Effect of Reference Age (z): ERR = $w5\_\ e^{\beta}\ e^{\gamma_1 z} c$ (Ib)
Modifying Effect of Reference Age: ERR = $(w5\_15 + \beta_1 w15\_)\ e^{\beta 0}\ e^{\gamma_i z} c$ (IIb)

The surrogate role of duration of employment was also considered in conjunction with a dichotomous variable defining exposure status as ever or never effectively exposed. Both TSE windows models fitted, though only slightly improving the fit with duration of employment alone, consistently estimated RR 46 times higher among those ever exposed as against those never exposed. However, the inclusion of the new variable on exposure status did not prove any marked improvement in the model fit (deviances of 283.273 and 283.009).

7.6 Further Examination of the Protracted Exposure Effect

A *protracted exposure effect* - where, for a given level of cumulative Rn exposure the risk of lung cancer mortality declined with increasing Rn concentrations - identified during person-years based Poisson regressions analyses in chapter 6, was further examined using the proportionate hazards modelling under nested case-control analyses. Analysis was based on complete risk sets corresponding to those exposed to Rn progeny at RH (similar to the dataset used in section 7.3.1). The effect of protracted exposure was examined by modelling intensity of exposure (Rn progeny concentrations) as a modifier of the effect of cumulative relevant Rn progeny exposure. Results of these examinations summarized in table 7.6.1 showed that the intensity of exposure had a modifying effect on the association between cumulative relevant Rn progeny exposure and lung cancer mortality.

The modifying effect of Rn progeny concentration was best represented by a linear additive model. Estimates from this model showed that for a given cumulative exposure, the ERR declined by 6% with each increasing WL of exposure.

The intensity of exposure modelled in this section was derived from cumulative relevant exposure and total duration of exposure to Rn progeny at RH, which therefore represent the overall average intensity of exposure. Analyses presented in chapter 6 however, were based on person-years data where person-years, events and cofactors were accumulated by each individual's exact age (integer) at each calendar year and covariates were obtained through weighted averages for each cell in the person-years and event tables. The person-years based analyses would therefore have captured and retained the essence of exposure intensities over the time they were experienced; whereas, in exposure intensities derived from cumulative exposure and duration, fluctuations in exposure conditions would have been lost in the summarization of a single average. Person-years based analyses are therefore preferred to nested case-control analyses based on life-time average exposure concentrations for examining protracted exposure effect.

Table 7.6.1: Further Examination of the Protracted Exposure Effect Base on Nested Case-Control Analyses

Risk Function	Parameter Estimates			Model Fit		LR Test		
	Par	Est	95% CB	Dev	DF	χ^2	DF	p
NULL:				181.45	32			
Linear Model:								
RR = $(1+\beta w)$	β	0.029		177.93	31	3.52^n	1	**0.06**
RR = $(1+\beta w + \gamma wc)$	β	0.114		173.81	30	3.12^l	1	**0.08**
	γ	-0.056						
Exponential Model:								
RR = $e^{\beta w}$	e^{β}	1.018	1.00-1.04	177.61	31	3.84^n	1	**0.05**
RR = $e^{\beta w + \gamma wc}$	e^{β}	1.054	1.01-1.10	174.50	30	3.11^e	1	**0.08**
	e^{γ}	0.972	0.94-1.00					
Linear-Exponential Model:								
RR = $(1+\beta w)$	β	0.029		177.93	31	3.52^n	1	**0.06**
RR = $(1+\beta we^{\gamma c})$	e^{λ}	0.181	0.01-2.44	174.14	30	3.79^l	1	**0.05**
	β	0.261						
Power Model:								
RR = $(1+\beta wc^{\nu})$	ν	-1.56		174.28	30	3.65^l	1	**0.06**
	β	0.040						

Note: *w* - Rn Progeny Exposure; β - ERR/WLM; c - Intensity of Rn Progeny Exposure.
^{n} LRT: Improvement on Null Model; ^{e} LRT Improvement on Exponential Model;
^{l} LRT Improvement on Linear Model.

Other more powerful methods of examining the protracted exposure effect using nested case-control analyses are available; these include the modelling of individual intensities by age or calendar time, the identification and modelling of peak exposures. In an exploratory analysis, annual exposures were modelled using the proportionate hazard procedure (PHGLM) in SAS (SAS 1990). Limitations of available data - identified in section 7.4 - hampered this approach and this analysis was no more informative than PYRS based analyses. Analyses based on peak exposures could not be undertaken for the same reason. Such analyses must await improved exposure data and radio-biological guidelines on effective exposure.

7.7 Examination of the Effect Other Occupational Exposures

Other occupational exposures relevant to the risk of Rn associated lung cancer mortality studied in this work comprise other occupational exposure to radioactive material and to asbestos. Descriptive findings on these exposures were summarized in chapter 4. Of the 54 lung cancer deaths in the RH study, data on other occupational exposure to radioactive material were available for 29 (54%) cases - 2 were exposed to radioactive material outside of RH - and data on occupational exposure to asbestos were available for 24 cases (44%) - 8 had been exposed to asbestos, and 5 of this number were exposed to Rn at RH. Jointly, of 30 lung cancer deaths 10 had other relevant occupational exposures and 7 were exposed to Rn at RH.

Considering the limited number of informative risk sets available for this part of the study and the lack of complete histories on other occupational exposures amongst controls, all available matched controls within each risk set who responded to the questionnaire survey queries on other occupational exposures were included in the analyses. The nested case-control analyses of other occupational exposures was therefore not based on randomly sampling the risk sets as in the previous sections of this work.

For the purpose of this analysis, other occupational exposure to radioactive material and to asbestos were treated as independent dichotomous variables. Initial examination showed that data available on these variables were adequate to independently address their effects. It was not necessary to combine them to form a composite variable on other occupational exposures as suggested in the descriptive epidemiology section of this thesis (chapter 4). Independent evaluation of these factors was preferred in order to avoid the difficulties in interpreting results based on a composite variable.

Analyses commenced with simple categorical evaluation based on multiplicative (log-linear) RR models; this was followed by more detailed examination of variations in ERR/WLM with Rn progeny exposure modelled using a smooth parametric function.

7.7.1 The Effect of Occupational Exposure to Radioactive Material Other Than at Radium Hill

Table 7.7.1a shows the distribution of cases and controls within the nested case-control dataset constructed for examining the effect of other occupational exposure to radio active material on the association between Rn progeny exposure at RH and lung cancer mortality.

Table 7.7.1a: Distribution of Cases and Controls By Exposure Status

Other Occupational Exposure to Radio Active Material	Radon Progeny Exposure at Radium Hill					
	Unexposed		Exposed		All	
	Cases	Controls	Cases	Controls	Cases	Controls
Unexposed	10	332	17	470	27	802
Exposed	0	59	2	85	2	144
All	10	391	19	555	29	946

The absence of any (table 7.7.1a) cases amongst those unexposed to Rn progeny at RH and occupationally exposed to radioactive material elsewhere, made categorical evaluation based on this dichotomous categorization of Rn progeny exposure at RH unfeasible. Therefore, examination of the effect of other occupational exposure to radioactive material was based on variations in ERR/WLM by modelling Rn progeny exposure at RH quantitatively.

Results of analyses summarized in table 7.7.1b showed that occupational exposure to radioactive material other than at RH had a significant residual effect on lung cancer mortality after accounting for the effect of Rn progeny exposure at RH ($p=0.05$). Other occupational exposure to radioactive material was found to have a significant confounding effect on the risk of lung cancer associated with exposure to Rn progeny at RH. After controlling for the confounding effect of other occupational exposures to radioactive material, the estimated increase in ERR/WLM associated with Rn progeny exposure experienced as RH was 11%; this estimated effect was slightly (not significantly) modified by the effect of attained age.

The effect of other occupational exposure to radioactive work examined with Rn exposure at RH expressed under TSE windows, improved the fit of the model considerably; ERR/WLM estimates were much closer to the overall estimates derived from previous examination. Risk of lung cancer mortality associated with distant exposures was 7% ERR/WLM, and was further modified by the influence of attained age (3% ERR/WLM in those aged <65 years); this modifying effect of attained age was however, not statistically significant.

Table 7.7.1b: Examining the Confounding Effect of Occupational Exposure to Radioactive Material Other Than at Radium Hill

Risk Function	Parameter Estimates			Model Fit		LR Test		
Models of Relevant Exposure:	**Par**	**Est.**	**95% CB**	**Dev**	**DF**	**χ^2**	**DF**	**p**
Null				187.36	27			
$(1+w_{5\_}e^{\beta})$	e^{β}	0.09	0.03-0.32	177.71	26	9.65	1	**<0.01**
$e^{\alpha x}(1+w_{5\_}e^{\beta})$	e^{α}	0.26	0.05-1.26	173.93	25	3.78	1	**0.05**
	e^{β}	0.11	0.03-0.38					
$e^{\alpha x}(1+w_{5\_}e^{\beta}\, e^{\gamma z})$	e^{α}	0.24	0.05-1.25	173.74	24	0.19	1	0.66
	e^{β}	0.09	0.02-0.42					
	e^{γ}	1.70	0.12-24.28					
Models Under TSE Windows:								
Null				187.36	27			
$[1+(w_{5\text{-}15}+\beta_1 w_{15\_})e^{\beta_0}]$	e^{β_0}	2.72	0.31-23.54	171.72	25	15.64	2	**<0.01**
	β_1	0.02						
$e^{\alpha}[1+(w_{5\text{-}15}+\beta_1 w_{15\_})e^{\beta_0}]$	e^{α}	0.29	0.06-1.43	168.69	24	3.03	1	**0.08**
	e^{β_0}	3.54	0.34-3.51					
	β_1	0.02						
$e^{\alpha x}[1+(w_{5\text{-}15}+\beta_1 w_{15\_})e^{\beta_0 e^{\gamma z}}]$	e^{α}	0.26	0.05-1.37	167.87	23	0.82	1	0.37
	e^{β_0}	2.94	0.30-29.13					
	β_1	0.01						
	e^{γ}	3.74	0.14-101.0					

Note: x - Other occupational exposure to radioactive material: α - confounding parameter;
w - Rn Progeny Exposure: e^{β} ERR/WLM
z - Attained age (reference age): γ – effect modification parameter

The modifying effect of other occupational exposure to radioactive material (table 7.7.1c) could not be effectively examined due to the sparseness of the data in the distribution of cases.

Table 7.7.1c: Examining the Modifying Effect of Occupational Exposure to Radioactive Material Other Than at Radium Hill

Risk Function	Parameter Estimates			Model Fit		LR Test		
	Par	Est.	95% CB	Dev	DF	χ^2	DF	p
Models of Relevant Exposure:								
$e^{\alpha x}$	e^{α}	0.39	0.09-1.69	185.32	26			
$e^{\alpha x}(1+w_5 e^{\beta})$	e^{α}	0.26	0.05-1.26	173.93	25	11.39	1	**<0.01**
	e^{β}	0.11	0.03-0.38					
$e^{\alpha x}(1+w_5 e^{\beta} e^{\theta x})$	e^{α}	*	*	172.27	24	1.66	1	0.20
	e^{β}	0.09	0.02-0.37					
	e^{θ}	*	*					
$e^{\alpha x}(1+w_5 e^{\beta} e^{\theta y} e^{\gamma z})$	e^{α}	*	*	172.02	24[a]			
	e^{β}	0.07	0.01-0.36					
	e^{θ}	*	*					
	e^{γ}	1.84	0.13-26.73					
Models Under TSE Windows:								
$e^{\alpha x}$				185.32	26			
$e^{\alpha x}[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta 0}]$	e^{α}	0.29	0.06-1.43	168.69	24	16.63	2	**<0.01**
	$e^{\beta 0}$	3.45	0.34-35.00					
	β_1	0.02						
$e^{\alpha x}[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta 0} e^{\theta x}]$	e^{α}	*		166.96	23	1.73	1	0.19
	$e^{\beta 0}$	2.19	0.24-19.89					
	β_1	0.03						
	e^{θ}	*						
$e^{\alpha x}[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta 0} e^{\theta x} e^{\gamma z}]$	e^{α}	*		165.89	23[a]			
	$e^{\beta 0}$	1.84	0.21-16.00					
	β_1	0.01						
	e^{θ}	*						
	e^{γ}							

Note: x - Occupational exposure to radioactive work other than at Radium Hill:
α - confounding parameter; θ - effect modification parameter;
w - Rn Progeny Exposure: e^{β} ERR/WLM
z - Attained age (reference age): γ – effect modification parameter
IR - Informative Risk Sets; * - > 0 But, Very Nearly 0.00; [a] - Aliased, Not Estimable.

7.7.2 The Effect of Occupational Exposure to Asbestos

The distribution of cases and controls by asbestos exposure and Rn progeny exposure status within the nested case-control dataset derived for this analysis is given in table 7.7.2a.

Table 7.7.2a: Distribution of Cases and Controls By Exposure Status

Occupational Exposure to Asbestos	Radon Progeny at Radium Hill					
	Unexposed		Exposed		All	
	Cases	Controls	Cases	Controls	Cases	Controls
Unexposed	5	279	11	380	16	659
Exposed	3	27	5	65	8	92
All	8	306	16	445	24	751

Results of analysis based on multiplicative RR models with categorical data (table 7.7.2b) showed that those occupationally exposed to asbestos were at significantly greater risk (over four times) of dying from lung cancer compared to those unexposed to asbestos.

Table 7.7.2b: Examining the Effect of Occupational Exposure to Asbestos Using Multiplicative RR Models and Categorical Data
(95% Confidence Limits)

Occupational Exposure to Asbestos	Relevant Exposure to Rn Progeny at RH			
	No	Yes	Overall	
			Unadjusted	Adjusted[r]
No	1 (Fixed)	1.91 (0.7 - 5.6)	1 (Fixed)	1 (Fixed)
Yes	8.16 (1.6 - 40.4)	6.38 (0.7 - 4.3)	4.65 (1.8 - 11.9)	4.44 (1.7 - 11.3)
Overall: Unadjusted	1 (Fixed)	1.73 (0.7 - 4.3)		
Overall: Adjusted[a]	1 (Fixed)	1.5 (0.6 - 3.7)		

Note: [a] Adjusted for Asbestos; [r] Adjusted for Rn Progeny Exposure

The interaction between Rn progeny exposure at RH and occupational exposure to asbestos was more additive than multiplicative. It was also noted that the high RR estimate amongst those occupationally exposed to asbestos alone - being based only on 3 risk sets and lung cancer cases and subjected to considerable lack of precision - should be interpreted with caution. Analysis of deviance for this categorical evaluation showed that occupational exposure to asbestos alone contributed the major effect on lung cancer mortality. After accounting for the main effect of occupational exposure to asbestos, neither the main effect of Rn exposure at RH nor the interaction effect were statistically significant.

Further examination of the effect of asbestos with Rn progeny exposure at RH being modelled as a continuous variable showed that occupational exposure to asbestos was a significant confounder of the risk of lung cancer mortality associated with Rn exposure (table 7.7.2c). However, exposure to Rn progeny (represented quantitatively) was significantly associated with lung cancer mortality after controlling for the confounding effect of occupational exposure to asbestos; the ERR/WLM was estimated at 8% before the modifying influence of attained age was accounted for and 6% thereafter. This effect had been masked by the dichotomous categorization of Rn progeny exposure in the previous analyses (table 7.7.2b). Risk evaluation under TSE windows showed that recent exposures to Rn progeny were associated with over twice the risk of distant exposures. Inclusion of an indicator of the temporal essence of Rn exposure showed an ERR/WLM of between 2% - 4% (in the distant and recent windows of time since exposure). Attained age was not found to have a significant modifying effect; however, there was some indication that those above the age of 65 were at greater risk of dying from Rn progeny associated lung cancer than those younger.

Findings summarized in table 7.7.2d showed that other occupational exposure to asbestos was not a significant effect modifier after it had been controlled for as a confounder. Other findings on effect modification arising from asbestos and attained age did not appear to be stable enough to draw any worthwhile inferences.

Table 7.7.2c: Examining the Confounding Effect of Occupational Exposure to Asbestos

Risk Function	Parameter Estimates			Model Fit		LR Test		
	Par	Est	95% CB	Dev	DF	χ^2	DF	p
Models of Relevant Exposure:								
Null				152.44	24			
$(1+w_{5\text{-}}e^{\beta})$	e^{β}	0.08	0.02-0.33	145.03	23	7.41	1	**0.01**
$e^{\alpha x}(1+w_{5\text{-}}e^{\beta})$	e^{α}	4.83	1.84-12.64	135.97	22	9.06	1	**<0.01**
	e^{β}	0.08	0.02-0.34					
$e^{\alpha x}(1+w_{5\text{-}}e^{\beta}\, e^{\gamma z})$	e^{α}	4.96	1.89-13.05	135.34	21	0.63	1	0.43
	e^{β}	0.06	0.01-0.36					
	e^{γ}	3.25	0.15-69.03					
Models Under TSE Windows:								
Null				152.44	24			
$[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta_0}]$	e^{β_0}	2.08	0.22-19.65	141.80	22	10.64	2	**<0.01**
	β_1	0.03						
$e^{\alpha x}[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta_0}]$	e^{α}	4.99	1.86-13.40	132.71	21	9.09	1	**<0.01**
	e^{β_0}	2.69	0.29-25.25					
	β_1	0.03						
$e^{\alpha x}[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta_0}\, e^{\gamma z}]$	e^{α}	5.14	1.90-13.87	131.72	20	0.99	1	0.32
	e^{β_0}	2.22	0.23-20.71					
	β_1	0.02						
	e^{γ}	5.05	0.14-178.9					

Note: x - Occupational exposure to Asbestos: α - confounding parameter;
w - Rn Progeny Exposure: e^{β} ERR/WLM
z - Attained age (reference age): γ – effect modification parameter

Table 7.7.2d: Examining the Modifying Effect of Occupational Exposure to Asbestos

Risk Function	Parameter Estimates			Model Fit		LR Test		
	Par	Est	95% CB	Dev	DF	χ^2	DF	p
Models of Relevant Exposure:								
$e^{\alpha x}$	e^{α}	4.65	1.01-11.93	143.57	23			
$e^{\alpha x(1+w_5 e^{\beta})}$	e^{α}	4.83	1.84-12.64	135.97	22	7.60	1	**<0.01**
	e^{β}	0.08	0.02-0.34					
$e^{\alpha x(1+w_5 e^{\beta} e^{\theta x})}$	e^{α}	5.25	1.41-19.45	135.93	21	0.04	1	0.84
	e^{β}	0.09	0.02-0.48					
	e^{θ}	0.76	0.03-16.3					
$e^{\alpha x(1+w_5 e^{\beta} e^{\theta x} e^{\gamma z})}$	e^{α}	2.85	0.79-10.29	134.23	20	1.70	1	0.19
	e^{β}	0.00	0.00-0.07					
	e^{θ}	48.93	1.75-1367					
	e^{γ}	183.2	2.77-12120					
Models Under TSE Windows:								
$e^{\alpha x}$	e^{α}	4.65	1.01-11.93	143.57	23			
$e^{\alpha x[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta_0}]}$	e^{α}	4.99	1.86-13.40	132.71	21	10.86	2	**<0.01**
	e^{β_0}	2.69	0.29-25.25					
	β_1	0.03						
$e^{\alpha x[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta_0} e^{\theta x}]}$	e^{α}	7.57	1.95-29.42	132.13	20	0.58	1	0.45
	e^{β_0}	6.41	0.59-70.34					
	β_1	0.02						
	e^{θ}	0.34	0.02-4.89					
$e^{\alpha x[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta_0} e^{\theta x} e^{\gamma z}]}$	e^{α}	6.58	1.82-23.83	131.529	19	0.60	1	0.44
	e^{β_0}	4.2	0.38-46.93					
	β_1	0.01						
	e^{θ}	0.45	0.03-7.37					
	e^{γ}	3.81	0.10-141.2					

Note: x - Occupational exposure to Asbestos: α - confounding parameter;
w - Rn Progeny Exposure: e^{β} ERR/WLM
z - Attained age (reference age): γ – effect modification parameter

7.8 Examination of the Role of Smoking

7.8.1 Construction of Categorical Variables on Smoking

Quantitative estimates of smoking in the RH study comprised the average duration, the average number of cigarettes smoked per day and the age at commencing regular smoking. Each individual was identified by current smoking status at the time of the questionnaire survey *i.e.* as *current smokers* or as *ex-smokers*. Methods of collecting smoking data were outlined in chapter 3 and descriptive epidemiological findings on individual smoking habits were comprehensively summarized in chapter 4.

Available data on individual smoking histories were limited; of the 54 lung cancer cases identified in the RH cohort, some data on smoking were available for 29 (54%) in varying degrees of completion; these comprised complete smoking histories - with quantitative estimates including average duration and amount smoked - for 12 (22%) lung cancer cases, quantitative estimates on one smoking parameter - either average duration or amount smoked - for a further 4 (8%), and the remaining 17 (24%) being identified only as to whether they had ever been regular smokers.

In the analyses of smoking, individuals were categorized as life-time non-smokers, those who quit smoking at least ten years prior to the time of observation, various categories based on cigarette consumption and smokers with unknown consumption. The basic variable used to quantify smoking was the composite variable of amount and duration - *pack years*, defined as the multiple of average number of packs (20 cigarettes to a pack) smoked per day and the duration of regular smoking (in years). Individual smoking characteristics used in this study are summarized in column 1 of table 7.8.1a.

Table 7.8.1a: Summary of Smoking Characteristics

Category	Cases		Controls		Total		
	Number	(%)	Number	(%)	Number	(%)	Pack-Years
Never Smokers	1	(3%)	236	(14%)	237	(14%)	0
Ex-smokers: Quit > 10 Years	0	(0%)	275	(17%)	275	(16%)	10.57
Cumulative Consumption:							
0<-6 Pack-Years	3	(10%)	32	(2%)	35	(2%)	3.33
6<-45 Pack-Years	4	(14%)	539	(33%)	543	(33%)	29.11
Over 45 Pack Years	4	(14%)	409	(25%)	413	(25%)	73.32
Unknown	17	(59%)	151	(9%)	168	(10%)	
Total	29	(100%)	1642	(100%)	1671	(100%)	

Note: No Smoking Data Available For: 25 cases and 289 controls.

The limited amount of smoking data available lead to improvisations in methods of studying the risk associated with smoking and Rn progeny exposure, so that the available data on smoking could be used to the fullest. Two such improvisations were adopted: firstly, a suitable approach to epidemiological design for risk assessment; and secondly, methods of estimating missing values for quantitative parameters on smoking.

- **Study Design:** Risk assessment was based on proportional hazards models and nested case-control analyses, where all possible matched controls in each risk set were included in the analyses, instead of sampling the risk sets and selecting a fixed number of randomized controls for each case; the analytical dataset included only the 29 risk sets with known smoking status for cases; within each of these risk sets, only controls with known smoking status were included; thus, 25 risk sets and a total of 289 controls were excluded from this part of the study.

- **Adjustment for Missing Values:** Two methods of estimating missing values for smoking parameters were examined and used in the case of the 17 individuals who were known to be smokers, but lacked complete quantitative smoking histories. These methods basically comprised using average estimates for those individuals with missing smoking parameters. Smokers with missing smoking parameters were identified as current or ex-smokers and the missing parameters were substituted with the median value observed within these groups. Adjustment for missing values were done in two stages: *single adjustment* - where only one quantitative smoking parameter was missing and therefore substituted with an estimate; *dual adjustment* - where both quantitative smoking parameters necessary for computing pack-years were missing and substituted for. The distribution of cases and controls categorized after estimation for missing values are summarized in table 7.8.1b, according to the methods of adjustment.

Table 7.8.1b: Distribution of Cases and Controls in Risk Set Data By Methods of Adjustment for Missing Values

Category	Method of Adjustment					
	None		Single		Dual	
	Cases	Controls	Cases	Controls	Cases	Controls
Never Smokers	1	236	1	236	1	236
Ex-smokers: Quit > 10 Years	0	275	0	275	0	275
Cumulative Consumption:						
0<-6 Pack-Years	3	32	7	52	7	52
6<-45 Pack-Years	4	539	4	539	17	670
Over 45 Pack Years	4	409	4	409	4	409
Unknown	17	151	13	131	-	-
Total	29	1642	29	1642	29	1642

Note: No Smoking Data Available For: 25 cases and 289 controls.

The number of cases amongst categories of *lifetime non-smokers* and *long term quitters* - quit over 10 years before reference age - were too small (only 1 case) for any worth while analyses; therefore, these categories were combined with *low smokers* (0<-6 pack-years) to form the baseline reference category for risk assessment and interpreted as *virtual non-smokers*.

7.8.2 Examining the Association Between Smoking, Radon Progeny Exposure and Lung Cancer Mortality

Table 7.8.2a shows the distribution of cases and controls by smoking and Rn progeny exposure status within the nested case control dataset derived for this analysis. Smokers were dichotomized as *virtual non-smokers* (defined in section 7.8.1) or other known smokers.

Table 7.8.2a: Distribution of Cases and Controls By Smoking and Exposure Status

Smoking Status	Radon Progeny at Radium Hill					
	Unexposed		Exposed		All	
	Cases	Controls	Cases	Controls	Cases	Controls
Virtual Non-Smoker	1	117	3	133	4	250
Smoker	9	235	16	298	25	533
All	10	352	19	431	29	783

Risk evaluation based on multiplicative (log-linear) RR models (table 7.8.2b) showed that smokers were at over twice the risk (p = 0.06) of lung cancer mortality than virtual non-smokers. The interaction between smoking and Rn progeny exposure was supra-additive and sub-multiplicative. Results of LRT showed that the interaction effect was not statistically significant after removing the main effects of Rn progeny exposure and smoking.

Table 7.8.2b: RR of Lung Cancer Mortality by Smoking and Exposure Status
(95% Confidence Limits)

Smoking Status	Relevant Exposure to Rn Progeny at RH			
	No	Yes	Overall	
			Unadjusted	Adjusted[r]
Virtual Non-Smoker	1 (Fixed)	3.25 (0.3 - 31.8)	1 (Fixed)	1 (Fixed)
Smoker	3.68 (0.4 - 30.2)	7.10 (0.9 - 54.6)	2.66 (0.9 - 7.6)	2.56 (0.9 - 7.6)
Overall: Unadjusted	1 (Fixed)	2.15 (0.9 - 5.0)		
Overall: Adjusted[s]	1 (Fixed)	2.08 (0.9 - 4.9)		

Note: [s] Adjusted for Smoking; [r] Adjusted for Rn Progeny Exposure.

7.8.3 The Role of Smoking as a Confounder

The role of smoking as a confounder of the exposure-response relationship between Rn progeny and lung cancer mortality was examined using smoking as a categorical variable in a log excess relative risk model; Rn exposure was considered first, as cumulative relevant exposure and then under the time-since-exposure windows of 5-15 years and 15 years or more. Estimates of risk unconfounded by smoking were then examined in conjunction with the modifying effect of attained age using the variable *reference age*, previously defined. Each of these factors examined were included in the model in a step-wise fashion, thereby enabling the evaluation of their additional contribution to the model fit. The effect of each factor added to the model was tested using the likelihood ratio tests. The confounding role of smoking was also examined with the smoking variable being categorized using the three categorizations outlined in the previous section - no adjustment, single adjustment and dual adjustments for missing values. Results of these analyses are summarized in tables 7.8.3a and 7.8.3b. In addition to the examination of smoking as a confounder, these analyses were also used as a means of evaluating the efficacy of the adjustment for missing values and for comparing between the two alternative methods of missing value estimation proposed.

Likelihood ratio test results summarized in table 7.8.3a showed that smoking was a significant confounder of the effect of Rn exposure on lung cancer mortality ($p<0.01$) and that, attained age did not significantly modify the effect of Rn exposure on lung cancer mortality after controlling for the confounding effect of smoking.

Risk estimates obtained with Rn progeny exposure expressed as cumulative effect exposure, showed a 7% increase in ERR/WLM before controlling for smoking; this estimate increased to 9% ERR/WLM after controlling for smoking as a confounder. With the inclusion of attained age into the model thereafter, the ERR/WLM in the reference group (attained age < 65 years) was estimated as 6%.

Table 7.8.3a: Examining the Confounding Effect of Smoking and Comparing Methods of Adjustment For Missing Values in Smoking Variables

Risk Characteristics	No Adjustment	Single Adjustment	Dual Adjustment
Null Deviance (DF)	179.92 (29)	179.92 (29)	179.92 (29)
RR = $(1+w_5 \cdot e^{\beta})$			
e^{β} (95% CI)	0.07 (0.02-0.28)	0.07 (0.02-0.28)	0.07 (0.02-0.28)
Deviance (DF)	172.20 (28)	172.20 (28)	172.20 (28)
LRT χ^2 (DF) p-value	7.72 (1) p=0.01	7.72 (1) p=0.01	7.72 (1) p=0.01
RR = $e^{\alpha_i x_c}(1+w_5 \cdot e^{\beta})$			
e^{α_0} (95% CI)	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{α_1} (95% CI)	1.31 (0.32-5.37)	2.24 (0.63-7.89)	4.13 (1.36-12.57)
e^{α_2} (95% CI)	1.27 (0.31-5.24)	1.56 0.40-6.03)	1.16 (0.40-6.04)
e^{α_3} (95% CI)	11.97 (3.59-39.92)	8.64 (2.57-29.0)	
e^{β} (95% CI)	0.09 (0.02-0.34)	0.08 (0.02-0.31)	0.09 (0.02-0.33)
Deviance (DF)	147.87 (25)	155.80 (25)	162.72 (26)
LRT χ^2 (DF) p-value	27.33 (3) p<0.01	16.40 (3) p<0.01	9.48 (2) p=0.01
RR= $e^{\alpha_i x_c}(1+w_5 \cdot e^{\beta} e^{\gamma z})$			
e^{α_0} (95% CI)	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{α_1} (95% CI)	1.41 (0.34-5.92)	2.35 (0.65-8.46)	4.29 (1.38-13.39)
e^{α_2} (95% CI)	1.31 (0.31-5.48)	1.60 (0.41-6.25)	1.58 (0.40-6.08)
e^{α_3} (95% CI)	13.53 (3.84-47.64)	9.33 (2.67-32.62)	
e^{β} (95% CI)	0.06 (0.01-0.35)	0.06 (0.01-0.34)	0.08 (0.01-0.37)
e^{γ} (95% CI)	3.11 (0.20-48.20)	2.26 (0.14-13.81)	1.71 (0.10-29.27)
Deviance (DF)	144.13 (24)	155.41 (24)	162.54 (25)
LRT χ^2 (DF) p-value	0.74 (1) p=0.39	0.39 (1) p=0.53	0.18 (1) p=0.67

Note: x - Smoking: α - confounding parameter.
w - Rn Progeny Exposure: e^{β} ERR/WLM.
z - Attained age (reference age): γ – effect modification parameter.

Even though attained age was not a statistically significant modifier of the exposure-response relationship after controlling for the confounding effect of smoking, those aged above 65 years at the time of reference (reference age or attained age) were seen to be at greater risk or Rn progeny associated lung cancer mortality than those below the age of 65. The influence of the potentially modifying effect of attained age increased the risk due to smoking and reduced the risk due to Rn progeny exposure; this indicated that the effect of attained age caused a slight shift of the onus of risk from Rn progeny exposure to smoking.

Estimates of RR showed that smokers were at greater risk of dying from lung cancer, even after accounting for their exposure to Rn progeny. However there appeared to be little difference between the categories of smokers whose smoking consumption could be quantified; *medium* (6<-45 pack-years) and *heavy* (>45 pack-years) smokers were about 30% greater risk than *virtual non-smokers,* but this excess was not statistically significant. The significance of the effect of smoking as a confounder stemmed from the category of smokers for whom complete quantitative estimates of smoking were not available; those in this category were at much greater risk than all other categories of smokers (12 times greater risk than virtual non-smokers). This finding suggested that category of smokers for whom quantitative estimates were not available may have included considerably more heavy smokers.

The effect of substituting estimates for missing values in parameters quantifying life-time smoking experience was a greater risk among medium smokers, whilst leaving the effect of cumulative relevant Rn progeny exposure relatively unaffected. Goodness-of-fit criteria for parallel models fitted under the three methods of quantifying smoking showed no improvement in model fit due to methods of adjusting for missing values; in fact, the model based on available data alone (without adjustment for missing values) provided the best fit. Results also indicated that serious biases may have stemmed from the more extreme case of dual adjustment.

Results on the confounding effect of smoking bore close resemblance under models of cumulative relevant Rn progeny exposure and Rn progeny exposure measured under TSE windows (table 7.8.3b). This indicated that the experience of Rn progeny exposure accounted for in time had little influence on the impact of smoking on lung cancer mortality.

Examination of risk under TSE windows showed that recent exposures to Rn progeny were more hazardous than distant exposure; even though estimates of RR of recent exposure compared to distant exposure were not statistically significant, this is a pertinent finding and the possibility of increased power for detecting a truly significant effect by improved smoking data should not be disregarded. Furthermore, risk estimates obtained under TSE models showed that the risk due to Rn progeny exposure was less than that estimated from cumulative relevant exposure - ERR/WLM estimates obtained from TSE models showed ERR/WLM estimates of 5% (unmodified by attained age) and 2% after accounting for effect modification by attained age; in the window of more distant exposures the corresponding estimates of ERR/WLM were 2% and 1%, respectively. This decrease in the ERR due to cumulative Rn progeny exposure when expressed under TSE windows is an expression of time-since-exposure itself - the temporal expression of the effect of Rn progeny exposure - where, after the specified latent period, the excess risk associated with cumulative Rn progeny exposure declines with time since exposure.

Table 7.8.3b: Examining the Confounding Effect of Smoking and Comparing Methods of Adjustment For Missing Values in Smoking Variables With Radon Progeny Exposure Expressed in TSE Windows

Risk Characteristics		No Adjustment		Single Adjustment		Dual Adjustment	
Null Model: Deviance(DF)		179.92	(29)	179.92	(29)	179.92	(29)
$RR=[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta_0}]$							
e^{β_0}	(95% CI)	3.09	(0.28-34.32)	3.09	(0.28-34.32)	3.09	(0.28-34.32)
β_1	(95% CI)	0.02		0.02		0.02	
Deviance	(DF)	167.92	(27)	167.92	(27)	167.92	(27)
LRTχ^2	(DF) *p*-value	12.00	(2) $p<0.01$	12.00	(2) $p<0.01$	12.00	(2) $p<0.01$
$RR=e^{\alpha_i x_c}[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta_0}$							
e^{α_0}	(95% CI)	1	(Fixed)	1	(Fixed)	1	(Fixed)
e^{α_1}	(95% CI)	1.22	(0.29-5.06)	2.03	(0.57-7.20)	3.19	(1.28-11.93)
e^{α_2}	(95% CI)	1.22	(0.29-5.04)	1.47	(0.39-5.71)	1.47	(0.38-5.69)
e^{α_3}	(95% CI)	11.09	(3.32-37.10)	9.15	(2.67-31.30)		
e^{β_0}	(95% CI)	2.60	(0.22-30.94)	9.72	(0.76-125.1)	4.26	(0.34-54.16)
β_1	(95% CI)	0.02		0.01			
Deviance (DF)		141.47	(24)	150.98	(24)	159.09	(25)
LRT: χ^2 (DF) *p*-value		26.45	(3) $p<0.01$	16.94	(3) $p<0.01$	8.83	(2) $p=0.01$
$RR=e^{\alpha_i x_c}[1+(w_{5\text{-}15}+\beta_1 w_{15\text{-}})e^{\beta_0}e^{\gamma z}$							
e^{α_0}	(95% CI)	1	(Fixed)	1	(Fixed)	1	(Fixed)
e^{α_1}	(95% CI)	1.37	(0.32-5.80)	2.20	(0.61-8.00)	4.23	(1.35-13.24)
e^{α_2}	(95% CI)	1.29	(0.31-5.38)	1.54	(0.39-60.6)	1.52	(0.39-5.95)
e^{α_3}	(95% CI)	13.35	(3.76-47.36)	10.49	(2.92-37.65)		
e^{β_0}	(95% CI)	1.85	(0.16-20.82)	7.56	(0.59-96.00)	3.14	(0.26-38.58)
β_1	(95% CI)	0.01		0.00*		4.38	(0.10-187.0)
e^{γ}	(95% CI)	9.43	(0.13-710.7)	6.45	(0.11-389.3)		
Deviance (DF)		139.67	(23)	149.71	(23)	158.26	(24)
LRT: χ^2 (DF) *p*-value		1.8	(1) $p=0.18$	1.27	(1) $p=0.25$	0.83	(1) $p=0.36$

Note: * <0.01 (Estimate = 0.003).

x - Smoking: α - confounding parameter.

w - Rn Progeny Exposure: e^{β} ERR/WLM.

z - Attained age (reference age): γ – effect modification parameter.

7.8.4 The Role of Smoking as an Effect Modifier

After controlling for its confounding effect smoking was not found to be a significant modifier of the effect of Rn progeny exposure (tables 7.8.4a and 7.8.4b). However, it was interesting to note that medium and heavy smokers were at much greater risk. The smaller modifying component of the RR estimate for smokers with unknown consumption indicated that this category may in fact, have comprised a fair distribution of reasonably low or medium smokers, contrary to the impression gleaned from the confounding component represented within the base-line risk. On the other hand, this may be an over interpretation of the findings, considering the borderline significance of the modifying effect.

Despite the lack of statistical significance in the modifying effect of smoking (after controlling for its confounding effect) examination based on cumulative relevant exposure showed an increasing trend in RR with increasing smoking - compared to virtual non-smokers, medium smokers were at about 3 times greater risk and heavy smokers were at over 10 times greater risk (table 7.8.4a). The inclusion of attained age to this model showed that those above the age of 65 faced over a three-fold greater risk; this modifying effect of attained age too, was statistically non-significant.

Models of Rn exposure in TSE windows showed an interesting change in the trends observed previously (analyses based on cumulative exposure - table 7.8.4a). Once Rn exposure was represented in TSE windows, after controlling for the confounding effect of smoking, there was little to be accounted for by smoking or attained age as effect modifiers; considering that smoking and attained age had demonstrated some effect modification on the effect of cumulative relevant Rn exposure (table 7.8.4a), the disappearance of these effects under the TSE windows suggested that the important effect had been explained by time since exposure to Rn progeny.

Table 7.8.4a: Examining the Modifying Effect of Smoking

Risk Characteristics	No Adjustment	Single Adjustment	Dual Adjustment
Null Deviance (DF):	179.92 (29)	179.92 (29)	179.92 (29)
RR = $e^{\alpha_i x_c}$			
e^{α_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{α_1} (95% CI):	1.18 (0.29-4.80)	1.99 (0.57-6.92)	3.57 (1.19-10.73)
e^{α_2} (95% CI):	1.19 (0.29-4.84)	1.42 (0.37-5.4)	1.34 (0.35-5.12)
e^{α_3} (95% CI):	10.88 (3.23-36.64)	7.94 (2.34-26.92)	
Deviance (DF):	153.70 (26)	164.39 (26)	171.70 (27)
LRT χ^2 (DF) p-value:	26.22 (3) $p<0.01$	15.53 (3) $p<0.01$	8.82 (2) $p=0.01$
RR = $e^{\alpha_i x_c(1+w_5 \cdot e^{\beta})}$			
e^{α_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{α_1} (95% CI):	1.31 (0.32-5.37)	2.24 (0.63-7.89)	4.13 (1.36-12.57)
e^{α_2} (95% CI):	1.27 (0.31-5.24)	1.56 0.40-6.03)	1.55 (0.40-6.04)
e^{α_3} (95% CI):	11.97 (3.59-39.92)	8.64 (2.57-29.0)	
e^{β} (95% CI):	0.09 (0.02-0.34)	0.08 (0.02-0.31)	0.09 (0.02-0.33)
Deviance (DF):	144.87 (25)	155.80 (25)	162.72 (26)
LRT χ^2 (DF) p-value:	... 8.83 (1) $p<0.01$	... 8.59 (1) $p<0.01$	... 8.98 (1) $p<0.01$
RR= $e^{\alpha_i x_c(1+w_5 \cdot e^{\beta} e^{\theta_i x_c})}$			
e^{α_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{α_1} (95% CI):	1.0 (0.14-7.06)	1.83 (0.35-9.42)	3.31 (0.82-13.41)
e^{α_2} (95% CI):	0.47 (0.04-6.20)	0.89 (0.12-6.50)	0.81 (0.11-5.89)
e^{α_3} (95% CI):	10.19 (2.18-47.56)	7.26 (1.54-34.33)	
e^{β} (95% CI):	0.04 (0.00-1.72)	0.04 (0.00-1.62)	0.04 (0.00-1.50)
e^{θ_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{θ_1} (95% CI):	2.49 (0.02-336.9)	2.12 (0.02-188.8)	2.16 (0.04-129.0)
e^{θ_2} (95% CI):	10.27 (0.08-1247.0)	5.30 (0.06-479.6)	7.13 (0.08-645.8)
e^{θ_3} (95% CI):	1.53 (0.02-128.8)	1.69 (0.02-149.2)	
Deviance (DF):	143.00 (22)	154.85 (22)	161.45 (24)
LRT χ^2 (DF) p-value:	1.87 (3) $p=0.60$	0.95 (3) $p=0.81$	1.27 (2) $p=0.53$
RR= $e^{\alpha_i x_c(1+w_5 \cdot e^{\beta} e^{\theta_i x_c} e^{\gamma z})}$			
e^{α_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{α_1} (95% CI):	0.93 (0.13-6.92)	1.89 (0.37-9.71)	3.44 (0.85-13.94)
e^{α_2} (95% CI):	0.48 (0.04-6.28)	0.89 (0.12-6.59)	0.81 (0.11-5.94)
e^{α_3} (95% CI):	11.60 (2.53-53.25)	7.95 (1.68-37.56)	
e^{β} (95% CI):	0.02 (0.00-1.16)	0.03 (0.00-1.40)	0.03 (0.00-1.35)
e^{θ_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{θ_1} (95% CI):	3.70 (0.03-397.0)	2.17 (0.02-196.4)	2.11 (0.04-123.3)
e^{θ_2} (95% CI):	10.32 (0.11-976.0)	5.56 (0.07-427.7)	7.40 (0.09-595.6)
e^{θ_3} (95% CI):	1.23 (0.01-120.5)	1.50 (0.01-152.3)	
e^{γ} (95% CI):	3.58 (0.26-49.78)	2.50 (0.16-39.13)	2.01 (0.13-31.59)
Deviance (DF):	141.97 (21)	154.27 (21)	161.13 (23)
LRT χ^2 (DF) p-value:	1.03 (1) $p=0.31$	0.58 (1) $p=0.45$	0.32 (1) $p=0.57$

Note: x - Smoking: α - confounding parameter. w - Rn Progeny Exposure: e^{β} ERR/WLM.
z - Attained age (reference age): γ – effect modification parameter.

Table 7.8.4b: Examining the Modifying Effect of Smoking With Radon Progeny Exposure Expressed Under TSE Windows

Risk Characteristics	No Adjustment	Single Adjustment	Dual Adjustment
Null Deviance (DF):	179.92 (29)	179.92 (29)	179.92 (29)
RR = $e^{\alpha_i x_c}$			
Deviance (DF):	153.70 (26)	164.39 (26)	171.7 (27)
LRT χ^2 (DF) p-value:	26.22 (3) $p<0.01$	15.53 (3) $p<0.01$	8.82 (2) $p=0.01$
RR = $e^{\alpha_i x_c [1+(w_{5-15} + \beta_1 w_{5-15})\, e^{\beta_0}]}$			
e^{α_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{α_1} (95% CI):	1.22 (0.30-5.06)	2.08 (0.59-7.33)	4.07 (1.33-12.46)
e^{α_2} (95% CI):	1.22 (2.06-5.04)	1.53 (0.40-5.84)	1.57 (0.40-6.08)
e^{α_3} (95% CI):	11.00 (1.85-37.10)	8.58 (2.55-28.80)	
e^{β_0} (95% CI):	2.52 (3.53-29.78)	2.45 (0.18-32.82)	1.88 (0.15-24.26)
β_1 (95% CI):	0.02	0.06	0.05
Deviance (DF):	141.47 (24)	153.27 (24)	159.89 (25)
LRT χ^2 (DF) p-value:	12.23 (2) $p<0.01$	11.2 (2) $p<0.01$	11.81 (2) $p<0.01$
RR= $e^{\alpha_i x_c [1+(w_{5-15} + \beta_1 w_{15-}) e^{\beta_0} e^{\theta_i x_c}]}$			
e^{α_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{α_1} (95% CI):	1.86 (0.20-16.91)	2.08 (0.28-15.55)	4.07 (0.77-21.44)
e^{α_2} (95% CI):	1.23 (0.10-14.53)	1.53 (0.18-13.82)	1.57 (0.21-11.44)
e^{α_3} (95% CI):	20.02 (2.84-141.0)	8.58 (1.30-56.74)	
e^{β_0} (95% CI):	9.11 (0.27-308.6)	2.45 (0.06-93.30)	1.88 (0.04-79.38)
β_1 (95% CI):	0.02	0.06	0.05
e^{θ_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{θ_1} (95% CI):	0.24 (0.00-13.87)	1.00 (0.03-30.81)	1.00 (0.04-23.30)
e^{θ_2} (95% CI):	0.87 (0.02-36.49)	1.00 (0.03-37.97)	1.00 (0.02-44.13)
e^{θ_3} (95% CI):	0.15 (0.00-4.68)	1.00 (0.04-28.33)	
Deviance (DF):	139.46 (21)	153.27 (21)	159.89 (23)
LRT χ^2 (DF) p-value:	2.01 (2) $p=0.37$	0.00 (2) NI	0.00 (2) NI
RR= $e^{\alpha_i x_c [1+(w_{5-15} + \beta_1 w_{15-}) e^{\beta_0} e^{\theta_i x_c} e^{\gamma z}]}$			
e^{α_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{α_1} (95% CI):	1.86 (0.02-16.93)	2.08 (0.28-15.59)	4.07 (0.77-21.45)
e^{α_2} 95% CI):	1.23 (0.10-14.53)	1.53 (0.18-12.82)	1.57 (0.21-11.45)
e^{α_3} (95% CI):	20.02 (2.84-141.2)	8.58 (1.30-56.81)	
e^{β_0} (95% CI):	9.11 (0.26-316.0)	2.45 (0.06-100.3)	1.88 (0.04-86.70)
β_1 95% CI):	0.02	0.06	0.05
e^{θ_0} (95% CI):	1 (Fixed)	1 (Fixed)	1 (Fixed)
e^{θ_1} (95% CI):	0.24 (0.00-14.16)	1.00 (0.03-32.56)	1.00 (0.04-25.02)
e^{θ_2} (95% CI):	0.87 (0.87-36.68)	1.00 (0.03-38.4)	1.00 (0.02-44.80)
e^{θ_3} (95% CI):	0.15 (0.00-4.70)	1.00 (0.03-29.19)	
e^{γ} (95% CI):	1.00 (0.05-19.11)	1.00 (0.07-14.86)	1.0 (0.05-20.64)
Deviance (DF):	139.46 (20)	153.27 (20)	159.89 (22)
LRT χ^2 (DF) p-value:	0.00 (1) NI	0.00 (1) NI	0.00 (1) NI

Note: x- Smoking: α - confounding parameter. w- Rn Progeny Exposure: e^{β} ERR/WLM. z- Attained age (ref. age): γ– effect modification parameter; NI- No Improvement.

Results from these analyses on the modifying effect of smoking after controlling for the confounding effect of smoking showed a 5% ERR/WLM associated with cumulative relevant Rn exposure; this estimate was modified by attained age and was 2% in the reference category of those aged less than 65 (table 7.8.4a). When these analyses were repeated with exposure expressed under TSE windows, it appeared that distant exposures to Rn progeny continued to yield a 2% ERR/WLM but, that recent exposures were much more (about 9 times) hazardous and together, Rn progeny exposures represented in time since exposure - *i.e.* the temporal effect of Rn progeny exposure - and the confounding effect of smoking, accounted for the major component of the variation observed in patterns of lung cancer mortality amongst those with known smoking histories. Compared to the findings from the TSE model in examining the confounding effect of smoking, it appeared that the influence of introducing smoking as an effect modifier over and above its confounding effect was to shift the onus of risk from attained age (RR reduced by 9 fold) to recent Rn exposure (RR increased by 9 fold). It is acknowledged that these results are based on extremely limited data. Therefore, it may be conservatively concluded that these findings should at least stimulate and justify further examination with improved data and further follow-up.

7.8.5 Modelling Variations in Excess Relative Risk Associated With Smoking and Radon Progeny Exposure

Continuous evaluation was based on pack years computed from life-time smoking experience; therefore, no distinction was made between current smokers and those who had quit smoking for more than 10 years - they too were represented by their life-time cumulative cigarette consumption. Due to the subjectivity of the proposed techniques of missing value estimation, analyses were based only on available data; this resulted in 11 risk sets, each comprising of smokers alone. Variations in ERR associated with smoking and Rn progeny exposure were examined using models comprising linear, exponential and power modification terms for smoking variables. In addition to the effect of cumulative smoking consumption represented in pack-years, the role of age at commencement of smoking was also examined.

Findings from these analyses (table 7.8.5) revealed risk estimates that were generally consistent with previous findings. The exponential effect modification model provided the best fit and estimated a 6% exponential (multiplicative) increase in ERR per pack-year, for a given level of Rn exposure. This modifying effect was marginally significant (p=0.13). The power model estimated a maximum likelihood estimate of the power parameter to be 0.8 made little improvement to the goodness-of-fit in the model, indicating that pack-years in itself (power parameter of 1) was an adequate representation of smoking consumption. These analyses supported the notion that the interaction of smoking and Rn progeny exposure was more multiplicative than additive.

Further effects of cumulative cigarette consumption could not effectively be examined in conjunction with other modifiers - attained age and age at commencement of smoking - or with time-since-exposure due the limited number of risk sets available for this purpose. Evaluation of the role of smoking is therefore concluded at this point.

Table 7.8.5: Variations in Excess Relative Risk Associated With Smoking and Radon Progeny Exposure

Risk Function	Parameter Estimates			Model Fit		LR Test		
	Par	ERR	95% CB	Dev	DF	χ^2	DF	p
Linear Models:								
$e^{\alpha x}$	e^{α}	0.99	0.96-1.01	60.64	10			
$e^{\alpha x}(1+w_5 e^{\beta})$	e^{α}	0.99	0.96-1.01	51.37	9	9.27	1	**<0.01**
	e^{β}	0.30	0.04-2.25					
$e^{\alpha x}[1+(w_5+\beta_1 w_5 x)e^{\beta_0}]$	e^{α}	0.97	0.93-1.02	50.41	8	0.96	1	0.33
	e^{β_0}	0.10	0.00-43.12					
	β_1	0.08						
Exponential Models:								
$e^{\alpha x}$	e^{α}	0.99	0.96-1.01	60.64	10			
$e^{\alpha x}(1+w_5 e^{\beta})$	e^{α}	0.99	0.96-1.01	51.37	9	9.27	1	**<0.01**
	e^{β}	0.30	0.04-2.25					
$e^{\alpha x}(1+w_5 e^{\beta} e^{\theta x})$	e^{α}	0.95	0.87-1.03	49.13	8	2.24	1	**0.13**
	e^{β}	0.06	0.00-1.25					
	e^{θ}	1.06	0.97-1.16					
Power Models:								
$e^{\alpha x}$	e^{α}	0.99	0.96-1.01	60.64	10			
$e^{\alpha x}(1+w_5 e^{\beta})$	e^{α}	0.99	0.96-1.01	51.37	9	9.27	1	**<0.01**
	e^{β}	0.30	0.04-2.25					
$e^{\alpha x}(1+w_5 e^{\beta} x^{\theta})$	e^{α}	0.97	0.92-1.03	50.76	8	0.61	1	0.43
	e^{β}	0.02	0.00-33.81					
	θ	0.81						

Note: w - Rn Progeny Exposure; e^{β} ERR/WLM.
Smoking: x - pack-years:
α - confounding parameter; θ - effect modification parameter.

7.8.6 Conclusions on the Role of Smoking

Risk evaluation based on simple multiplicative (log-linear) RR models using categorical data showed that smokers were found to be at significantly greater risk of lung cancer mortality than virtual non-smokers, after explaining the main effect of Rn progeny exposure. The interaction between smoking and Rn progeny exposure on the outcome of lung cancer mortality was supra-additive and sub-multiplicative.

Smoking was identified as a significant confounder of the association between Rn progeny exposure and lung cancer mortality when Rn progeny exposure was modelled as a continuous variable and smoking was classified into multiple categories. Though medium and high smokers did tend to be at a higher risk than virtual non-smokers, the significance of the confounding effect of smoking stemmed from the group of individuals who were known to be smokers but for whom detailed smoking histories were not available. Many of these individuals included those who had died during the course of the study. Analyses including estimates of missing smoking consumption showed no advantage over available data. After controlling for its confounding effect, smoking did not appear to be a modifier of the effect of Rn progeny exposure on lung cancer mortality.

Evaluation based on quantitative estimates of smoking consumption showed a 6% increase in ERR for each increasing pack-year of smoking, given a fixed cumulative exposure; though statistically not highly significant (p = 0.13), this estimate of 6% increase in ERR/pack-year indicated a reasonably strong modifying effect. Other potential modifiers of the Rn exposure-time-smoking-lung cancer relationship - particularly other occupational exposures and various manifestations of age related temporal variables - could not be investigated further due to the limited number of risk sets available for this part of the study. Since there are important unanswered questions, extended follow-up and more data on smoking and other occupational histories should be sought.

CHAPTER 8

8. Global Perspectives: Summary, Interpretation and Implications of Findings from the Radium Hill Study

Tables and Figures

8. Global Perspectives: Summary, Interpretation and Implications of Findings from the Radium Hill Study

8.1 Summary of Findings from the Radium Hill Study

• Aim

The purpose of this study was to examine the risk of lung cancer mortality associated with exposure to Rn progeny amongst former workers at the RH uranium mine in South Australia.

• Study Design

Individuals who were on the pay-roll of the former RH uranium mine between 1948 and 1962 were followed-up over a forty year span - till the end of 1987 - through a retrospective cohort study. Study design characteristics were comprehensively detailed and discussed in chapter 3 which included critical appraisal of study trace rates. A total of 2,521 males and 53 females were included in the nominal roll of the study. The study trace rate was 74%; 24% of the traced male cohort are known to have died during the follow-up period.

• Demographics

Descriptive characteristics of the study cohort were detailed and discussed in chapter 4. The average male worker was entered the study at 31 years of age when commencing employment at RH, worked there for a total of 17 months, held 2.4 different jobs and was aged 32 years at termination of employment; he was followed-up for 19.5 years and aged 50 years at the end of follow-up. The average underground worker at RH was exposed to 7.7 WLM of Rn progeny, at a concentration of 0.83 WL, over a 12 month duration. Surface workers at RH were considered not to have been exposed. Lung cancer was the underlying cause of death of 54 males who comprised 9% of all deaths and 43% of the deaths from malignant neoplasms - the second leading cause of death which accounted for 22% of all deaths amongst males at RH.

- **Assessment of Analytical Methods Used**

- New statistical methodology for epidemiological cohort analysis based on modelling is, to some extent, a generalization of the traditional standardized mortality ratio (SMR) methods. It entails improved and formalized examination of the dependence of the relative risk on exposure level, time since exposure, age at risk, age at exposure, gender, and other relevant risk factors. It also provides a unified approach for testing the validity of models and for estimating the value of parameters.

- A particular strength of methods based on modelling techniques is that they permit the analysis of cohort data with purely internal comparisons, in addition to comparison with external population rates (which is the focus of traditional standardization methods). Methods based on internal comparisons have the advantage of avoiding potential biases due to differences between the cohort and the reference population, other than the exposure being studied (BEIR IV 1988).

- Methods based on standardization are commonly applied and remain important for cohort analyses; therefore, special efforts were made to compare the two methods. Modelling was then embarked upon to provide increasingly clear descriptions of the cohort experience.

- In using modelling techniques it was noted that lung cancer may be caused by several different agents that may act independently or through combined effects. The process of carcinogenesis is very complex, and any model to describe it will remain, at best, a rather rough approximation. Furthermore, even with fairly simple models that incorporate some of the factors determining lung cancer risk, the available data are not always detailed enough to provide clear answers to some of the most important questions, such as the effect of age at first exposure or time since exposure. Efforts have been made throughout this work to recognize these issues.

- In some parts of this work, it may appear that the assessment of confounding and effect modification have been overly associated with statistical tests of significance. I address this issue in the context of my own understanding and appraisal of some of the fundamental epidemiological concepts underlying this work, in Appendix C.

- All the analyses presented in this work are based on models that describe lung cancer risk relative to age- and calendar year-specific background rates. This does not amount to an *a priori* acceptance of what is commonly meant by a RR model in the radon literature (Fabrikant 1987), i.e., that the RR is constant with age, and perhaps with other factors such as gender, smoking habits and locality. It is believed that the RR model may be simpler and more applicable than models based on absolute excess risk (AER), because in other studies the AER has been found to increase substantially with age and/or time since exposure in a way that requires complex modelling. The RR, even if not constant in relation to changes in these factors, is often less dependent on them and therefore, can be modelled more simply. Moreover, statistical methods for analysis of the RR are more fully developed than those for analysis of the AER. However, when the RR and the AER are allowed to depend rather arbitrarily on other relevant factors, the two models are merely alternative expressions of the excess risk (BEIR IV 1988; BEIR V 1990).

- Analyses of other occupational exposures and smoking (chapter 7) were based on risk sets comprising only non-missing values for each of the variables examined (*i.e.*, records with missing values were excluded from analyses; where cases lacked complete data, the entire risk set was excluded). Evaluation based on these *selective* subsets of data yielded higher risk estimates than analyses based on complete data sets. Though such analyses would tend to carry inherent biases, there are few options available for analyses when the available data are limited. Even if the sub groups studied are at elevated risk, important questions of effect modification can still be addressed through the examination of variations in ERR. Therefore, in the absence of complete data, the procedure of nested case control analyses on risk sets excluding records with missing values offers a valuable exploratory tool.

• Confounders

Attained age and calendar time of observation - measured by calendar year - were found to be independent confounders of the exposure-response relationship. Their confounding effects were identified empirically through stratification and standardization techniques, and these techniques were also used to control their confounding effects. Methods based on both external and internal reference populations were evaluated in chapter 5. Once identified, these confounding effects were controlled for in all subsequent analyses. The Australian national population was found to be an unsuitable reference population and therefore all subsequent analyses were based on internal references.

• Exposure-Response Relationships

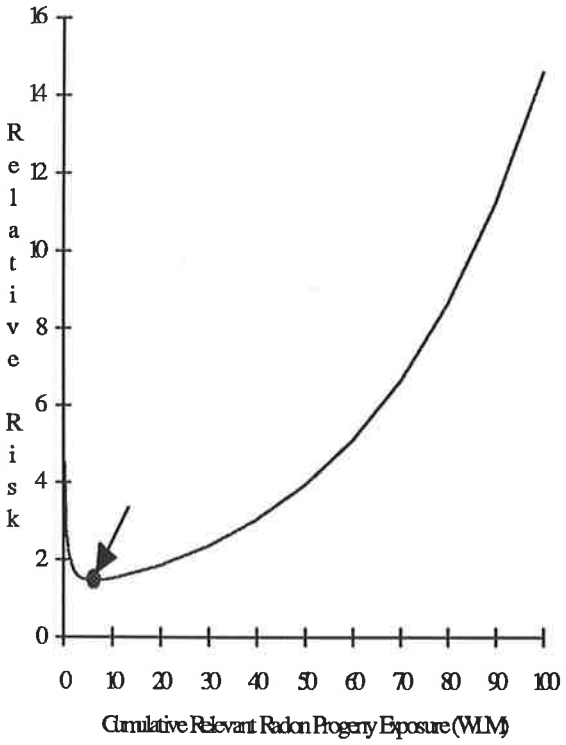
Findings from this study showed a significantly elevated risk (RR = 1.85) of lung cancer mortality amongst male RH workers compared to the Australian national population, after controlling for confounders. Though only underground workers were thought to have been exposed to Rn progeny at RH, underground workers when grouped together were not seen to be at a statistically significantly greater risk than surface workers at RH. Compared to surface workers at RH, workers who experienced 20-40 WLM of cumulative relevant Rn progeny exposure were at twice the risk of lung cancer mortality, and those exposed to over 40 WLM faced a greater than four-fold increase in RR. A significant trend in the Rn progeny exposure-response relationship was identified.

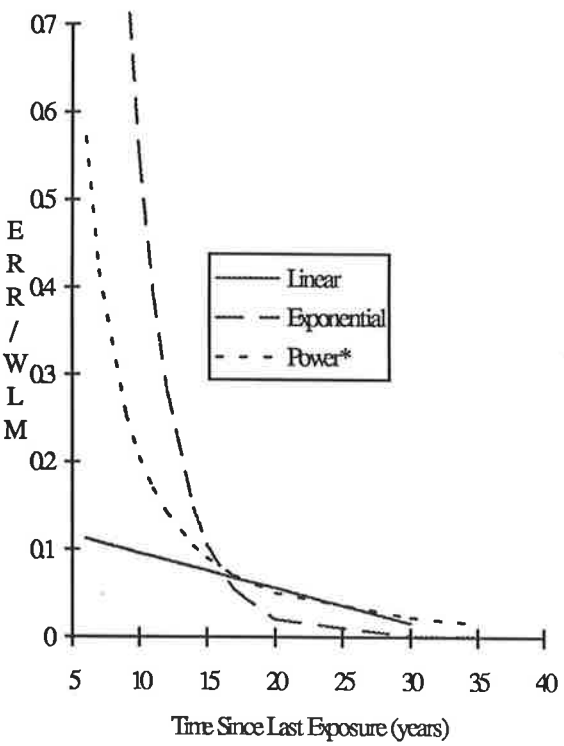
The linear model provided a reasonably good fit of the exposure-response data. Fitted independently, neither the exponential modification component nor the power function contributed any significant improvement to the linear fit. However, the combined expression of a general model which incorporated an exponential modification component and a power function provided a marginally significant improvement over the linear fit, an important contribution being in the improved risk estimates for lower levels of cumulative exposure (below the average level of exposure at RH).

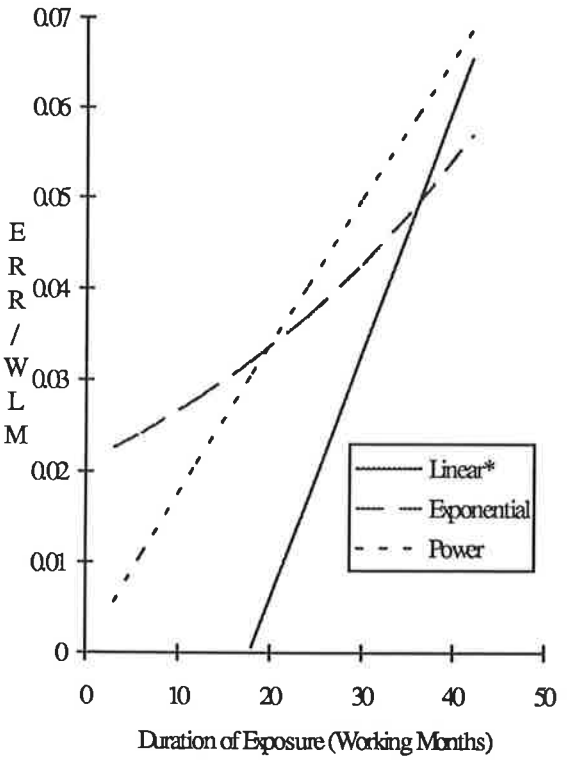
A doubling exposure of approximately 25 WLM was estimated from the best fitting model - the general model incorporating an exponential modification component and a power function (figure 8.1). Risk estimates obtained from the linear model showed an overall estimate of 4.3% ERR/WLM for all workers at RH (including those unexposed) and an estimate of 5.4% ERR/WLM amongst underground workers at RH.

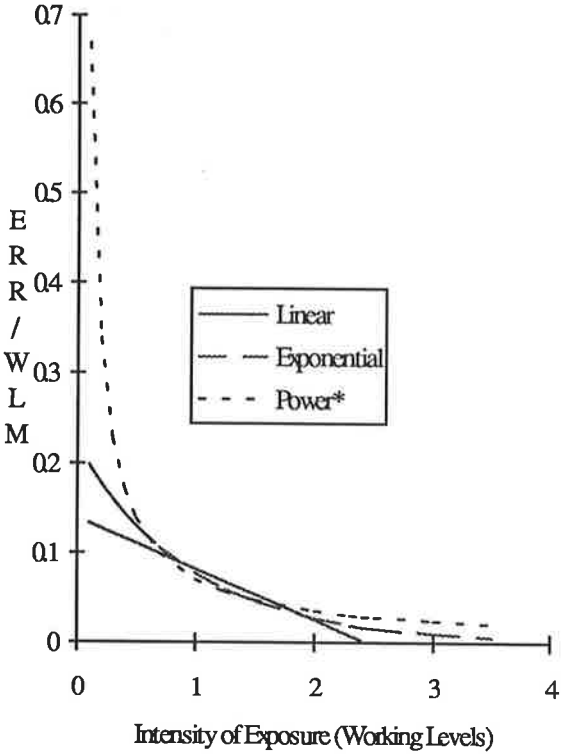
It must be noted that estimates of ERR/WLM provided in various sections of this thesis differ with the selection of each analytical data set. Estimates provided in chapter 5 were based on all workers and estimates in chapter 6 only on exposed workers; all analyses presented in both these chapters were based on person-years data and Poisson regression techniques. Estimates of risk derived in chapter 7 varied with the individual data sets; nested case-control analyses based on random sampling of each risk set and on complete risk sets provided estimates that were very similar, as expected. Each of the other analyses was based on specific subsections of the cohort e.g., those who responded to the questionnaire survey on specific questions viz., other occupational exposures and smoking. Estimates of Risk (ERR/WLM) derived from each of these sub cohorts are summarized at the end of this section.

Figure 8.1: Variations in Risk









Note: * Best Fitting Models

- **Exposure-Time-Response Relationships**

The temporal expression of Rn progeny exposure was identified as a strong modifier of the exposure-response relationship; the risk of lung cancer mortality decreased with increasing time since last exposure. The burden of lung cancer mortality risk was greatest with Rn progeny exposures experienced 5-15 years prior to the time of observation; more distant exposures had impacts that were 1-2 orders of magnitude smaller. However, it was these distant exposures that steered risk estimation towards the overall estimate of 2-4% ERR/WLM. Temporal effects of Rn progeny exposure were further examined in conjunction with other effect modifiers.

- **Exposure-Modifier-Response Relationships**

Time since last exposure, duration of exposure and intensity of exposure were identified as significant modifiers of the exposure-response relationship. For a given level of cumulative relevant exposure (hereafter, referred to as cumulative exposure or CE) to Rn progeny, the risk of lung cancer mortality decreased with increasing intensity of exposure, and increased with increasing duration of exposure. The Rn progeny related risk of lung cancer mortality also increased with increasing age at first exposure, but this modifying effect was not statistically significant. The functional expressions of risk associated with each of these modifiers are graphically depicted in figure 8.1, which also indicates the best fitting functions.

- **Exposure-Time-Modifier-Response Relationships**

The role of effect modifiers was further examined with exposure expressed in windows of time since exposure. Exposure was modelled under three windows of time since exposure viz., 5-15 years, 15-25 years and Over 25 years. Models examining the modifying effect of duration of exposure and intensity of exposure showed no significant effect modification with cumulative exposure when expressed under time since exposure windows. The reference model showed that risk declined steadily with increasing time since exposure. However, due to the sparseness of data in PYRS tables, these analyses provided rather unstable estimates.

Analyses of Rn progeny exposure under time since exposure windows were continued using nested case control analyses. The modifying effect of attained age on exposure expressed under time since exposure windows was examined using the BEIR IV TSE model (BEIR IV 1988). Analysis using a single window of time (corresponding to cumulative relevant exposure) showed that the risk of Rn progeny associated lung cancer mortality increased with attained age; however, this modifying effect was not statistically significant. When Rn progeny exposure was expressed in two windows of time (5-15 years and Over 15 years), the three level classification of attained age (below 55 years, 55-65 years and 65 years or more) had sparse data and could not yield any useful estimates. Therefore, attained age was modelled as a dichotomous variable thereafter (less than 65 years and 65 years or more); results of these analyses also indicated that the risk of Rn progeny associated lung cancer mortality increased with attained age; this pattern is contrary to the findings of the BEIR IV analyses (BEIR IV 1988), but was however, not statistically significant. The relative risk of lung cancer mortality associated with distant exposures (over 15 years) was about 20 times less than that associated with more recent exposure, which was about 10 times smaller than estimated in the BEIR IV analyses. Force-fitting the BEIR IV parameters to the RH data showed little difference in model deviances, a finding that was not unexpected, given the instability of parameter estimates in these models. Models such as the BEIR IV model can only be fitted effectively to the RH data after further follow-up when there are more data in distant windows of time since exposure and in higher categories of attained age.

- **Effect of Other Occupational Exposures**

Other occupational exposures relevant to this study comprised exposures to radioactive material other than at RH and exposures to asbestos. Each of these factors was examined through separate nested case control analyses comprising only workers with available data on the variables in question. In these subsets - risk sets with no missing values - estimates of the baseline ERR/WLM varied from those estimates derived for complete risk sets and randomized risk sets, as expected.

Occupational exposures to radioactive material other than at RH had a marginally confounding effect on the relationship between Rn progeny exposure at RH and lung cancer mortality. The baseline ERR/WLM was 9% before and 11% after controlling for the confounding effect of other occupational exposure to radioactive work; those who reported having experienced other occupational exposures to radioactive materials were at lower risk than those who reported no other such exposure.

Occupational exposure to asbestos was a strong confounder of the risk of lung cancer mortality associated with Rn progeny related lung cancer mortality. In this subset, those exposed to asbestos faced over a four-fold greater risk of lung cancer death than those who were not exposed to asbestos. The adjusted risk estimate derived from this study data set was 8% ERR/WLM.

Neither of these occupational exposures (experienced other than at RH) acted as effect modifiers, after controlling for their confounding or potentially confounding effects.

- **Role of Smoking**

Smoking was found to confound the relationship between Rn progeny exposure and lung cancer mortality. Close examination showed that the significance of the confounding effect arose mainly from the category of individuals who were known to be smokers but for whom complete smoking histories were not available. Some methods of estimating missing quantitative variables on smoking were explored; however, these attempts did not contribute to any improvement in the analyses.

The risk associated with the cigarette consumption was studied using 11 risk sets comparing cases and controls who had complete smoking histories. The effect of smoking alone did not adequately account for the risk of lung cancer mortality amongst the RH cohort; exposure to Rn progeny had a significant effect on lung cancer mortality after the confounding effect of smoking had been controlled. The ERR/WLM was estimated at 9% after controlling for the potentially confounding effect of smoking as a quantitative variable. Analyses based on quantitative estimates of smoking showed that its confounding effect was only of marginal statistical significance.

The modifying effect of smoking was then examined after controlling for its confounding effect. The modifying effect of smoking was best characterized by an exponential excess relative risk function (linear-exponential model); though only marginally significant, the ERR of lung cancer mortality for a given level of Rn progeny exposure, increased exponentially by 6% with each increasing pack-year of smoking. The modifying effect of smoking was thus found to be better characterized by a multiplicative model than an additive model. Risk estimates derived from an additive model showed an 8% increase in ERR with each increasing pack-year of smoking for a given cumulative level of Rn progeny exposure.

- **Surrogate Measures of Exposure**

The roles of duration of exposure and average Rn progeny concentrations (intensity of exposure) were examined as potential surrogate measures in both the person-years based analyses and the nested case control analyses. Results of these analyses were consistent with each other.

Duration of exposure provided a good surrogate measure. A comparison of goodness-of-fit characteristics between models showed that duration of exposure was a good predictor of lung cancer mortality; in fact, there was an indication that it may even be a slightly better predictor than CE - the model deviances with duration of exposure and cumulative exposure were 277.4 and 280.0 respectively from nested cases control analyses and 180.6 and 181.0 from person-years based analyses. Estimates of ERR per unit exposure derived from duration of exposure and CE were very similar - 2.7% ERR/WM and 2.9% ERR/WLM multiplicatively or 3.6% ERR/WM and 2.8% ERR/WLM additively. Intensity of exposure by itself, was not as good a predictor of lung cancer mortality as cumulative exposure or duration of exposure. The ERR of lung cancer mortality declined by 22% additively or 43% multiplicatively, with each increasing WL of exposure.

If these findings can be substantiated in other research, the implications are far reaching (see later; also Peto 1985). The equivalence of findings from duration of exposure and estimated cumulative exposure suggests that there is a large amount of measurement error overall in the estimates of Rn progeny concentrations.

- **Effect of Exposure Extrapolation Assumptions**

Estimates of exposures prior to 1953 and after 1961 were available in the original job exposure matrix constructed for this study, since there were no data available on radon gas concentrations in the mine during these years. These exposures were therefore estimated through direct extrapolation of Rn progeny concentrations from estimates closest in time i.e., from 1953 for previous years and from 1961 for the subsequent year. Such extrapolations were made in consultation with the Radiation Physicist who compiled the job exposure matrix. The validity of these assumptions were explored in this study with a simple form of sensitivity analyses.

Sensitivity analyses showed exposure extrapolation assumptions used may not have been totally appropriate. Exploratory analyses based on inductive methods were first used to examine various anecdotal suggestions that levels of exposure during the years in question may have been higher than those in the reference years for extrapolation. Maximum likelihood methods were then used to estimate *'best' multipliers* of reference exposures.

Maximum likelihood estimation showed that the best fit was obtained by extrapolations made on exposure multipliers greater than 0 but less than 1; i.e., those who worked underground at RH prior to 1953 and after 1961 were exposed to Rn progeny at concentrations that may have been less than those estimated for 1953 and 1961, respectively. Therefore, it was concluded that the ERR/WLM must lie between 1.76% (with simple extrapolation) and 3.08% (estimated assuming a zero multiplier *i.e.* no exposures prior to 1953 and following 1962). These estimates were made with allowance being made for the modifying effect of attained age - very early and very late workers at RH were thought to have been experienced older workers who worked on securing the mine shafts. These findings were supported by findings from the inductive exploratory analyses which preceded the maximum likelihood estimation of multipliers of exposure.

- ## Summary of Risk Estimates

Exposure-Response Relationships:

Cumulative Exposure (WLM): $RR = 1.029^{WLM}$

Duration of Exposure (WM): $RR = 1.029^{WM}$

Intensity of Exposure (WL): $RR = 0.57^{WL}$

Other Radioactive Material (*R*) $RR = (0.26)^{R} (1 + 0.11\ WLM)$

Asbestos (*A*) $RR = (4.83)^{A} (1 + 0.08\ WLM)$

#m Smoking (*P*) $RR = (0.95)^{P} (1 + 0.06 WLM * 1.06^{P})$

Exposure-Time-Response Relationships:

Exposure in TSE Windows: $RR = 1 + 0.334\ (W_{5\text{-}15} + 0.055\ W_{15\text{-}})$

Exposure-Modifier-Response Relationships:

#Age at First Exposure (AFE): $RR = 1 + 0.01\ WLM\ 1.04^{AFE}$

Time Since Last Exposure (TSLE): $RR = 1 + 22.44\ WLM\ (TSLE + 0.1)^{-2.03}$

Duration of Exposure (WM): $RR = 1 - 0.048\ WLM + 0.0027\ WLM * WM$

Intensity of Exposure (WL): $RR = 1.035^{WLM}\ 0.44^{WL}$

Exposure-Time-Modifier-Response Relationships:

#Attained Age (AA) and TSE : $RR = 1 + 0.289\ (1.884)^{AA}(W_{5\text{-}15} + 0.058\ W_{15\text{-}})$

Note: # Not Statistically Significant; m Marginally Significant Effect Modification;

8.2 Interpretations

• Measurement Error in Estimates of Radon Progeny Exposure

Results of sensitivity analyses on Rn progeny exposures estimated at RH showed some evidence of measurement error. The greater predictive power of the duration of exposure compared to cumulative Rn progeny exposure may be an indication of greater measurement error in estimates of cumulative exposure than in estimates of duration of exposure. Durations of exposure were estimated from pay-roll records that were maintained during the operation of the RH mine; therefore, these records of duration of exposure would have been reasonably accurate. In contrast, estimates of Rn progeny were obtained through reconstruction based on mine exposure models and estimates of Rn gas measurements - methods known to be fraught with uncertainties. The extent of the uncertainties in exposure measurement however, did not appear to be too great; risk estimates and the goodness-of-fit of models based on cumulative exposure and duration of exposure corresponded quite closely. Measurement error also possibly explains negative effect of intensity when modelled alone and the equally good fit of duration of exposure and cumulative exposure.

• Exposure-Response Relationships

The shape of the exposure-response function at the lower end of exposures could be a reflection the *short-stay worker effect* (Gilbert 1982, Doll and Peto 1985, Breslow and Day 1987). The opposite of the healthy worker effect, the short-stay worker effect implies that those who tended to work for short durations are at higher risk of lung cancer death than others, due to other aspects of lifestyle such as alcohol and tobacco consumption and *fast living*. Workers exposed to very low cumulative exposures do belong to this category of short-stay workers; also, risk estimates were unreliable at very low exposures; therefore, the patterns of increasing risk at very low levels of cumulative exposure should be interpreted with caution.

In interpreting the risk estimates from the RH study it must be realised that the RH mine was operated for a relatively short time compared to other mines overseas. For workers in mines overseas, mining could have been a life-time career; for workers at RH, this was seldom the case. Workers at the RH mine comprised individuals from a wide variety of backgrounds including many who were post Second World War migrants, who were brought to Australia from refugee camps in Europe under the obligation of working in areas they were allocated to; therefore, not all workers at RH were there by their own choice. Furthermore, few of these workers were in fact professional miners; former workers from RH include a diplomat, a high court judge, teachers, engineers, and many other professionals in diverse fields.

Though the finding of elevated risk at low cumulative levels and of the protracted exposure effect are relevant to the context of environmental exposures to Rn progeny, in drawing extrapolations it is necessary to remember that individuals who worked at RH had short durations of exposure. Therefore, exposures experienced by the RH cohort itself, are not strictly comparable to domestic exposures which generally occur over longer durations. These issues should be considered in any attempts at extrapolating findings from the RH study.

- **Radon Progeny Exposure as a Late-Stage Carcinogen**

Interpreted in terms of multistage models of carcinogenesis, the significant decrease in the risk of Rn progeny associated lung cancer mortality with increasing time since last exposure identified in this study, suggests that Rn progeny is a late-stage carcinogen. The role of age at first exposure could not be clearly identified in this study; however, the evident role of attained age as a modifier, where risk increases in those above the age of 65 lends further support to Rn progeny acting as a late-stage carcinogen. The lack of any significant modifying effect due to age at first exposure may be due to the limited variation observed in the distribution of age at first exposure, compared to other studies. Therefore, the RH cohort cannot effectively be used to address the modifying effect of age at first exposure.

• The Protracted Exposure Effect

The significant decrease identified in the risk of Rn related lung cancer mortality with increasing intensity of exposure (for a given cumulative exposure) could be interpreted as a *protracted exposure effect*. The protracted exposure effect suggests in essence that long exposures at low levels cumulative levels are more hazardous than short-lasting high exposures to Rn progeny. This finding was supported by the significant modifying effect of duration of exposure. The greater significance of the modifying effect of duration of exposure compared to that of intensity of exposure may reflect the fact that duration of exposure was subject to less measurement error than intensity of exposure. The greater significance of intensity of exposure as a modifier than as a predictor does imply that its effect is real rather than an artifact owing to measurement error.

This study has identified a significant protracted exposure effect amongst the RH cohort who were largely exposed to below 100 WLM of cumulative Rn progeny exposure; no other study reported to date has identified such an effect in this range of exposure.

In view of the protracted exposure effect, question arises as to the appropriateness of exposure limits (traditionally, dose limits) based on cumulative exposures. Recommendations on occupational exposure limits made in terms of average annual cumulative exposures, have been derived from models of minimum allowable risk for life-time occupational exposure. In subscribing to a protracted exposure effect, it would be appropriate to consider the joint implications of cumulative Rn progeny exposure and the concentrations at which these exposures were experienced. This means that a worker exposed to 1 WLM for a working life time of 50 years would have a relative risk of 4.18 of dying from lung cancer at that time using the model fitting cumulative exposure alone. However, when allowing for the protracted exposure effect this relative risk increases to 5.16. These calculations were based on models of best fit viz., $RR = 1.029^{WLM}$; $RR = 1.035^{WLM}\ 0.44^{WL}$; 1 WLM experienced over 1 year = 0.0971 WL - based on the conversion factor (170/1750) used in this study and in the joint analyses now being undertaken by the National Cancer Institute.

- **Smoking**

The quality of the data available on smoking made it rather difficult to examine the effects of smoking. Based on available data, smoking may have confounded the association between Rn progeny exposure and lung cancer mortality, but does not fully explain the observed association with Rn progeny. After adjusting for this potentially confounding effect of smoking, for a given level of Rn progeny exposure, the ERR increased exponentially by 6% with each increasing pack year of smoking. The modifying effect of smoking was only of marginal statistical significance. However, given the limited amount of data from which this finding was obtained, it was considered an indication of the importance of continued research based on improved data.

The finding that the modifying effect of smoking on the Rn progeny associated lung cancer mortality was more multiplicative than additive corresponded with findings from other studies (Whittemore and McMillan 1983; Hornung and Meinhardt 1987; BEIR IV 1988; Samet 1992; Sevc *et al.* 1993). However, the limited amount of data available on smoking provided very little power for further interpretation or more detailed examination.

• Other Relevant Occupational Exposures

Asbestos: Few studies have reported on other relevant occupational exposures that may have acted as confounders of the risk of Rn progeny related lung cancer mortality. No other studies on occupational radon epidemiology reported to date have reported the effects of asbestos exposure on the association between Rn progeny exposure and lung cancer mortality.

Though the amount of data available on these factors in the RH study was limited, occupational exposure to asbestos was found to be a statistically significant confounder of the exposure-response relationship. These preliminary analyses show an effect that should be explored further with more complete data. It is therefore important that occupational exposure to asbestos should not be ignored in future examination of the RH cohort and any other studies. The reverse is also of course true in that studies of asbestos-related risk of lung cancer should also consider added exposure to Rn progeny. Such a study is being planned with the Wittenoom Gorge study; joint analysis will focus on 99 workers who are on the nominal roll of both the Wittenoom Gorge and the RH studies.

Other Occupational Exposure to Radioactive Material: The confounding effect of other occupational exposure to radioactive material (other than at RH) was of marginal significance. Those who reported having worked with radioactive material other than at RH appeared to be at lower risk than those who had not. Possible explanations for this seemingly anomalous *protective effect* could include lack of data (there were no lung cancer cases amongst those unexposed to Rn progeny at RH, but exposed elsewhere to radioactive material), a chance effect, response bias (the reluctance to report other exposures in anticipation of future compensation payments to RH workers), the healthy worker effect (i.e., these workers may have been fitter than others). The apparent confounding may be an artifact of some other hidden or unidentified factor. Here again, it is important that future investigations should target specific questions on other relevant occupational exposure to radioactive material including some clear indications as to the nature of the exposures and their severity.

8.3 Implications

A summary of findings from the RH study corresponding to those summarized from other studies reported in chapter 2 is presented in table 8.3.1. Findings summarized in table 8.3.1 show that the RH cohort had the lowest average exposures (cumulative relevant exposures) of all studies. The number of lung cancer deaths in the RH cohort was lower than in some of the other studies, particularly in relation to the size of each individual cohort. The duration of follow-up in the RH study was longer than in most other studies. With the exception of the larger studies - Ontario, China, Beaverlodge and Czech - the size of the RH cohort was comparable with that of other studies. Since risk estimates in several of the studies reported were based only on workers who had worked underground for a specific minimum period, two separate sets of risk estimates were reported in table 8.3.1 for the RH study - viz., estimates based on the entire cohort and those based only on those exposed.

Risk estimates derived from the RH study are higher than those reported from any of the other studies reported. These - the highest reported estimates of risk - are associated with the lowest reported levels of cumulative exposure to Rn progeny. Such findings appear to be paralleled by preliminary investigations of the Wismut mining population in Germany. These findings imply that projections made from risk estimates obtained in studies with higher cumulative exposures, to populations exposed to low cumulative levels of Rn progeny may underestimate the risk to these populations. The need for studying *statistically respectable* (see chapter 2, page 67) populations of individuals exposed to low levels of Rn progeny exposure has long been emphasized. Attaining *statistical respectability* is a long and arduous research process, beyond the point of identifying suitable populations for such study. This study has confirmed that the RH cohort can make a valuable contribution to the study Rn progeny related lung cancer mortality, independently and jointly with other studies - for which the available data is suitable. Despite the limited amount of data currently available, the RH study has contributed valuable insight into the risk of Rn progeny associated lung cancer mortality; further statistical respectability is attainable, the quest for which must continue.

Table 8.3.1: Summary of Findings from All Reported Studies

Study	Follow up Period	Cohort Size	Cases Ca. Lung	PYRS	Mean Exposure (WLM)	Doubling Dose (WLM)	SMR Ca. Lung Mortality	ERR/WLM Ca. Lung Mortality (%)	Comments
Canada:									
Beaverlodge (Howe *et al.* 1986; SENES 1991)	1948-80	8,487	54[e] 65	56,942[e] [114,170] (BEIR IV 1988)	20.2 [44][s]	31[c]	1.90 (1.43-2.49)	3.28 (2.08-4.48) [1.3][s] (0.59-2.98)[s]	D-R: Linear Lag:10 yrs. Smk: Not Rep. [s]SENES 1991.
Port Radium (Howe *et al.* 1987)	1942-80	2,103	48[e] 57	34,673[e]	183.3	370[c]	2.10 (1.55-2.79)	0.27 (0.11-1.43)	D-R: Linear Lag: 10 yrs. Smk: Not Rep.
Newfoundland (Morrison *et al.* 1988)	1933-84	2,124	110[e] 113	38,509[e]	382.4	112	5.25[u] (4.33-6.32)	0.9 (0.6-1.2)	D-R: Linear Lag[e]:10 yrs. Smk: Risk Rep.
Ontario (Kusiak *et al.* 1991)	1955-86	54,128	378	1,706,103	[40-90[r]] (Muller *et al.* 1985)	83[c]	1.29 (1.15-1.45)	1.2 (0.02-2.4)	Lag: 5 yrs. Smk: Assessed. Mean Exp. NR.
United States:									
Colorado (Hornung and Meinhardt 1987)	1950-82	3,346	256	[73,642] (BEIR IV 1988)	834.0	71-111[cr]	[4.33][c] (BEIR IV 1988)	0.9-1.4[r]	Lag: Estimated. Smk: Risk Rep.
New Mexico (Samet *et al.* 1991)	1957-85	3,469	65	NR	111	55.6[c]	4.0 (3.1-5.1)	1.8 (0.7-5.4)	Smk: Synergystic with Rn Exp.
Czech.: Study S (Sevc *et al.* 1993)	1948-80	4,042	574	97,913	227.0	95	4.7 NR	0.6 NR	Lag: 5 years. D-R: Supra-linear, linear
Sweden: (Radford and Renard 1984)	1951-76	1,294	50	24,083	93.7	28[c]	3.9 (3.0-4.9)	3.6 (2.5-4.8)	Lag: 5 years. CI: 90%. Smk: Risk Rep.
France: (Tirmarche *et al.* 1993)	1946-85	1,785	45	44,005	70	58.8[c]	1.91 NR	0.59 (0-1.6)	Interim Report; Analyses Cont. Smk: No Data.
UK: (Hodgson and Jones 1990)	1941-86	3,082 3,010[t]	105	NR	NR	NR	1.58 NR	9.4 NR	Lag: Not Used; Est. 10 yrs. Smk: No Data.
China: (Xuan *et al.* 1993)	1976-87	17,143	981	175,405	275.4	161[c] 625[ca]	NR NR	0.62 (0.5-0.8) 0.16[a] (0.1-0.2)[a]	Lag: 5 yrs. Smk: Risk Rep. [a] Adj.: Arsenic
Australia: Radium Hill (This Study)	**1948-87**	**2,521**	**54**	**49,241**	**7.7**	**23.26** **18.52**	**1.85** **2.09**	**4.3** **(0.7-11.1)** **5.4** **(1.0-17.0)**	**All Workers** **D-R: Linear** **U/G Workers** **D-R: Linear**

Note: NR Not Reported in main reference.
e Excluding first 10 years of followup.
t Traced Cohort: only these analysed.
r Range covers uncertainity in estimates.
D-R Dose-Response Relationship.
u Unlagged data.
c Computed from published findings.
Smk. Smoking; Rep.: Reported.

BIBLIOGRAPHY

Abramson J.H. 1988.
Making Sense of Data.
A Self Instruction Manual on the Interpretation of Epidemiological Data.
Oxford University Press.

Adams G.E., Cox R. and Humphreys E.R. 1987.
Biological Effects of Ionising Radiation: Dose and Dose-Rate Relationships.
In: Health Effects of Low-Dose Ionising Radiation - Recent Advances and their Implications.
Proceedings of the International Conference Organised by the British Nuclear Energy Society
and Co-Sponsored by the European Nuclear Society.
British Nuclear Energy Society, London: pp. 19-23.

Agricola G. 1597.
De Re Metallica, Basel. Cited by Hueper W. 1942.
Occupational Tumors and Allied Diseases. Charles C. Thomas, Springfield, Illenois.
As translated by: Hoover, H.C. and Hoover, L.H. 1950. Dover Publications, New York.

Aitken M., Anderson D., Francis B. and Hinde J. 1989.
Statistical Modelling in GLIM.
Clarendon Press, Oxford.

Anthony R.D. 1988.
Opening Address.
In: Health Effects of Low-Dose Ionising Radiation - Recent Advances and their Implications.
Proceedings of the International Conference Organised by the British Nuclear Energy Society
and Co-Sponsored by the European Nuclear Society.
British Nuclear Energy Society, London: p. 1.

Archer V.E., Wagoner J.K. and Lundin F.E. 1973.
Lung Cancer Among Uranium Miners in the United States.
Health Physics 25: pp. 351-71.

Archer V.E., Gillam J.D. and Wagoner J.K. 1976.
Respiratory Disease Mortality Among Uranium Miners.
Annals of the New York Academy of Sciences 271: pp. 280-93.

Armitage P. and Doll R. 1961.
Stochastic Models for Carcinogenesis.
In: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability: Biology and Problems of Health.
Berkeley, University of California Press: pp. 19-38.

Armstrong B.K., De Klerk N.H., Musk A.W. and Hobbs M.S.T. 1988.
Mortality in Miners and Millers of Crocidolite in Western Australia.
British Journal of Industrial Medicine 45: pp. 5-13.

Barendson G.W. 1968.
Responses of Cultured Cells, Tumors, and Normal Tissues to Radiation of Different Linear Energy Transfer.
Current Topics in Radiation Research 4: 293-356.

Beckam R. and Holub R. 1979.
Radon Daughter Growth With Continuous Radon Influx.
United States Department of Labour, IR 1099.

BEIR III 1980.

The Effects on Populations of Exposure to Low Levels of Ionizing Radiation.

National Research Council, Committee on the Biological Effects of Ionizing Radiations;

Board on Radiation Effects Research; Commission on Life Sciences;

National Academy Press.

BEIR IV 1988.

Health Risks of Radon and Other Internally Deposited Alpha-Emitters BEIR IV

Committee on the Biological Effects of Ionizing Radiations;

Board on Radiation Effects Research; Commission on Life Sciences;

National Research Council. National Academy Press.

BEIR V 1990.

Health Effects of Exposure to Low Levels of Ionizing Radiation BEIR V

Committee on the Biological Effects of Ionizing Radiations;

Board on Radiation Effects Research; Commission on Life Sciences;

National Research Council. National Academy Press.

Berkson J. 1955.

The Statistical Study of Association Between Smoking and Lung Cancer.

In: Proceedings of a Staff Meeting, Mayo Clinic 30: pp. 319-48.

Blot N.J. and Day N.E. 1979.

Synergism and Interaction: Are They Equivalent.

American Journal of Epidemiology 110: pp. 99-100.

Borzsonyi M., Day N.E., Lapis K. and Yamasaki H. (Editors). 1984.
Models, Mechanisms and Etiology of Tumour Promotion.
IARC Scientific Publications No. 56.
World Health Organization, International Agency for Research on Cancer, Lyon, France.

Breslow N.E. 1984.
Elementary Methods of Cohort Analysis
International Journal of Epidemiology 13-1: 112-5.

Breslow N.E. and Day N.E. 1980.
Statistical Methods in Cancer Research.
Volume 1 - The Analysis of Case-Control Studies.
IARC Scientific Publications No. 32.
World Health Organization, International Agency for Research on Cancer, Lyon, France.

Breslow N.E. and Day N.E. 1987.
Statistical Methods in Cancer Research.
Volume II - The Design and Analysis of Cohort Studies.
IARC Scientific Publications No. 82.
World Health Organization, International Agency for Research on Cancer, Lyon, France.

Breslow N.E. and Langholz B. 1987.
Nonparametric Estimation of Relative Mortality Functions.
Journal of Chronic Diseases 40, Supplement 2: pp. 89S-99S.

Breslow N.E., Lubin J.H., Marek P. and Langholz B. 1983.
Multiplicative Models and Cohort Analysis.
Journal of the American Statistical Association 78-381: pp. 1-12.

Broerse J.J. and Barendson G.W. 1973.
Relative Biological Effectiveness of Fast Neutrons for Effects on Normal Tissue.
Current Topics in Radiation Research 8: pp. 305-50.

Brown C.C. and Chu K.C. 1983.
Implications of the Multi-Stage Theory of Carcinogenesis Applied to Occupational Arsenic Exposure.
Journal of the National Cancer Institute 70: pp. 455-63.

Cameron J.R. and Skofronick J.G. 1978.
Medical Physics.
John Wiley and Sons.

Chambers D.B., Stager R.H., Benard F.D. and Low L.M. 1992.
Reconstruction of the Radon Daughter Exposure of Beaverlodge Uranium Miners.
In: Proceedings of the Eighth International Congress of IRPA, Montreal: pp. 1586-89.

Charlesworth M. 1982.
Science, Non-Science and Pseudo-Science
Bacon, Popper, Lakatos Kuhn and Feyerabend on Defining Science.
Deakin University Press.

Checkoway H., Pearce N.E. and Crawford-Brown D.J. 1989.
Research Methods in Occupational Epidemiology.
Monographs in Epidemiology and Biostatistics Volume 13.
Oxford University Press.

Chovil A. 1981.
The Epidemiology of Primary Lung Cancer in Uranium Miners in Ontario.
Journal of Occupational Medicine 23: pp. 417-21.

Chu K.C. 1987.
A Non Mathematical View of Mathematical Models for Cancer.
Journal of Chronic Diseases 40, Supplement 2: pp. 163S-70S.

Clayton D. and Hills M. Draft 1992.
Statistical Models in Epidemiology.
MRC Unit of Biostatistics, Cambridge, U.K.

Coggle J. E. 1983.
Biological Effects of Radiation.
Taylor and Francis Ltd. London.

Cook P.J., Doll R. and Fellingham S.A. 1969.
A Mathematical Model for the Age Distribution of Cancer in Man.
International Journal of Cancer 3: pp. 93-112.

Cox D.R. 1972 .
Regression Models and Life-Tables (With Discussion).
Journal of the Royal Statistical Society B 38: pp. 187-220.

Criqui M.H., Barrett-Conner E. and Austin M. 1978.
Differences Between Respondents and Non-Respondents in a Population-Based
Cardiovascular Disease Study.
American Journal of Epidemiology 108: pp. 367-72.

Crouch P. and Corani C. 1986.
Radon Daughter Exposures During Mining at Radium Hill 1953-1961.
An Internal Working Report.
Radiation Control Section, Occupational Health and Radiation Control Branch,
Environmental and Occupational Health Division, South Australian Health Commission.

Crouch D. 1987.
The Role of Predictive Modelling: Social and Scientific Problems of Radiation Risk Assessment.
In: Radiation and Health - The Biological Effects of Low-Level Exposure to Ionizing Radiation.
Jones R.R. and Southwood R. (Editors).
John Wiley and Sons Limited: pp. 47-63.

Day N.E. 1984.
Epidemiological Data and Multistage Carcinogenesis.
In: Models, Mechanisms and Etiology of Tumour Promotion,
IARC Scientific Publications No. 56, Lyon: pp.339-57.

Day N.E. and Brown C.C. 1980.
Multistage Models and the Primary Prevention of Cancer.
Journal of the National Cancer Institute 64: pp. 977-89.

De Klerk N.H., Armstrong B.K., Musk A.W. and Hobbs M.S.T. 1989.
Cancer Mortality in Relation ot Measures of Occupational Exposure to Crocidolite at Wittenoom Gorge in Western Australia.
British Journal of Industrial Medicine 46: pp. 529-36.

Dobson A.J. 1983.
Introduction to Statistical Modelling.
Chapman and Hall.

Doll R. and Peto R. 1981.
The Causes of Cancer.
Quantitative Estimates of Avoidable Risks of Cancer in the United States Today.
Journal of the National Cancer Institute 66: pp. 1197-308.
Also: Oxford University Press.

Doll R. and Peto R. 1978.
Cigarette Smoking and Bronchial Carcinoma: Dose and Time Relationships Among Regular Smokers and Lifelong Non-Smokers.
Journal of Epidemiology and Community Health 32: pp. 303-13.

Doll R. and Peto J. 1985.
Asbestos: Effects on Health of Exposure to Asbestos.
Health and Safety Commission
London: Her Majesty's Stationery Office.

Enterline P.E. and Henderson V. 1973.
Type of Asbestos and Respiratory Cancer in the Asbestos Industry.
Archives of Environmental Health 27: pp. 312-17.

EPICURE 1992.
Preston D.L., Lubin J.H. and Pierce D.A.
EPICURE: Risk Regression and Data Analysis Software - Users Guide.
HiroSoft International Corporation, Seattle, Washington, U.S.A.

Fabrikant J.I. 1987.

The BEIR IV Report.

In: Proceedings of the ASA Conference on Radiation and Health. Coolfont VII.

American Statistical Association. Alexandria, Virginia: pp. 3-7.

Franks L.M. and Teich N. 1986.

Introduction to the Cellular and Molecular Biology of Cancer.

Oxford Science Publications, Oxford University Press.

Fleiss J.L. 1986.

Significance Tests Have a Roll in Epidemiologic Research: Reaction to A.M. Walker.

American Journal of Public Health 76: pp. 559-60.

Fox A.J., Goldblatt P. and Kinlen L.J. 1981.

A Study of the Mortality of Cornish Tin Miners.

British Journal of Industrial Medicine 38: pp. 7378-80.

Freund J.E. 1971.

Mathematical Statistics.

Prentice Hall International, Inc., London.

Frome E.L. 1983.

The Analysis of Rates Using Poisson Regression Models.

Biometrics 39: pp. 665-74.

Frome E.L. and Checkoway H. 1985.

Epidemiologic Programs for Computers and Calculators

Use of Poisson Regression Models in Estimating Incidence Rates and Ratios.

American Journal of Epidemiology 121-2: pp. 309-23.

Frost S.E. 1983.
Port Radium Working Level Month Calculations.
Ottawa: Eldorado Resources Limited: Draft No. 1.

Gilbert E.S. 1982.
Some Confounding Factors in the Study of Mortality and Occupational Exposures.
American Journal of Epidemiology 116-1: pp.177-188.

Grace M., Larson M. and Hanson J. 1980.
Bronchogenic Carcinoma among Former Uranium Mine Workers at Port Radium, Canada - A Pilot Study.
Health Physics 38: pp. 657-61.

Gray L.H. 1965.
Radiation Biology and Cancer.
In: Cellular Radiation Biology: pp. 18-9.
Baltimore: Williams and Wilkins.

Greenberg M. and Selikoff I.J. 1993.
Lung Cancer in the Schneeberg Mines: a Reappraisal of the Data Reported by Harting and Hesse in 1879.
Annals of Occupational Hygiene, 38-1: pp. 5-14.

Greenland S. 1977.
Response and Follow-up Bias in Cohort Studies.
American Journal of Epidemiology 106: pp. 184-7.

Guess H.A. and Hoel D.G. 1977.

The Effect of Dose on Cancer Latency Period.

Journal of Environmental Pathology and Toxicology, 1: pp. 279-86.

Haldane J.B.S. 1955.

The Estimation and Significance of the Logarithm of a Radio of Frequencies.

Annals of Human Genetics 20: pp. 309-11.

Ham J.M. 1976.

Report of the Royal Commission on the Health and Safety of Workers in Mines.

Toronto, Ministry of the Attorney General.

Harrington M. and Kakoschke K. 1991.

We Were Radium Hill.

Stories and Memories of People Who Once Lived at Radium Hill.

Malcolm Harrington, South Australia.

Harting F.H. and Hesse W. 1879.

Lung Cancer, the Disease of Miners in the Schneeberg Mines.

Vrtljhrssch f Genrichtl Med 30: pp. 296-309; 31: p. 102-29; and 31: p. 313-37.

NIH Library Translations: NIH-77-262C; NIH-77-247C; and NIH-77-246C.

Cited by Hueper 1942; Greenberg and Selikoff 1993; Schulttman 1993.

In: Occupational Tumors and Allied Diseases. Charles C. Thomas, Springfield, Illenois.

Hennekens C.H. and Buring J.E. 1987.

Epidemiology in Medicine.

Little, Brown and Company. Boston / Toronto.

Hennings H., Shores R., Wenk M.L., Spangler E.F., Tarone R. and Yuspa S.H. 1983.
Malignant Conversions of Mouse Skin Tumors is Increased by Tumor Initiators and Unaffected by Tumor Promoters.
Nature (London) 304: pp. 67-9.

Hewitt D. 1976.
Radiogenic Lung Cancer in Ontario Uranium Miners, 1955-74.
In: Report of the Royal Commission on the Health and Safety of Workers in Mines.
Ham J.M. (Editor).
Toronto, Ministry of the Attorney General. Appendix C: pp. 319-29.

Hodgson J.T. and Jones R.D. 1990.
Mortality of a Cohort of Tin Miners 1941-86.
British Journal of Industrial Medicine 47: pp. 665-76.

Hornung R.W. and Samuels S. 1981.
Survivorship Models for Lung Cancer Mortality in Uranium Miners.
Is Cumulative Dose an Appropriate Measure of Exposure?
In: Radiation Hazards in Mining: Control, Measurement, and Medical Aspects.
Gomez M. (Editor).
New York, Society of Mining Engineers of American Institute of Mining, Metallurgical, and Petroleum Engineers pp. 363-8.

Hornung R.W. and Meinhardt T.J. 1987.
Quantitative Risk Assessment of Lung Cancer in U.S. Uranium Miners.
Health Physics 52-4: pp. 417-30.

Howe G.R., Nair R.C., Newcombe H.B., Miller A.B. and Abbatt J.D. 1986.
Lung Cancer Mortality (1950-80) in Relation to Radon Daughter Exposure in a Cohort of Workers at the Eldorado Beaverlodge Uranium Mine.
Journal of the National Cancer Institute 77-2: pp. 357-62

Howe G.R., Nair R.C., Newcombe H.B., Miller A.B., Burch J.D. and Abbatt J.D. 1987.
Lung Cancer Mortality (1950-80) in Relation to Radon Daughter Exposure in a Cohort of Workers at the Eldorado Port Radium Uranium Mine: Possible Modification of Risk by Exposure Rate.
Journal of the National Cancer Institute 79-6: pp. 1255-60.

IAEA 1986.
Facts about Low-Level Radiation.
International Atomic Energy Agency.

IARC 1988.
IARC Monographs on the Evaluation of Carcinogenic Risks to Humans
Man-made Mineral Fibres and Radon. Volume 43.
IARC Working Group on the Evaluation of Carcinogenic Risks to Humans.
World Health Organization, International Agency for Research on Cancer, Lyon, France.

IARC 1990.
Cancer: Causes, Occurrence and Control.
IARC Scientific Publications: No. 100.
World Health Organization, International Agency for Research on Cancer, Lyon, France.

ICD 9 1978.

Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death.

9th Revision. World Health Organization, Geneva.

Radon Daughter Growth With Continuous Radon Influx.

United States Department of Labour: IR 1099.

ICRP 32 1981.

Limits for Inhalation of Radon Daughters by Workers: ICRP Publication 32.

International Commission on Radiological Protection.

Oxford Pergamon Press.

ICRP 50 1987.

Lung Cancer Risk from Indoor Exposures to Radon Daughters: ICRP Publication 50.

International Commission on Radiological Protection.

Oxford Pergamon Press.

ICRP 60 1991.

1990 Recommendations of the International Commission on Radiological Protection.

Pergamon Press.

ICRP Draft 1993.

Protection Against Radon-222 At Home and At Work.

Draft for Consultation.

International Commission on Radiological Protection.

Jacobi W., Henrichs K. and Barclay D. 1992.
Verursachungs-Wahrscheinlichkeit von Lungenkrebs durch die berufliche Strahlenexposition von Uran-Bergarbeitern der SDAG WISMUT.
GSF Research Centre for Environment and Health.
GSF-Report S-13/92. Munich/Neuherberg. Federal Republic of Germany.

Jacobsen M. 1976.
Against Popperized Epidemiology.
International Journal of Epidemiology 5: 9-11.

Kaldor J.M. and Day N.E. 1989.
Interpretation of Epidemiologcial Studies in the Context of the Multistage Model of Carcinogenesis.
In: Mechanisms of Environmental Carcinogenesis:
Volume II, Multistep Models of Carcinogenesis. Barrett, J.C. (Editor).
Boca Raton, Florida, CRC Press: pp. 21-57.

Kellerer A.M. and Rossi H.H. 1972.
The Theory of Dual Radiation Action.
Current Topics in Radiation Research 8: pp. 85-158.

Kelsey J.L., Thompson W.D. and Evans A.S. 1986.
Methods in Observational Epidemiology.
Monographs inEpidemiology and Biostatistics Volume 10.
Oxford University press.

Kleinbaum D.G., Morgenstern H. and Kupper L.L. 1981.
Selection in Epidemiological Studies.
American Journal of Epidemiology 113: pp. 452-63.

Kunz E., Sevc J., Placek V. and Horacek J. 1979.
Lung Cancer in Man in Relation to Different Time Distribution of Radiation Exposure.
Health Physics 35: pp. 579-80.

Kusiak R.A., Springer J., Ritchie A.C. and Muller J. 1991.
Carcinoma of the Lung in Ontario Gold Miners: possible Aetiological Factors.
British Journal of Industrial Medicine 48: pp. 808-17.

Last J.M. 1988.
A Dictionary of Epidemiology. Second Edition (Editor).
Oxford University Press.

Lewis P.D. 1987.
Variation in Individual Sensitivity to Ionizing Radiation.
In: Radiation and Health: The Biological Effects of Low-Level Exposure to Ionizing Radiation.
Jones R.R. and Southwood R. (Editors).
John Wiley and Sons Limited. pp. 167-77

Ludewig P. and Lorenser E. 1924.
Untersuchungen der Grubenluft in den Schneeberger Gruben auf den Gehalt an Radiumemanation.
Ein Beitrag zur Frage nach der Enststehng des Schneeberger Lungenkrebses.
Strahlentherapie 17: pp. 428-435.
Z. Physik 22: p. 178.

Lundin F.E.Jr., Lloyd J.W., Smith E.M., Archer V.E. and Holaday D.A. 1969.
Mortality of Uranium Miners in Relation to Radiation Exposure, Hard-Rock Mining and Cigarette Smoking - 1950 Through September 1957.
Health Physics 16: pp. 571-8.

Lundin F.E.Jr., Wagoner J.K. and Archer V.E. 1971.
Radon Daughter Exposure and Respiratory Cancer: Quantitative and Temporal Aspects.
National Institute of Occupational Safety and Health, National Institute of Environmental Health Sciences: Joint Monograph No. 1.
United States Department of Health, Education and Welfare, Public Health Service.

Lundin F.E.Jr., Archer V.E. and Wagoner J.K. 1979.
An Exposure-Time-Response Model for Lung Cancer Mortality in Uranium Miners: Effects of Radiation Exposure, Age and Cigarette Smoking.
In: Energy and Health. Proceedings of a SIMS Conference.
Breslow N.E. and Whittemore A.S. (Editors).
Philadelphia, PA: Society for Industrial and Applied Mathematics pp. 243-64.

MacMahon B., Pugh T.F. and Ipsen J. 1960.
Epidemiologic Methods.
Boston: Littli, Brown and Company.

MacMahon B. and Pugh T.F. 1970.
Epidemiology: Principles and Methods.
Little, Brown and Company, Boston.

Mantel N. and Haenszel W. 1959.
Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease.
Journal of the National Cancer Institute 22: 719-48.

McCullagh P. and Nelder J.A. 1989.
Generalized Linear Models.
Chapman and Hall.

Mettler F.A.Jr. and Moseley R.D.Jr. 1985.
Medical Effects of Ionizing Radiation.
Grune and Stratton Inc.

Miettenen O.S. and Cook E.F. 1981.
Confounding: Essence and Detection.
American Journal of Epidemiology 114: 593-603.

Mood A.M., Graybill F.A. and Boes D.C. 1974.
Introduction to the Theory of Statistics.
McGraw-Hill Kogakusha, Limited.

Moolgavkar S.H. 1978.
The Multistage Theory of Carcinogenesis and the Age Distribution of Cancer in Man.
Journal of the National Cancer Institute 61: pp. 49-52.

Moolgavkar S.H and Venzon D.J. 1979.
Two-Event Models for Carcinogenesis: Incidence Curves for Childhood and Adult Tumors.
Mathematics in the Biosciences 47: pp. 55-77.

Moolgavkar S.H and Knudson A.G.Jr. 1981.
Mutation and Cancer: A Model for Human Carcinogenesis.
Journal of the National Cancer Institute 66: pp. 1037-52.

Morrison H.I., Semenciw R.M., Mao Y. and Wigle D.T. 1988.
Cancer Mortality Among A Group of Flourspar Miners Exposed to Radon Progeny.
American Journal of Epidemiology 128-6: pp. 1266-75.

Muller J. and Wheeler W.C. 1974.
Causes of Deaths in Ontario Uranium Miners.
Radiation Protection in Mining and Milling of Uranium and Thorium.
International Labour Office, Geneva: pp. 29-42.

Muller J., Wheleler W.C., Gentleman J.F., Suranyi G., Kusiak R. and Smith M. 1981.
The Ontario Miners Mortality Study, General Outline and Progress Report.
In: Radiation Hazards in Mining: Control Measurement and Medical Aspects.
Gomez M. (Editor).
New York, Society of Mining Engineers of American Institute of Mining, Metallurgical and Petroleum Engineers: pp. 359-62.

Muller J., Wheleler W.C., Gentleman J.F., Suranyi G. and Kusiak R. 1983.
Study of Mortality of Ontario Miners, 1955-1977. Part 1.
Toronto, Ontario, Minstry of Labour.

Muller J., Wheleler W.C., Gentleman J.F., Suranyi G. and Kusiak R. 1985.
Study of Mortality of Ontario Miners.
In: Occupational Radiation Safety in Mining, Volume 1.
Stocker E. (Editor).
Toronto, Ontario, Canadian Nuclear Association: pp. 335-43.

Muller J., Kusiak R. and Ritchie A.C. 1989.
Factors Modifying Lung Cancer Risk in Ontario Uranium Miners 1955-81.
Ontario Ministry of Labour.
Workers' Compensation Board of Ontario.
Atomic Energy Control Board of Canada.

Murphy E.A. 1978.
Epidemiologic Strategies and Genetic Factors.
International Journal of Epidemiology 7: pp. 7-14.

Mylvaganam A. 1993a.
Implications of Findings from the Radium Hill Study: The Protracted Exposure Effect,
Smoking, Asbestos and Temporal Modifiers of the Exposure-Response Relationship.
Accepted: Australian Radiation Protection Society 18th Annual Conference, Sydney.

Mylvaganam A. and Woodward A.J. 1993a;
Health of Uranium Mining on the Health of Workers: Findings From Radium Hill.
Presented at: The International Congress on Applied Mineralogy 1993:
Minerals in the Service of Mankind.
Perth, Australia.

Mylvaganam A. and Woodward A.J. 1993b;
Radon Exposure Limits: Implications of The Protracted Exposure Effect.
Accepted: 7 Years and Counting: Health Beyond 2000.
25th Annual Conference of the Public Health Association of Australia Inc. Sydney, Australia.

Mylvaganam A., Woodward A.J., Crouch P. and Thomas D.C. 1992;
Lung Cancer Mortality Amongst Former Workers at the Radium Hill Uranium Mine in South Australia.
In: Proceedings of the 9th International Symposium on Epidemiology in Occupational Health.
U.S. Department of Health and Human Services and University of Cincinnati: pp. 147.

Mylvaganam A., Woodward A.J. and Crouch P. 1993a.
A Study of Lung Cancer Mortality Amongst Workers in a South Australian Uranium Mine.
Accepted: International Epidemiological Association 13th Scientific Meeting,
Sydney Australia.

Mylvaganam A., Woodward A. and Crouch P. 1993b.
Risk of Lung Cancer Mortality Amongst Uranium Miners - Interpreting Findings from the Radium Hill Study in a Global Perspective.
Accepted: Australian Radiation Protection Society 18th Annual Conference, Sydney.

Nair R.C., Abbatt J.D., Howe G.R., Newcombe H.B. and Frost S.E. 1985.
Mortality Experience Among Workers in the Uranium Industry.
In: Proceedings of the International Conference on Occupational Radiation Safety in Mining.
Stocker H. (Editor).
Canadian Nuclear Association, Toronto, Ontario.

NCI 1992.
Radon-Exposed Underground Miners Workshop.
National Cancer Insitiute, National Institutes of Health, Bethesda, Maryland.

NCRP 78 1984.
Evaluation of Occupational and Environmental Exposues to Radon and Radon Daughters in the United States: NCRP Report 78.
United States National Council of Radiation Protection and Measurements.
Bethesda, Maryland.

NIH 1985.
Report of the National Institutes of Health Ad Hoc Working Group to Develop Radioepidemiological Tables.
Publication No. 85-2748. Bethesda, Maryland.

NIOSH 1987.
Criteria for a Recommended Standard ....
Occupational Exposure to Radon Progeny in Underground Mines.
National Institute for Occupational Safety and Health.
United States Government Printing Office DHSS Publication No. 88-101.

NRPB 1986.
Living With Radiation.
National Radiological Protection Board.

Pathak D.R., Morgan M.V. and Samet J.M. 1988.
Lung cancer epidemiology in New Mexico Uranium Miners.
In: Proceedings of the American Statistical Association Conference on Radiation and Health, Coolfont VII: pp. 24-5.

Pearce N. 1989.
Analytical Implications of Epidemiological Concepts of Interaction.
International Journal of Epidemiology 18-4: pp. 976-80.

Pearce N. and Checkoway H. 1987.
Epidemiologic Programs for Computers and Calculators.
A Simple Computer Program for Generating Person-Time Data in Cohort Studies Involving Time-Related Factors.
American Journal of Epidemiology 125-6: 1085-91.

Peto J. 1984.
Early- and Late-Stage Carcinogenesis in Mouse Skin and in Man.
In: Models, Mechanisms and Etiology of Tumour Promotion.
Borzsonyi M., Day N.E., Lapis K. and Yamasaki H. (Editors).
IARC Scientific Publications 56: pp. 359-371.
World Health Organization,
International Agency for Research on Cancer, Lyon, France.

Peto J. 1985.
Some Problems in Dose-Response Estimation in Cancer Epidemiology.
Methods for Estimating Risk of Chemical Injury:
Human and Non-Human Biota and Ecosystems.
In: Voug V.B., Butler G.B., Hoel D.G. and Peakall D.B. (Editors).
New York: Wiley: pp. 361-80.

Peto R., Roe F.J., Lee P.N., Levy L. and Clark J. 1975.
Cancer and Aging in Mice and Men.
British Journal of Cancer 32: pp. 422-6.

Preston D.L., Kato H., Kopecky K.J. and Fujita S. 1986.
Cancer Mortality Among A-bomb Survivors in Hiroshima and Nagasaki, 1950-1982.
Life Span Study, Report 10, Part 1. Technical Report 1-86.
Hiroshima, Japan: Radiation Effects Research Foundation.

Radford E.P. and Renard K.G.St.C. 1984.

Lung Cancer in swedish Iron Miners Exposed to low Doses of Radon Daughters.

New England Journal of Medicine 310: pp. 371-7.

Rothman K.J. 1986.

Modern Epidemiology.

Little, Brown and Company. Boston / Toronto.

Rothman K. and Keller A.Z. 1972.

The Effect of Joint Exposure to Alcohol and Tobacco on Risk of Cancer of the Mouth and Pharynx.

Journal of Chronic Diseases 25: pp. 711-16.

Rothman K.J., Greenland S. and Walker A.M. 1980.

Concepts of Interaction.

American Journal of Epidemiology 112: pp. 467-70.

SADM 1952.

The Radium Hill Mine.

South Australian Department of Mines.

SADM 1954.

Radium Hill South Australia.

Official Opening Souvenir.

South Australian Department of Mines.

SAHC 1986.
Deaths from Lung Cancer and Other Causes Among Former Workers at Radium Hill - An Interim Assessment.
South Australian Health Commission.

Samet J.M. 1989.
Radon and Lung Cancer.
Journal of the National Cancer Institute 81: pp. 745-57.

Samet J.M. 1992.
Lung Cancer Epidemiology in New Mexico Uranium Miners.
In: Radon Research Program.
United States Department of Energy, Office of Energy Research, Office of Health and Environmental Research: pp. 225-7.

Samet J.M., Kutvirt D.M., Waxweiler R.J. and Key C.R. 1984.
Uranium Mining and Lung Cancer in Navajo Men.
New England Journal of Medicine 310-23: pp.1481-4.

Samet J.M., Pathak D.R., Morgan M.V., Marbury M.C., Key C.R. and Valdivia A.A. 1989.
Radon Progeny Exposure and Lung Cancer Risk in New Mexico Uranium Miners: A Case-Control Study.
Health Physics 56: pp. 415-21.

Samet J.M. and Hornung R.W. 1990.
Review of Radon and Lung Cancer Risk.
Risk Analysis 10-1: pp. 65-75.

Samet J.M., Pathak D.R., Morgan M.V., Key C.R., Valdivia A.A and Lubin J.H. 1991.
Lung Cancer Mortality and Exposure to Radon Progeny in a Cohort of New Mexico Underground Uranium Miners.
Health Physics 61-6: pp. 745-52.

SAS 1990.
SAS Users Guides for Personal Computers.
Version 6.
SAS Institute Inc. U.S.A.

Saracci R. 1980.
Interaction and Synergism.
American Journal of Epidemiology 112: 465-66.

Schulttmann W. 1993.
Historical Perspectives in Occupational Medicine
Schneeberg Lung Disease and Uranium Mining in the Saxon Ore Mountains (Erzgebirge).
American Journal of Industrial Medicine 23: pp. 355-68.

SENES 1984.
Assessment of the Scientific Basis for Existing Federal Limitations on Radiation Exposure to Underground Uranium Miners.
Prepared for the American Mining Congress.
SENES Consultants Limited, Richmond Hill, Ontario.

SENES 1990.
An Evaluation of the Risk of Exposure to Radon Daughters.
Report Prepared for the American Mining Congress.
SENES Consultants Limited, Richmond Hill, Ontario.

SENES 1991.
Detailed Reconstruction of Radon Daughter Exposures of Eldorado Beaverlodge Uranium Mine Employees.
Report Prepared for the Atomic Energy Control Board of Canada.
SENES Consultants Limited, Richmond Hill, Ontario.

Sevc J., Placek V. and Jerabek J. 1971.
Lung Cancer Risk in Relation to Long-Term Radiation Exposure in Uranium Miners.
In: Proceedings of the Fourth Conference on Radiation Hygiene
CSSR - 1971 JASNA POD CHOPKOM: pp. 315-325.

Sevc J., Kunz E. and Placek V. 1976.
Lung Cancer in Uranium Miners and Long-Term Exposure to Radon Daughter Products.
Health Physics 30: pp. 433-7.

Sevc J., Kunz E., Tomasek L., Placek V. and Horacek J. 1988.
Cancer in Man After Exposure to Rn Daughters.
Health Physics 54: pp. 27-46.

Sevc J., Tomasek L., Kunz E., Placek V., Chmelevsky D., Barclay D. and Kellerer A.M. 1993.
A Survey of the Czechoslovak Follow-up of Lung Cancer Mortality in Uranium Miners.
Health Physics 64-4: pp. 355-69.

Siemiatycki J. and Thomas D.C. 1981.
Biological Models and Statistical Interactions: An Example from Multistage Carcinogenesis.
International Journal of Epidemiology 10-4: 383-7.

Southwood R. 1987.

Opening Remarks

In: Radiation and Health - The Biological Effects of Low-Level Exposure to Ionizing Radiation.

Jones R.R. and Southwood R. (Editors).

John Wiley and Sons Limited. pp. 3-6.

Strong J.C., Laidlaw A.J. and O'Riordan M.C. 1975.

Radon and its Daughters in Various British Mines.

Chilton: National Radiological Protection Board: Publication No. R39.

Thomas D.C. 1981.

Are Dose-Response, Synergy and Latency Confounded?

Annual Meeting of the American Statistical Association, Detriot.

Thomas D.C.1982.

Temporal Effects and Interactions in Cancer: Implications of Carcinogenic Models.

Environmental Epidemiology: Risk Assessment

In: Prenctice R.L. and Whittemore A.S. (Editors).

Philadelphia, SIAM: pp. 107-21.

Thomas D.C. 1983.

Statsitical Methods for Analyzing Effects of Temporal Patterns of Exposure on Cancer Risks.

Scandinavian Journal of Work Environmental Health 9: pp. 353-66.

Thomas D.C. 1987.

Pitfalls in the Analysis of Exposure-Time-Response Relationships.

Journal of Chronic Diseases 40-2: pp. 71S-8S.

Thomas D.C. 1988.
Models for Exposure-Time-Response Relationships With
Applications to Cancer Epidemiology.
Annual Review of Public Health 9: pp. 451-82.

Thomas D.C. 1990.
A Model for Dose Rate and Duration of Exposure Effects in Radiation Carcinogenesis.
Environmental Health Perspectives 87: pp.163-71.

Thomas D.C., McNeill K.G. and Dougherty C. 1985.
Estimates of Lifetime Lung Cancer Risks Resulting from Rn Progeny Exposures.
Health Physics 49: pp. 825-46.

Thomas D.C., Brown C.C., Chu K.C., Goldsmith D.F. and Saracci R. 1987.
Time Related Factors in Cancer Epidemiology.
Journal of Chronic Diseases 40-2: pp. V-VI.

Thomas D.C. and Whittemore A.S. 1988.
Methods for Testing Interactions, With Appliations to Occupational Exposures, Smoking, and Lung Cancer.
American Journal of Industrial Medicine 13: pp. 131-47.

Thomas D.C., Blettner M. and Day N.E. 1992.
Case-Control Study of Acute and Nonlymphocytic Leukemia.
Journal of the National Cancer Institute 84-20: pp. 1600-1601.

Tirmarche M., Brenot J., Piechowski J., Chameaud J. and Paradel J. 1984.
The Present State of an Epidemiological Study of Uranium Miners in France.
In: Proceedings of the International Conference on Occupational Radiation Safety in Mining.
Stocker H. (Editor).
Canadian Nuclear Association, Toronto, Canada 1: pp. 344-9.

Tirmarche M., Raphaelen A., Allin F., Shameaud J. and Bredon P. 1993.
Mortality of a Cohort of French Uranium Miners Exposed to Relatively Low Radon Concentrations.
British Journal of Cancer 67: pp. 1090-7.

UNEP 1985.
United Nations Environment Programmme 1985.
Radiation Doses, Effects, Risks.
United Nations Publication.

UNSCEAR 1977.
United Nations Scientific Committee on the Effects of Atomic Radiation
Sources and Effects of Ionizing Radiation.
New York, United Nations.

UNSCEAR 1982.
Ionizing Radiation: Sources and Biological Effects.
Report to the General Assembly, with Annexes.
New York, United Nations.

UNSCEAR 1988.
Sources, Effects and Risks of Ionizing Radiation.
United Nations Scientific Committee of the Effects of Atomic Radiation.

Upton A.C. 1987.
Health Effects of Low-Level Ionising Radiation: Biological Basis for Risk Assessment.
In: Health Effects of Low-Dose Ionising Radiation - Recent Advances and their Implications.
Proceedings of the International Conference Organised by the British Nuclear Energy Society and Co-Sponsored by the European Nuclear Society.
British Nuclear Energy Society, London. pp. 3-13.

US DOE 1988.
United States Department of Energy, Office of Energy Research, Office of Health and Environmental Research. Literature Survey Series.
Radon Epidemiology: A Guide to the Literature.

US DOE 1992.
United States Department of Energy, Office of Energy Research, Office of Health and Environmental Research. Literature Survey Series.
Radon Research Program: Radon.

US EPA 1987.
Radon Reference Manual.
United States Environmental Protection Agency, Office of Radiation Programs.
EPA Publication No. 520/1-87-20.

US PHS 1964.
Report of the Advisory Committee to the Surgeon General of the U.S.Public Health Service.
PHS Publication 1103 US-DHEW: pp. 96-8, 113-6.

Wagoner J.K., Archer V.E., Lundin F.E.Jr., Holaday D.A. and Lloyd J.W. 1965.
Radiation as a Cause of Lung Cancer Among Uranium Miners.
New England Journal of Medicine 273: pp. 181-8.

Waxweiler R.J., Roscoe R.J., Archer V.E., Thun M.J., Wagoner J.K. and Lundin F.E.Jr. 1981.
Mortality Follow-up Through 1977 of the White Underground Uranium Miners Cohort Examined by the United States Public Health Service.
In: Radiation Hazards in Mining: Control, Measurement, and Medical Aspects.
Gomez M. (Editor).
New York, Society of Mining Engineers of American Institute of Mining, Metallurgical, and Petroleum Engineers pp. 823-30.

Weast R.C. 1985.
CRC Handbook of Chemistry and Physics, 66th edition (Editor).
Boca Raton, Florida, CRC Press: p. B-133.

White C. and Bailar J.C. 1956.
Retrospective and Prospective Methods of Studying Association in Medicine.
American Journal of Public Health 46: pp. 34-44.

Whittemore A.S. 1977.
The Age Distribution of Human Cancer for Carcinogenic Exposures of Varying Intensity.
American Journal of Epidemiology 106: pp. 418-32.

Whittemore A.S. and Keller J.B.1978.
Quantitative Theories of Carcinogenesis.
SIAM Review 20: pp. 1-30.

Whittemore A.S. and McMillan A. 1983.
Lung Cancer Mortality Among U.S. Uranium Miners: A Reappraisal.
Journal of the National Cancer Institute 71: pp. 489-99.

Wolfenden H.H. 1923.
On the Methods of Comparing the Mortalities of Two or More Communities, and the Standardization of Death Rates.
Journal of the Royal Statistical Society 86: 399-411.

Woodward A., Roder D.M., McMichael A.J. and Lewis S.Y. 1988.
Ionising Radiation and Cancer Incidence in Uranium Miners: An Historical Cohort Study.
In: Progress in Occupational Epidemiology.
Hogstedt C. and Reuterwall C. (Editors).
Elsevier Science Publishers B.V. (Biomedical Division).

Woodward A.J., Roder D., McMichael A.J., Crouch P. and Mylvaganam A. 1991.
Radon Daughter Exposures at the Radium Hill Uranium Mine and Lung Cancer Rates Among Former Workers, 1952-1987.
Cancer Causes and Controls 2:213-30.

Xuan X.Z., Lubin J.H., Li J.Y., Yang L.F., Luo Q.S., Yang L., Wang J.Z. and Blot W.J. 1992.
A Cohort Study in Southern China of Tin Miners Exposed to Radon and Radon Decay Products.
Health Physics 64-2: pp. 120-31.

Yule G.U. 1934.
On Some Points Relating to Vital Statistics, More Especially Statistics of Occupational Mortality.
Journal of the Royal Statistical Society 97: 1-84.

Zhang J.Y., Wang X.H., Huang X.H. 1981.
Investigations of the Radioactive Etiological Factors of Lung Cancer Among Miners.
Acta Acad. Med. Sinicae 3; pp. 127-30. (In Chinese).

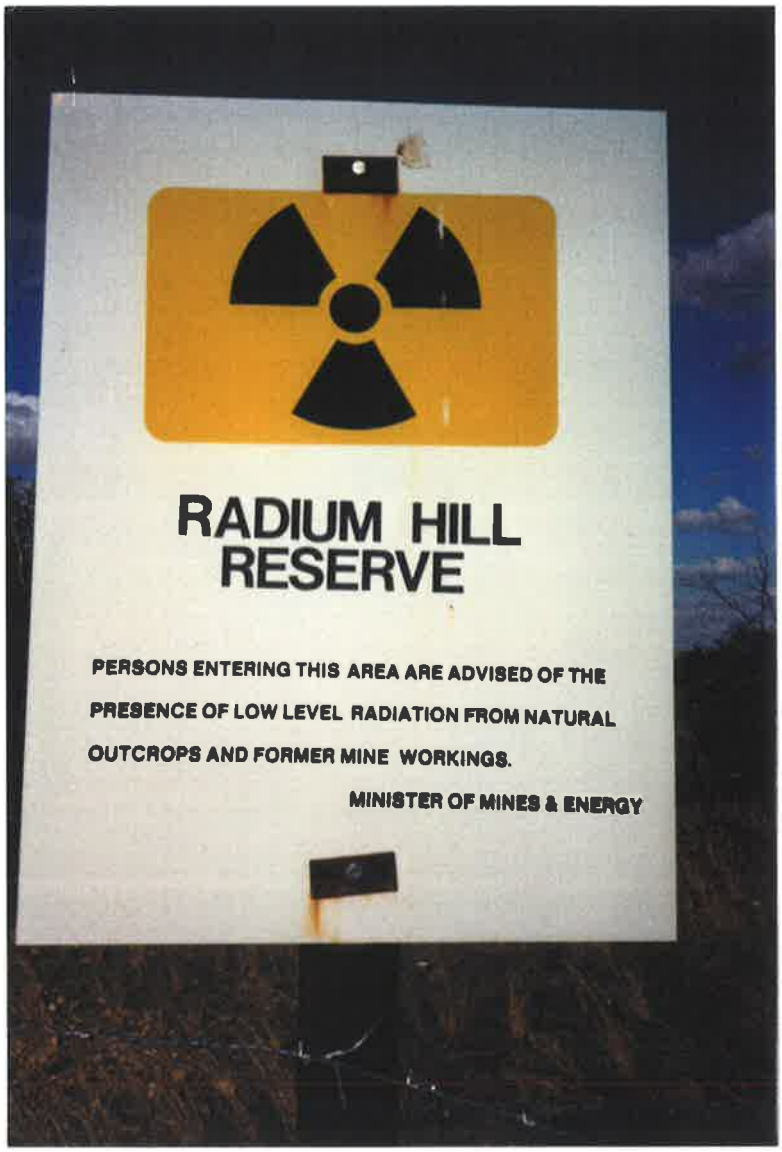
APPENDIX A

A. Album of Photographs

Contents

Radium Hill - 1991

What Remains of a Once Busy Mining Town

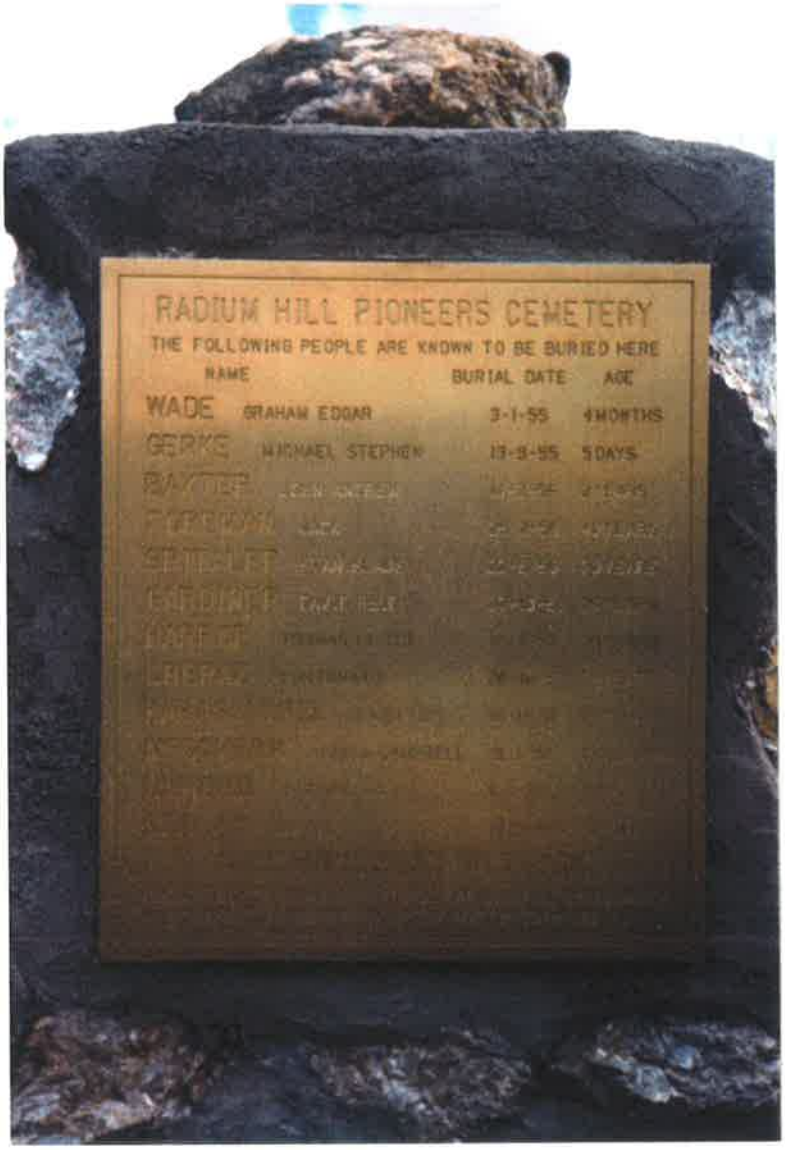

RADIUM HILL
RESERVE
PERSONS ENTERING THIS AREA ARE ADVISED OF THE
PRESENCE OF LOW LEVEL RADIATION FROM NATURAL
OUTCROPS AND FORMER MINE WORKINGS.
MINISTER OF MINES & ENERGY

The Radium Hill Reunion 1998 and the Old Mine Site


RE-UNION '91
RADIUM HILL
WELCOMES
YOUR RETURN

Radium Hill Pioneers Cemetry

Unveiling of the Commemoration Stone 199[illegible]


RADIUM HILL PIONEERS CEMETERY
THE FOLLOWING PEOPLE ARE KNOWN TO BE BURIED HERE
NAME
BURIAL DATE
AGE
WADE GRAHAM EDGAR 3-1-55 4MONTHS
GERKE MICHAEL STEPHEN 13-3-55 5DAYS

Some Innovative Attempts at Cohort Trace

'Grave Yard Trace'

A Trace From A Grave Stone at the Radium Hill Cemetry


LJP

The Missing Persons' Bureau At The Radium Hill Reunion

The Missing Persons' Bureau

At The Adelaide Reunion of Radium Hill Residents

Life and Times at Radium Hill 1948-1962

HE SOUTHER
CROSS
ORGAN OF THE CATHOLIC
HURCH IN SOUTH AUSTRALIA
THE OTHER
OF YOUR
Published W
Established 1889
THURSDAY, APRIL 24, 1958
Registered at the G.P.O., Adelaide, for transmission by post as a newspaper.
XX, No. 3521

THE WILDERNESS TAKES HEART, AND BLOSSOMS

(BY OUR OWN REPRESENTATIVE)

WASN'T Radium Hill that Isaias was talking about when he said, "The wilderness takes heart, and blossoms, fair as the lily." But it could ve been. At least that is what the visitor thinks as he goes down the ee-lined streets of Radium Hill and looks over front fences at flourishing rubs and green lawns and bright flowers that surround the homes. But at's only part of the story, and, in a way, the least part of it.

don't know whether uranium has a magnetic power ttraction or not. My impression is that it has not. ertheless, Radium Hill was a magnet with great drawing r on Wednesday night last week, when the Bishop of Pirie, Most Rev. B. Gallagher, blessed and opened the Catholic School there.

But let's begin at the beginning.

began when they started for uranium in a big way Hill, ten years or so Carmody, who was orbit- the Catholic Inland Mis- took Radium Hill into calling and saying regularly. In mining work expanded population grew, Fr. W. the first resident Fr. V. Shiel was priest. Keeping down, he was building a church.

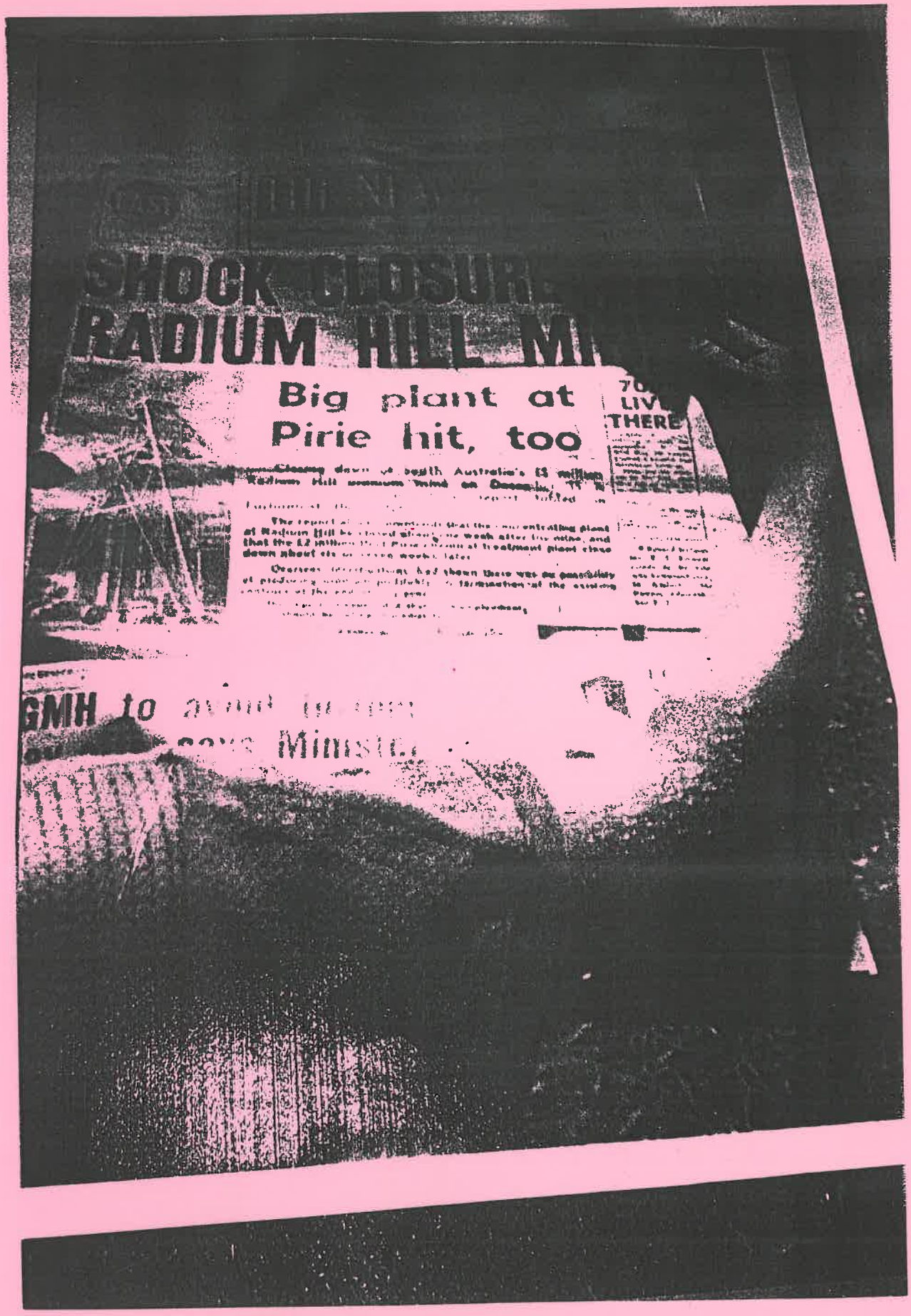
DO IT YOURSELF

architect, he was a skilled tradesman in some six trades associated with building. There used to be an oil called three-in-one. Fr. Shiel might well have been called six-in-one. He was a bricklayer, a carpenter, a painter, a plumber, a joiner, a plasterer.

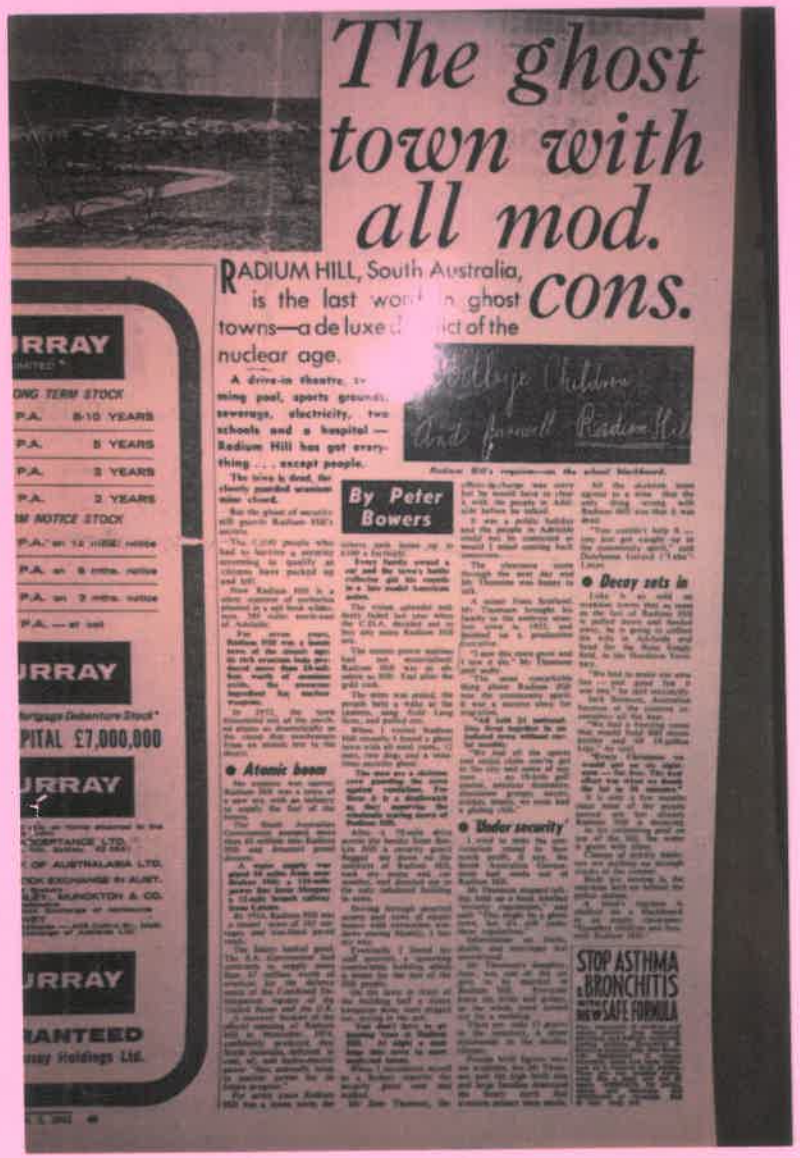
ONE-MAN BUILDING TEAM

So he set this one-man building team to work—after first having drawn the plans for his church and done a bit of preliminary laboring. The laboring work was the making of some 6,000 cement bricks.

News Paper Clippings: Life at Radium Hill



News Paper Clippings: Closure of the Radium Hill Mine



The ghost town with all mod. cons.

RADIUM HILL, South Australia, is the last word in ghost towns—a de luxe district of the nuclear age.

A drive-in theatre, swimming pool, sports grounds, sewerage, electricity, two schools and a hospital — Radium Hill has got everything . . . except people.

The mine is dead, the cleverly guarded uranium mine closed.

But the ghost of security still guards Radium Hill's secrets.

By Peter Bowers

Radium Hill's requiem—on the school blackboard.

• Atomic boom

• Decay sets in

• Under security

STOP ASTHMA & BRONCHITIS WITH THE SAFE FORMULA

THE SUNDAY ADVERTISER PICTURE SUPPLEMENT — SATURDAY NIGHT, NOVEMBER 13, 1954—PAGE 5E

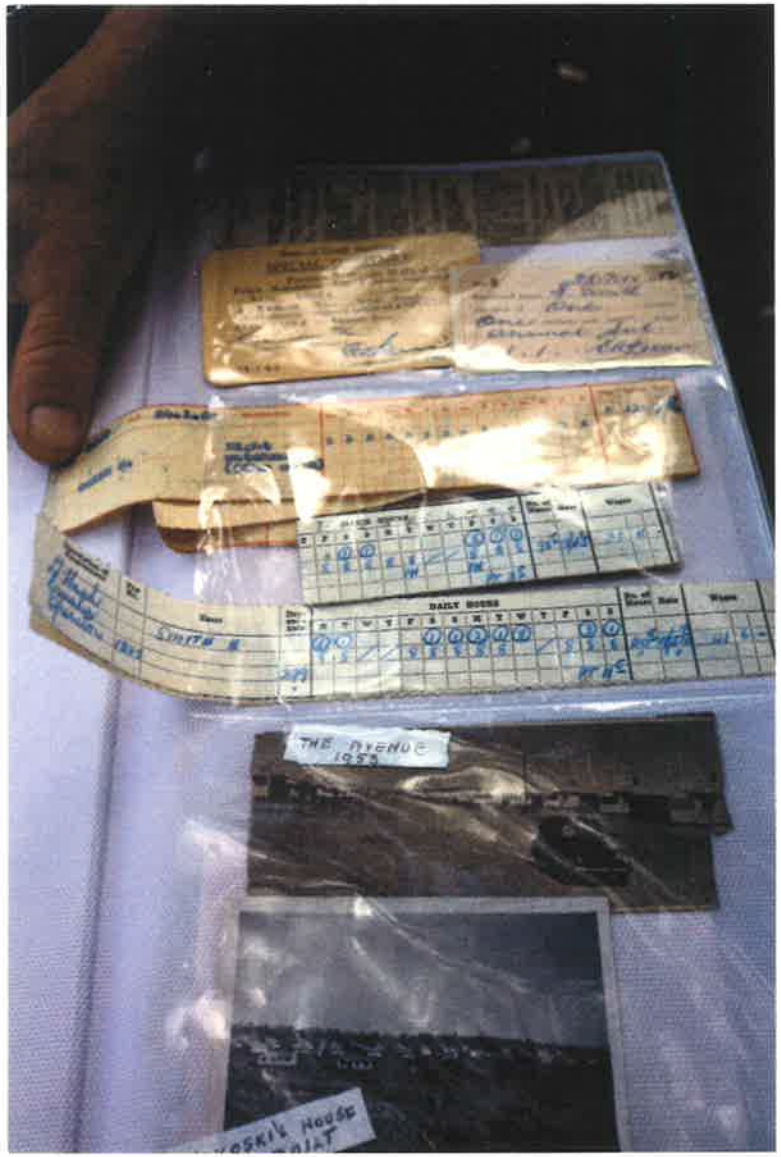
S.A. URANIUM PROJECT.—The Governor-General (Sir William Slim) (on Wednesday) pressed a button from the dais at the foot of this 330 ft. headframe (right) over Radium Hill's main 750 ft. shaft, as started crushers on SA's £5m. uranium projects. He was watched by miners (above) and 350 guest

ALL MOD CONS.—The Premier (Mr. Playford), above, chief architect of the project, among cooks and waitresses who prepared and served lunch for the guests. Mr. Playford has announced that within three years a similar mine 70 miles north-west of Radium Hill will also supply concentrate to the almost completed £13m. Port Pirie refinery. Right—Mrs. T. O'Connell in her 3-fabulous home near the mine.

News Paper Clippings: Farewell To Radium Hill

The Radium Hill Golf Course:
An Ex-Resident Finds The Hole Named After Him 40 Years Previously

Memorabilia: A Radium Hill Scrap Book


THE AVENUE
1958

APPENDIX B

B. Instruments of Data Collection

Contents

FOR COMPLETION BY PREVIOUS EMPLOYEE OF THE

RADIUM HILL MINE

(Please telephone Kaye Robinson on if you require help answering these questions)

KING HISTORY

CIGARETTES

If you have EVER smoked cigarettes regularly (i.e., at least one cigarette per day for six months or longer), please complete either Section 1 OR 2.

If you have NOT smoked cigarettes regularly, go directly to B.

Section 1 - IF YOU STILL SMOKE CIGARETTES REGULARLY

(1) Approximately how old were you when you started? .....yrs. of age.

(2) For about how many years have you smoked cigarettes regularly? (If less than one year, indicate "1") .....yrs.

(3) About how many cigarettes do you usually smoke daily? .....cigarettes.

(4) About how many cigarettes did you usually smoke daily at Radium Hill? (indicate '0' if none) .....cigarettes.

Section 2 - IF YOU HAVE GIVEN UP SMOKING CIGARETTES REGULARLY

(1) Approximately how old were you when you started smoking cigarettes regularly? .....yrs. of age.

(2) For about how many years did you smoke cigarettes regularly? (If less than one year, indicate "1") .....years.

(3) About how many cigarettes did you usually smoke daily? .....cigarettes.

(4) About how many cigarettes did you usually smoke daily at Radium Hill? (indicate'0' if none) .....cigarettes.

(5) How long ago did you give up smoking cigarettes? (indicate '0' if less than six months) .....years.

CIGARS

Have you ever smoked cigars regularly (at least weekly)? Yes ☐ No ☐

If YES - About how old were you when you started? .....yrs. of age.

Did you smoke cigars regularly at Radium Hill? Yes ☐ No ☐

If YES -

Indicate how many per week. .....cig./wk.

Do you still smoke cigars regularly? Yes ☐ No ☐

If NO -

Indicate how long since you stopped (Indicate '0' if less than six months) .....years.

PIPE

Have you ever smoked a pipe regularly (at least weekly)? Yes ☐ No ☐

If YES - About how old were you when you started? .....yrs.of age

Did you smoke a pipe regularly at Radium Hill? Yes ☐ No ☐

Do you still smoke a pipe regularly? Yes ☐ No ☐

If NO -

Indicate how long since you stopped (Indicate '0' if less than six months) .....years

---

EMPLOYMENT HISTORY

A. URANIUM OR OTHER RADIOACTIVE MATERIALS

(1) BEFORE WORKING AT RADIUM HILL

Did you work in a uranium mine or with uranium ore or another radioactive material before going to Radium Hill? Yes ☐ No ☐

If YES - for how many years? .....years

(2) AFTER WORKING AT RADIUM HILL

Did you work in a uranium mine or with uranium ore or another radioactive material after being at Radium Hill? Yes ☐ No ☐

If YES, give names of companies and approximate years involved.

Companies	Years Involved
.....	from 19.... to 19....
.....	from 19.... to 19....
.....	from 19.... to 19....
.....	from 19.... to 19....

B. ASBESTOS

Have you ever worked for an asbestos mining company? Yes ☐ No ☐

Have you ever worked in the areas of asbestos milling/ asbestos insulating. Yes ☐ No ☐

If YES to either of the above questions, indicate names of companies and approximate years involved.

Companies	Years Involved
.....	from 19.... to 19....
.....	from 19.... to 19....
.....	from 19.... to 19....
.....	from 19.... to 19....

---

you for your assistance. Please sign and return to the ment of Community Medicine in the "Reply Paid" envelope.

.....
(signature)

APPENDIX C

C. Overview of Analytical and Inferential Methodology

Page

C.1 Epidemiological Concepts Relevant to This Work

The aim of this appendix is to provide an overview of the analytical and inferential methodology underlying this work and has not been detailed in the main text of this thesis.

C.1.1 Outline of Research Methodology

The purpose of this work is to study the effects of ionizing radiation on lung cancer mortality. It is known that lung cancer does not have a unique causative agent (Doll and Peto 1981). The *risk* of an individual developing lung cancer is influenced by the individual's own susceptibility to cancer induction and exposure to one or more causative agents. Isolating the effect of a single causative factor such as ionizing radiation on lung cancer is therefore complicated by the presence of *multifactorial causation* i.e., by the effect of other potential causative agents and their interactions. Moreover, it is also known that there are factors that may *confound*, and factors that may *modify*, the *effect of exposure to the risk factor* - ionizing radiation - on the *response* - lung cancer mortality. Therefore, before attempting to isolate the effect of radiation exposure on lung cancer, the roles of effect modifiers and other covariates must be examined and adjustments made for the existence of potential confounding factors.

This section (C.1) provides an overview of the epidemiological process of enquiry pertaining to such an investigation and introduces the epidemiological concepts and quantitative measures which provide the basis for and govern the use of the quantitative techniques presented in the analytical chapters of this work. These issues are described in detail to help bridge the gaps (in my understanding) between epidemiological research questions, the application of statistical methods in epidemiology and the process of drawing epidemiological inferences from the statistical methods applied. However, detailed appraisal of issues that have been well documented in the literature are avoided here. I include this appendix in my thesis because it is submitted for a doctoral degree based on a research submission alone.

C.1.2 The Process of Enquiry and its Rationale

Epidemiology has been broadly defined as the study of the distribution and determinants of health-related states or events in specified populations (Last 1988). The principles and methods of epidemiology are based on two fundamental assumptions: that human disease does not occur at random, and that human disease has causal and preventive factors that can be identified through systematic study. These assumptions give rise to a comprehensive working definition that encompasses the epidemiological principles and methods relevant to this work - *"the study of the distribution and determinants of disease frequency" in human populations* (MacMahon and Pugh 1970; Hennekens and Buring 1987). The goal of epidemiology may therefore be described as: making measurements - measurements of descriptors and associations, cause and effect; and formulating explanations - especially causal explanations.

The field of epidemiology has traditionally been considered as comprising two branches viz., descriptive and analytic epidemiology. Descriptive epidemiology deals with the measurement of disease frequency or health-related states, and analytical epidemiology concerns, measuring the nature and the role of the determinants of these health-related states.

The process of epidemiological enquiry may broadly be described as involving three logical phases, viz. *the determination and summarization of facts*, *the formulation of possible explanations for features observed*, and *the testing of these explanations and identification of any additional information required* for this purpose. It is generally an iterative process of enquiry starting from any of the logical stages mentioned above. Depending on the circumstances of the study this quest can be approached either *deductively* or *inductively*. In keeping with traditional demarcations, descriptive studies were regarded as a basis for *hypothesis generation* and often followed a deductive approach of enquiry whilst, an inductive approach was considered more suited to analytical studies aimed at *hypothesis testing*.

This notion of dividing epidemiological research into descriptive and analytical compartments evolved from earlier conceptualisations of the philosophy of science based on the empirical views of *Bacon* and *Hume*. The process of hypothesis generation and hypothesis testing however, does not necessarily follow a solely *Baconian* inductive approach, nor a deductive approach as advocated by *Hume*. In its applications, the practical process of hypothesis generation and testing is not only iterative in nature, but also suggests that the underlying scientific philosophy is governed more by the *Popperian* notion of *conjectures* and *refutations,* rather than by purely inductive or deductive approaches (Charlesworth 1982). In the light of this broader understanding of the scientific thought process, the notion that descriptive studies generate hypotheses whilst analytical studies test them, does not seem to be very appropriate (Rothman 1986). Moreover, in practice epidemiological studies are often seen to have a combination of descriptive and analytical components, and to utilise a mixed approach of enquiry.

In the traditional sense, the study of exposure in relation to disease outcome in individuals, as in this work, has been regarded as the domain of analytical epidemiology (IARC 1990). However, the analytical methods used in these studies are often based on basic descriptive quantitative measures and techniques. The use of a combined approach of enquiry is therefore inevitable in this work which will continue, independent of the specific distinctions between descriptive and analytical methodology. This appendix will proceed by outlining the epidemiological and statistical concepts relevant to the process of enquiry adopted in this work.

It is obvious from the foregoing that statistical methods play an important role in epidemiological studies. The role of statistics in epidemiology must be determined on the basis that epidemiological objectives should set the requirements for study design and analysis of data and steer the course of statistical applications in epidemiology. Therefore, statistical methods are merely the tools to achieve these epidemiological objectives and should be adopted or if necessary invented for this purpose only (Rothman 1986).

C.1.3 The Quest: Causal Explanations

The aim of epidemiology is to describe disease frequency, identify determinants of diseases and provide *causal explanations*. Any factor whose modification alters the outcome under consideration may be regarded as a *potentially causal factor* (Abramson 1988). *Causes* are generally multiple and *potential causes* need not be necessary or sufficient. Potential causes may be predisposing, enabling, precipitative, reinforcing, concomitant or intermediate.

The role of causes or potential causes are evaluated through their *effect* with relation to the occurrence of disease. In a quantitative sense, an effect may be described as the difference in disease occurrence between groups of people who differ with respect to a particular causal characteristic, generally referred to as an *exposure* (Rothman 1986). The effect of potential cause is basically measured through the association between cause and effect, which may be referred to as *cause-effect association* or *causal association*. A *causal association* may be defined as an *association* between two categories of events where the change in the frequency or quality of one event follows alteration in the other event (MacMahon *et al*. 1960).

For causation to be possible it is necessary that an *association* exists between the outcome of interest and the potential cause; this follows directly from the fundamental statistical requirement that causation can only exist in the presence of correlation. It is therefore necessary to identify association before attempting to establish causation.

C.1.4 The Process of Formulating Causal Explanations

I. Identifying Association

An association is said to be present between two variables if the presence or magnitude of one variable depends on that of the other variable. If the rate of an outcome of interest in one group is the same as in another group, then it could be said that there is no association between the grouping variable and the outcome variable. A difference in the rates of the outcome variable indicates an association.

II. Association and Causation

The existence of an association however, does not imply causation just as correlation does not imply causation. Association can also result from an artifact of the study design, be due merely to chance, or reflect the influence of other related factors.

Therefore, once an association has been identified, before explanations as to the role of specific determinants can be reached, it is necessary to identify whether the association could be *artifactual* rather than *factual.* It is also important to examine the possibility of merely *fortuitous association*, by considering the *role of statistical significance.* Furthermore, the *influence of extraneous factors* on the *strength* and *consistency* of the association must be critically examined. Causal explanations can only be considered thereafter.

The logical process of establishing causal explanations may therefore, be said to consist of two phases, that of *identifying association*, and then, *examining the possibility of exceptions to a causal association.* The latter process of *excluding causal explanations* will now be described by examining its logical phases viz., *the role of artifactual associations*, *the role of statistical significance* and *the influence of extraneous factors present.*

III. The Role of Artifactual Associations

Associations may arise as a result of artifacts of study design and implementation. The major sources of artifactual association include selection bias, response and non-response bias, measurement bias, measurement validity and reliability, and information bias. Artifactual associations may therefore be identified through a critical appraisal of the design and implementation features of the study.

IV. The Role of Statistical Significance

After an association has been established as factual rather than artifactual, it is necessary to examine whether the association may have risen merely by chance. Fortuitousness of an association can be examined by testing its statistical significance. A statistical test of significance provides a *p-value* which gives the probability that, even if an association actually did not exist, chance processes alone would produce an association as strong as, or stronger than, the one estimated or measured. Statistical significance based on *p-values*, is appraised through a pre-defined *critical value* α (eg., $\alpha = 0.05$). A *p-value* of less than α (eg., 0.05 or *a 1 in 20 chance*), is often regarded as justification for regarding an association as nonfortuitous. If *p* is very small, the association may be regarded as *highly significant*; one which cannot be reasonably attributed to chance alone. It should be noted however, that, such statistical tests have built-in errors. If a *critical value* of 0.05 is used, chance processes alone would produce a verdict of *statistically significant* in about 5 out of every 100 tests performed, *even if no real associations exist.*

Statistical significance does not indicate whether an association actually exists; it only *guards against chance*, and yet, does not prove that an association is *not due to chance.* Statistical significance merely implies that an association is unlikely to be due to chance alone and gives a *degree of confidence* in regarding it as nonfortuitous.

A statistically non-significant result does not *prove* that the association is due to chance; it implies that a chance process may easily produce such an association. A statistically non-significant association, particularly when based on a large number of observations, indicates that there is probably *no strong nonfortuitous association.*

Most importantly, it must be noted that even a statistically significant association does not imply causation because it could be *artifactual* or a consequence of the influence of other related factors.

It can be seen from the above mentioned issues that unless the role of statistical significance is critically examined, it can often be misleading. The presentation of *p*-values alone may lead to the blatant misinterpretation of statistical significance which can be avoided by using *standard errors* or *confidence intervals* which are more safely and easily interpretable by the unwary. Rothman (Rothman 1986) states: *"the term 'statistical significance' can be expunged from the lexicon of the epidemiologist with no loss."* This however, is arguable. The notion of statistical signifcance is of fundamental importance in any quantitative research that is subject to stochastic variability; despite its dubiousness, statistical significance is perhaps, the only available tool for evaluating the role of chance. In weighing the *pros* and *cons,* the most satisfying practical solution may therefore lie in reporting both, confidence intervals and statistical significance *p-values.*

V. The Role of Extraneous Factors

It is often difficult to identify the role of specific determinants on the outcomes studied even after excluding artifactual explanations and establishing statistical significance. This happens in the presence of *multifactorial causation*, a situation where there is more than one factor that determines the outcome which can be complicated further by some of these factors being inter-related, thereby making it difficult to isolate the effects of specific determinants.

In epidemiological investigations, special importance is given to two specific phenomena which are manifestations of extraneous factors and referred to as *confounding* and *effect modification,* which are described below.

i. The Role of Confounding:

The association between two variables can be *confounded* by any third varable that influences the *outcome* of interest and is associated with the *predictor variable*, without being an intermediate link in the chain of causation connecting the other two variables. A confounding factor is one that is associated with the potential cause being examined and is an independent cause of the outcome of interest without being an intermediate cause linking the potential cause and the outcome. In epidemiological investigations, "*universal variables*" (Abramson 1988), such as age, sex, parity, ethnic group, social class, religon, marital status etc., should always be considered as potential confounders.

Confounding may diminish, reverse or exaggerate an association. Furthermore, confounding may mask an existing association or produce an apparent association when one does not exist. Therefore, when confounding occurs, an undistorted picture of the effect of the potential cause can be obtained only by controlling for the confounder.

Confounding may be considered *a nuisance effect* that interferes with the fundamental research question or association under consideration. This nuisance effect can be prevented through study design features or identified and controlled whereever possible, in the analytical stage of the study.

ii. The Role of Effect Modification:

Effect modification arises when the association between two factors is modified by the state of another factor i.e., when the magnitude of an effect measure is changed by the state of a factor other than exposure and disease, which is generally referred to as an *effect modifier*.

Rothman (Rothman 1986), suggests that effect modification should not be regarded as a nuisance effect. Unlike confounding which is a bias that needs to be prevented, controlled or removed from the data, effect modification is an elaborated description of the effect itself. Effect modification is thus, a finding to be reported rather than a bias to be avoided. The general aim at the analytical stage of epidemiological studies is therefore to identify and eliminate confounding, and to identify and describe effect modification.

Effect modification can be identified by elaborating an association through stratification of the potential effect modifier. Effect modification is deemed to exist when the stratum specific measures of association are not constant and found to vary between stratum.

Constant or slightly varying stratum specific measures of association need not necessarily confirm the absence of effect modification. They may be reflect a lack of refinement in the association where the degree of stratification exercised is not adequate to demonstrate the existence of the modifying effect. In terms of statistical modelling, effect modification is a corollary to *statistical interaction* between two independent variables in their association with a third variable.

iii. Confounders and Effect Modifiers:

A factor may be a confounder and an effect modifier, a confounder alone, an effect modifier alone, or neither a confounder nor a modifier. If it is a confounder as well as an effect modifier, associations between this factor and the determinant of interest or the outcome of interest may not necessarily be obvious. Associations may only be obvious at certain levels of the nuisance factor and may be drowned when it is not stratified or inappropriately stratified. When a modifying effect is extremely strong it can be argued that the confounding effect becomes irrelevant.

iv. Methods of Addressing Confounding and Effect Modification: An Overview

A confounding effect is determined by the presence, direction and the strength of the associations between potential confounders and other variables. Therefore, a variable can confound the association between two other variables only if it is associated with both of them. However, this association may not be ostensible. Although the statistical significance of these associations is irrelevant, for a confounding effect to be of any importance these associations between the potential confounders and the exposure and outcome variables must be strong. Weak associations even if statistically significant, are unlikely to produce an important confounding effect, whereas, strong associations even if statistically insignificant (usually because of the number of observations being small) may produce a substantial confounding effect (Abramson 1988). Though statistical significance is considered irrelevant in identifying potential confounders, it is useful in appraising the relative importance of potential confounders and deciding which to control (Fleiss 1986). It should however be noted that it is "simply incorrect" (Rothman 1986), to use statistical significance to assess confounding in any other sense.

Strategies for identifying confounding include the *exclusion test,* elaborating the association through *stratification,* and *standardization.*

The *exclusion test* is based on the surmise that since the conditions necessary for confounding to exist are known, if these conditions are definitely not met, then the possibility of confounding can be disregarded. The *exclusion test* however, is not foolproof and may be misleading if the suspected confounder is also an effect modifier in which case *conditional associations* may exist even though *crude associations* may be absent.

Confounding and effect modification can be identified at the analysis phase of a study through critical assessment of associations, which includes processes such as the *refinement* and *elaboration of associations*. Refinement refers to the procedure of using finer instead of broader categories of the potential *nuisance factors*, so as to throw additional light on the association under investigation. The simplest form of elaborating an association consists of *stratification* according to the categories of the potential nuisance factors. Stratification provides a means of identifying and demonstrating confounding and effect modification. Other strategies for identifying confounding at the time of analysis include the *exclusion test* and *standardization*. Stratification can be followed by methods of pooling stratum-specific measures of association, such as the *Mantel-Haenszel procedure*, the *method of inverse variance weighting* and *maximum likelihood methods,* to provide summary measures of association after controlling for the effect of potential confounders (Rothman 1986).

Standardization is a technique that ensues stratification and uses a set of *reference weights* to control for differences in the vital structure of a population. Standardization has traditionally been the basic technique for identifying and controlling confounding, for identifying effect modification, and for providing a tool that aids the comparison between study and reference populations and to some extent, between study populations.

Stratification and standardization are techniques that are useful not only in demonstrating confounding, but also in controlling its effect. A change in the strength of an association when a potential confounder is controlled indicates the presence of confounding. Stratification indicates confounding when the stratum specific measures of association differ from the

corresponding *overall* or *crude measure*. Standardization suggests confounding when the standardized measure of association differs from the crude measure.

Confounding can also be controlled or prevented at the design stage of a study, through features such as *matching,* restricting the study to homogeneous groups, and random or stratified allocation to study groups. When populations are matched according to variables that are similar in certain respects, these variables cannot have a confounding effect. Matching may be done at an individual level by selecting individuals who are similar in defined respects, or at a group level by ensuring that the groups as a whole are similar in certain respects. Such matching for controlling confounding is utilised in chapter 7 of this thesis, in the construction of risk sets for nested case control analyses of the Radium Hill cohort.

Another analytical method of identifying and controlling for confounding is through *multivariate analysis* using *modelling* techniques. Multivariate statistical modelling techniques are also useful in identifying and describing effect modification. Effect modification may be thought of as a biological interaction (Siemiatycki and Thomas 1981). Therefore, if an interaction between a covariate and an exposure is thought to be biologically plausible, then this effect may be elaborated and examined in terms of a statistical interaction in statistical modelling. This does not however imply, that all covariates presented as interactions in statistical modelling are necessarily effect modifiers. These aspects of effect modification and statistical interactions have been extensively examined in recent decades (Blot and Day 1979; Rothman *et al.* 1980; Saracci 1980; Siemiatycki and Thomas 1981; Pearce 1989).

Further description of stratification, standardization and multivariate modelling methods are presented in sections C.4, C.5, C.6 and in chapter 5, 6 and 7 respectively, in conjunction with their application to the Radium Hill data analysed in this thesis. These sections are preceeded by sections C.2 and C.3 which introduce some of the basic summarization techniques and quantitative measures used in this work and relevant methodological issues.

C.2 Summarization

C.2.1 Basic Measures of Summarization

The fundamental measures of summarizing statistical data consist of *measures of central tendency, measures of dispersion or variation* and *measures of association.* Commonly used *estimators* of these measures include *means, medians, and modes* which estimate central tendency, *ranges, standard deviations and variances* in estimating variation, and *regression and correlation* to evaluate association. In most disciplines these measures are generally based on counts, ratios, proportions or rates.

Fundamental measures of disease frequency used most frequently in epidemiology include *prevalence* and *incidence*. *Prevalence* quantifies the proportion of individuals in a population with a particular disease at a given point of time. Prevalence therefore, provides an estimate of the probability that an individual will be ill at a given point in time, and thus, is also an estimate of *risk*. *Incidence* quantifies the number of *new* cases of disease that develop in a population at risk during a specified period of time. Therefore, incidence provides an estimate of the risk or probability of an individual developing the disease during a specified period of time. Commonly used estimators of incidence include *cumulative incidence* and *incidence rates* or *densities*. These two quantities differ in terms of their denominators or basis of reference and will be described in the context of computations of rates.

C.2.2 Properties of Estimators

The quality of statistical estimators are evaluated on certain desirable properties of "good" estimators, viz., unbiasedness, sufficiency, efficiency and completeness (Mood *et al.* 1974). Of these properties, it is often desired that estimators provide *minimum variance unbiased estimat*es. These statistical notions of *minimum variance* and *unbiasedness* have important corollaries in the epidemiological concepts of *precision* and *accuracy*. The precision of an estimate or measurement depends on the variability seen in its sampling distribution, i.e., the variability observed in its repeatability. The *variance* of an estimate is therefore, found to be a suitable estimator of its precision. Consequently, the property of *minimum variance* could be said to imply *maximum precision*. Accuracy evaluates the degree to which an estimate or measurement represents the true value being measured. The property of *unbiasedness* may therefore be regarded as a useful indicator of accuracy.

The statistical property of *efficiency* in estimators, refers to their being asymptotically minimum variance unbiased estimators. Therefore, the *efficiency* of an estimator could also be regarded as an indicator of precision and accuracy.

Computation of estimators for summarizing disease frequency and cause-effect association pertinent to this thesis is presented in sections AC.2.3 and AC.2.4.

C.2.3 Estimators of Disease Frequency

Since epidemiological data are specific to study populations and involve comparison with other populations, summarization of epidemiological data often entails the use of comparative measures such as *ratios*, *proportions* and *rates*.

Ratios provide the simplest form of comparative measures and are obtained simply by dividing one quantity by another. Ratios may be defined as consisting of a numerator that compares with a denominator without necessarily implying any specific relationship between the numerator and the denominator e.g., number of still birts per thousand live births. Ratio is therefore a general term that includes more specific measures such as proportions, percentages and rates (Hennekens and Buring 1987).

Proportions are a particular form of ratios where the numerator is incorporated in the denominator e.g., the proportion of fetal deaths to the total number of births (live births plus fetal deaths). Proportions are therefore ratios of a part to the whole and are often expressed as percentages.

Rates include an additional dimension of time in its measure. Strictly defined, rates are ratios comprising a distinct relationship between numerator and denominator with a measure of time being an intrinsic part of the denominator.

Rates refer to a wide variety of measures of the frequency of a phenomenon in relation to the size of a population or some other quantity, for a given period of time. In epidemiology, rates have several different usages. Sometimes, used synonymously with ratios, rates refer to proportions e.g., cumulative incidence rate, prevalence rate, survival rate; in some other instances they refer only to ratios representing relative changes in two quantities.

In its most specific usage rates refer only to ratios representing *changes over time*, which do not include measures that refer to observations at a *point of time*; in this restricted usage, the commonly used epdimiological measure of prevalence - prevalence rates - fails to qualify as a "true" rate (Last 1988). Since all these definitions are in accepted usage and widely applied to epidemiological data, any further reference to rates in this work will imply a broad definition in keeping with the specific circumstances of their use.

Rates can be *crude* or *specific*. A crude rate provides an overall measure of the occurence of a phenomenon in a population whereas specific rates are specific to each stratum of the population.

A crude rate is a summary measure in itself. In summarizing stratum specific rates, the essential characteristics of the strata have to be taken into consideration. This is achieved by means of weighting the stratum specific rates with an appropriate factor that is representative of each stratum, before combining them into a summary measure. In addition to providing a suitable measure for summarizing epidemiological data, rates also provide a method of controlling for the effect of population size on the frequency of the event or attribute.

C.2.4 Estimators of Effect

When expressed as probabilities rates provide *risks*, another commonly used measure in epidemiology. Defined as the probability that an event will occur in a defined time period, risks provide a measure of cumulative incidence. The rate of incidence of an event in a population is an estimate of the average risk for its individual members. Cumulative incidence is defined as the proportion of people who become ill in a specified period of time and is based on the ratio of the number of new cases that occurred in the given period of time to the total population at risk during that time period. Cumulative incidence therefore provides an estimate of the probability, or risk, that an individual will develop a disease during a specified period of time.

Risks are of fundamental importance in epidemiology and are capable of multiple expressions. When seen individually, risks provide a tool for summarizing data; when expressed as *risk differences* they provide measures of *absolute effect* and when expressed as ratios, *risk ratios* provide an estimate of *relative effect.*

In addition to providing descriptive measures in epidemiology, rates and risks also provide the basis for computing *measures of association* such as *attributable risks (AR)* and *relative risks* (*RR*), which are also useful in measuring *effect* in comparative studies. Though used to describe several different concepts, attributable risks are based fundamentally on risk differences, and provide an estimate of the size of risk that can be attributed to a given risk factor. Defined as the ratio of risks amongst those exposed and those unexposed to a given risk factor, *RR* provides a measure of the relative risk of exposure to the risk factor.

C.3 Summarization Measures Pertinent to This Work

In epidemiological studies, the risk of cancer is generally measured in terms of incidence rates. When data on incidence is not available, as is the case in historical cohort studies, surrogate measures such as mortality rates must be used (IARC 1990).

The purpose of this work is to evaluate the risk of lung cancer associated with exposure to low cumulative levels of ionizing radiation amongst a specific population of uranium miners. Because this work is based essentially on data from a historical cohort study, the risk of lung cancer will be measured in terms of appropriate lung cancer mortality rates. Lung cancer mortality rates will therefore form the fundamental quantitative measure on which this work is based. The specific representation of rates and rate computation pertaining to this work are outlined in section AC.3.1 below.

C.3.1 Rates and Rate Computation

All rates are ratios multiplied by a suitable factor for convenience of use; *the numerator* indicating the number of events or individuals with a specific characteristic observed in a given period, *the denominator* consisting of a suitable basis of reference for the same period (eg. the size of the population studied), and *the multiplier* generally being a power of 10 which converts the rate from an awkward fraction or decimal to a whole number. The specific representation of these three components of rate computation, viz., the numerator, the denominator and the multiplier, will now be introduced in turn in the context of their primary use in this work. For the purpose of this introduction, references will be made to rates in general and not to *crude rates* or *stratum specific rates*, since the generalistion of the definitions are applicable to both crude and specific rates in the identical manner.

I. The Numerators

The primary outcome being studied in this work is lung cancer mortality. The numerators used in the computation of lung cancer mortality rates will therefore, comprise of the number of deaths from lung cancer.

II. The Denominators

Estimates of the *population at risk* generally form the denominator in the computation of mortality rates. Two commonly used estimators of the population at risk consist of *the number of persons at risk*, and *person-time at risk* which will be discussed in turn.

i. Number of Persons at Risk:

The simplest estimator of the population at risk that may be used in the computation of mortality rates or other rates in epidemiology is based on the number of persons observed in each stratum during the period of observation. Though this estimator has the advantage of computational ease, it is based on the intrinsic assumption that the number of individuals observed during the period considered remains constant; i.e., that all individuals observed at the beginning of the time period considered were followed up for the development of the disease throughout the entire period under consideration. This assumption may be reasonably valid in epidemiologcial studies based on short periods of observation (or stratums based on short calendar periods), or where all individuals start being observed at approximately the same time or at the start of a period of observation and also end being observed at the same time or at the end of a period of observation. The latter may be regarded the case in studies using certain techniques of survival analysis where for computational purposes, observations are considered on a daily basis, reducing each period of observation to its smallest observable unit and thereby regarding period of observation (measured in days) as a continuous variable.

This however, is seldom the case in followup studies where survival time is not the end point of interest. In such studies it is generally found that individuals join the study at different times and also leave the study at different times, both of which may not necessarilly be at the start or end of a period of observation. Furthermore, for computational purposes, the period of observation is not generally startified into very small intervals, and is therefore regarded as a categorical variable rather than as a continuous variable. In such cases an alternative denominator known as the *person years at risk (PYRS)* - a measure of *person-time at risk* - is found to be more appropriate. In the first analyses performed and published on the Radium Hill study (Woodward *et al.* 1991), I used the number of persons at risk denominator. However, all analyses thereafter were refined by using more appropriate person years at risk denominators.

ii. Person Time at Risk:

The computation of person-time at risk is based on first determining for each individual the amount of observation time contributed to a given stratum (eg., a given age - calender period category), and then summing up these contributions for all cohort members so as to obtain the total person-time of observation in that stratum. The denominator thus computed has the advantage that it retains all the essential information available regarding each individuals own period of followup unlike in the previous measure described.

A more precise method of computing mortality rates is therefore derived from the usual numerator which comprises simply of the number of deaths due to a given disease in the specified stratum with the corresponding person-time at risk contributed forming the denominator.

III. The Multiplier

The multiplier is a factor that is used for merely for convenience of use in interpreting rates. Multipliers generally depend on the dimension of the ratio formed by the rate numerator and denominator, and are chosen so as to make the computed rate an integer. Multipliers commonly used in the computation of mortality rates include 10^5 and 10^6.

C.3.2 Crude Measures of The Frequency of Lung Cancer

The fundamental measures of summarizing disease frequency with a view to evaluating risk in terms of a disease such as lung cancer comprise measures of *incidence*. Commonly used estimators of incidence include *cumulative incidence* and *incidence rates* or *densities*. These two quantities basically differ in terms of their denominators or basis of reference. *Cumulative incidence* is defined as the proportion of people who become diseased during a specified period of time. Cumulative incidence is based on a numerator that comprises the number of new cases of a disease during a given period of time and a denominator that comprises the number of persons at risk. Cumulative incidence therefore gives, an estimator of the probability or risk that an individual will develop a disease during a specified time period. The *Incidence rate* is based on the same numerator as the cumulative incidence but uses a person-time denominator and provides a more precise measure of the *force of morbidity* or *mortality*.

The outcome of interest in this work is lung cancer, and the only available data pertaining to this outcome relates to mortality particulars. In the absence of data on the onset of the disease, it is not possible to measure the *incidence* of lung cancer. This work is therefore based on *lung cancer mortality*, which will be used as a *surrogate for lung cancer incidence*, in measuring disease frequency and also as a basis for *measuring effect*. Consequently, measures of lung cancer mortality, measured by lung cancer mortality rates, will form the basic quantitative measure used in this work.

In using mortality rates as surrogates for incidence rates it seems appropriate the two parallel estimators of incidence defined above be introduced. In this work the terms *cumulative mortality rate* will be used as a corollary to the *cumulative incidence rate,* and *mortality rate* as analogous to *incidence rate.* In terms of the denominators described in section AC.3.1, the essential difference between these two surrogate estimators of incidence is that computationally, the cumulative mortality rate is based on on a denominator that comprises the *number of persons at risk* whilst, incidence rate is based on a *person-time* denominator.

The *number of person-years at risk (PYRS)* will be used to as an estimator of *person-time at risk* in this work. Though the *PYRS* provide a more precise estimator of the denominator, its computation depends on the nature of the data collected. In practice it is often found that *PYRS* are based on approximated dates of commencement and termination of exposure and followup. It is also possible that in certain cases the opportunity cost involved in gathering detailed data on followup dates and additional computational costs of computing *PYRS* could outweigh the gain in precision of the resulting rates and risks computed. It may therefore be of value to examine the differences between these two alternatives and evaluate the degree of improvement in the precision of the estimates of rates and risks computed.

C.3.3 Crude Measures of The Effects of Radon Exposure

The basic measures of effect used in this work comprise *rate differences (RD)* and *rate ratios* which estimate *relative risks (RR).*

The sampling distribution of RD could reasonably be expected to be symmetric. Approximate confidence intervals for RD can therefore be based on the standard normal deviate as:

95% CI. For Rate Differences: $\hat{RD} \pm 1.96 SD(\hat{RD})$

The sampling distribution of RR however, is not symmetrical and is known to be skewed. This is because estimates of RR can range from zero to infinity whilst, being more frequent in the lower regions of this range than at the higher regions and *random errors* can lead to larger discrepancies on the high side of the mean. Consequently the sampling distribution of RR will not be *normally distributed* unless it is based on a relatively large number of observations. Change of scale transformations of RR which provides a better approximation to the *normal distribution,* and are therefore used in hypothesis testing when test statistics are based on normal deviates. More accurate confidence intervals for RR are generally obtained after performing a *log transformation* to adjust for skewness, and confidence limits are then extracted through a antilogarithmic transformation. Confidence limits for RR may thus, be represented by:

95% CI. For Rate Ratios: $antilog\{ln(\hat{RR}) \pm 1.96 \mathrm{SD}[ln(\hat{RR})]\}$

Exact confidence intervals and tests of hypotheses based on *counting distributions* such as the *binomial* and *hypergeometric distributions* are also available for these measures of effect (Rothman 1986). These exact methods will not be used in this work and therefore, will not be discussed here.

C.4. Stratification

C.4.1 The Role of Stratification

When rates are compared between populations that vary in their fundamental vital structure (e.g., distribution of age, gender, social status), *overall* or *crude rates* do not provide an accurate means of comparison. A comparison of crude rates controls for the effect of the population size only; not for other specific factors. Therefore, crude rates are of little use for purposes of comparison in the presence of possible confounders and effect modifiers.

Stratification is the most basic tool available for identifying confounding and effect modification, describing effect modification when it exists and in the absence of effect modification, adjusting or controlling for confounding effects. The primary objective of stratification could therefore be identified as its role in dealing with confounding and effect modification.

This section concentrates on the specific issues of the role of stratification in addressing confounding and effect modification. In each case, in addition to describing the relevance and implication of stratification, methods of interpreting stratified data will be presented. Some methods of summarizing stratified data in the absence of effect modifiers will then be discussed; the methods will consist only of the simplest and most basic techniques which are *independent of external references*. Methods of summarization under stratification based on *external reference populations - standardization* - will be presented in sections AC.5. This section concludes with a summary of the statistical implications of stratification, and the relative efficacy of stratification options and summarization techniques. The practical implications of these aspects are evaluated using the Radium Hill dataset in chapter 5 of this work.

C.4.2 The Specific Relevance of Stratification to This Work

In some applications, particularly when the observation period is relatively short, the calendar-year is ignored and the rates are determined by age interval alone (Breslow and Day 1987). However, when the period of observation is extensive, it is important that the essential characteristics of age and calender-period be retained by the data and captured in the analysis. This too is achieved by stratification; stratification of the data according to age and calendar-period. This approach has been in controlling for confounding through out this thesis.

Age at observation is considered the fundamental *stratifier* in this work. However, the nature of the primary data-set being analysed, perticularly since it pertains to an extensive calendar period of followup, indicates the need for controlling simultaneously for two stratifiers, age and calendar period of observation.

In addition to stratification by the two basic variables considered above viz., age and calendar-period, stratified analyes will also be used for the preliminary examination of the role of other factors such as age at first exposure, time since cessation of exposure, cumulative exposure, intensity of exposure and duration of exposure, all of which play a potentially important role in radiation carcinogenisis.

C.4.3 The Role of Stratification in Confounding and Effect Modification

Stratification of the data according to the suitable categories of potential confounders controls for the effect of these potential confounders. The aim of stratification is to group the data into ranges that are narrow enough to be regarded as homogeneous categories of the confounder. The resulting lack of variability within each stratum of the confounding variable leads to each stratum providing and *unconfounded* estimate of the effect. Furthermore, *if* the estimates of effect *between-strata* are homogeneous (i.e. the stratum-specific effects are uniform across all strata), these estimates may be combined to provide an overall unconfounded estimate of effect over the entire range of the confounder. This however, is seldom the case in practice, due to the role of random error in the sampling distribution which reflects the precision of each stratum-specific point estimate. Therefore, even in the case of apparent homogeniety between stratum-specific estimates, summarization across strata has to be preceeded by an examination of the role of chance by estimating the random error.

Effect modification is signified by heterogeneity in stratum-specific measures, whilst uniform stratum-specific measures that vary from the crude overall measure signify confounding (the essence of confounding being the disparity betwwen crude and adjsted rates). If effect modification exists, some argue one should not combinine stratum-specific measures to derive an overall summary measure of disease frequency or effect. Heterogeneity in stratum-specific measures may arise not only as a result of extraneous effects such as confounding and effect modification, but also, as a result of random variation. Therefore, before confirming and describing or controlling for extraneous effects, it is once again necessary to quantify the random error.

Thus, the logical steps in the analysis of data are firstly, the identification, evaluation and control of potential confounders; secondly, ascertaining the homogeneity of stratum-specific estimates of effect by evaluating and describing effect modification if it exists; and finally, in the absence of effect modification, the summarization of stratum-specific estimates of effect.

C.4.4 The Identification, Evaluation and Control of Confounding Through Stratification

The three basic prerequisites for a factor to qualify as a confounder are: firstly, it must be a risk factor for the disease amongst the unexposed; secondly, it must be associated with the exposure being studied; and thirdly, it must not be a intermediate link in the cause-effect chain. Whilst adherence to the first two of these conditions can be confirmed quantitatively, confirmation of the third requisite has to be based on a knowledge of the disease and the process of causality.

One way of identifying a confounder is to examine each of these requisities in turn and confirm their validity. Stratification plays a useful role in validating the first two of conditions. The potential confounder can be identified as a risk factor for the disease amongst the unexposed if the risk amongst unexposed varies over stratum, ie., if the proportion of cases unexposed is not constant in each stratum. Its association with the exposure can be confirmed by variability in the stratum-specific risks of exposure amongst nondiseased.

A simpler and more direct way of identifying cofounding is to compare a pooled estimate of the unconfounded stratum-specific rates with the overall or crude rate. Any difference between these measures indicates the existence of a confounding effect and the degree of discrepency estimates the magnitude of the confounding effect.

Although, it is possible to detect the existence of confounding by examining whether the potential confounder is associated with the disease amongst the exposed and with exposure amongst nondiseased, this method does not provide a means of examining the magnitude of the confounding effect. The method of comparing the *crude* (unadjusted) estimates with the *unconfounded* (adjusted) estimates is thus preferred in assessing confounding because, it provides a clear and unambiguous indication not only of the presence but also the magnitude

of the confounding effect. Moreover, this method can easily be generalized to incorporate simultaneous control for multiple factors (Rothman 1986).

It is important to note that confirmation of confounding should not depend solely on quantitative reslts. Prior knowledge of the disease and its determinants should play the fundamental role in this process. Quantitative findings from the above mentioned methods of assessment should only complement such prior knowledge.

C.4.5 Summarization Under Stratification in the Absence of Effect Modification

Although stratification helps identify and control for the effect of potential confounders, the interpretation of a number of stratum-specific rates, unless they are identical, is often clumsy and difficult. If the stratum-specific rates are identical it is obvious that they represent the unconfounded estimate of the effect of exposure on the disease. This however, is seldom the case. Furthermore, even if estimates are identical across strata, they may vary within strata due to random error. It is therefore necessary that these stratified measures be combined into a single summary measure through a suitable method of pooling that incorporates the essence of the stratification and the inherent variability in each stratum-specific measure.

The basic methods of deriving overall estimates that encompass the essence of stratification and its inherent variability are *the method of inverse variance weighting, the method of maximum likelihood (ML),* and *the Mantel-Haenszel (M-H) method.*

The derivation of an overall summary estimate of disease frequency or effect for stratified data is typically based on a weighted average of the stratum-specific estimates. The aim of stratification is to produce strata that can be regarded as homogenous categories of the stratifier. Appropriate stratification should therefore, produce uniform parameter estimates

which should theoretically be estimates of the same summary parameter. Consequently, any variability observed in these stratum-specific measures can be regarded as representing the random variation within strata. It is therefore desirable that the method of pooling these values should give greater weight to more precise estimates ie., ones with smaller random variability and lesser weight to less precise estimates which have greater variability. Theoretically, this is best achieved by the inverse variance weighting procedure.

The method of inverse variance weighting provides the optimum procedure for reducing the variance of the summary measure, thereby enhancing its precision. An inverse variance weighted summary measure is therefore the weighted average of the stratum-specific measures, where the weights comprise the inverse of the stratum-specific variances. This method therefore requires a knowledge of stratum-specific measures and their variances. Though it may be an optimum procedure for enhancing the precision of the summary measure, summary estimates obtained by *the method of inverse variance weighting* are influenced by small numbers of observations in stratum, thus, affecting the validity of the summary measure. Furthermore, if there are cells with zero observed frequencies, the corresponding stratum-specific weights will be zero resulting in the variance estimates being infinity. Even though the problem of zero weights may be mitigated by adding a small constant correction factor such as 0.5 or 1.0 to zero cell frequencies or to each observed frequency (Haldane 1955), it does not overcome the inaccuracy arising in stratum-specific variance estimates due to small cell frequencies (Rothman 1986).

The *ML* method is an iterative method of obtaining estimators that are *sufficient* - it uses all the information (relevant to the estimation of the parameter) that is contained in the data (Freund 1971) - whenever they exist, and *asymptotically minimum variance unbiased.* In principle, the *ML* method consists of selecting the value of the parameter being estimated for which the probability of obtaining the sample values is a maximum ie., the value of the parameter that maximized the *likelihood* of the particular sample being drawn. The term *likelihood* is mathematically defined as the joint distribution of the random variables as a

function of the parameter being estimated (Mood *et al.* 1974) a *Maximum Likelihood Estimate (MLE)* is the value of the parameter that maximizes this *likelihood function.* The likelihood function is written as a function of a uniform effect measure, the observationas and whatever nuisance parameters may be involved. The *MLE* is then obtained by setting the derivative of the *log-likelihood* to zero.

The *ML* procedure is preferable when there are some cells with a small number of observed frequencies. Rather than treating each cell in isolation, the *ML* approach *'adjusts'* the observations in each stratum in a way that integrates the information among all strata (Rothman 1986). The computational complexity of the *ML* method has acted as the major deterrant to its use until the availability of powerful desktop computing facilities.

The *M-H* procedure suggested by Mantel and Haenszel in 1959 (Mantel and Haenszel 1959), made a profound contribution to analytical epdidemiology. The *M-H* method provides a pooled estimator based on explicit weights that are incorporated in its compuatational formulae . The *M-H* procedure is often favoured as a method of pooling because of its computational simplicity and becuase it has statistical properties as nearly as good as the *ML* estimators (Rothman 1986; Breslow 1984).

The computations facilities available to this work, provide adequate capabilities for the use of *ML* methods. In view of the desirable properties of *MLE* identified above, only methods of *ML* estimation will be used in the analytical component of this work.

C.4.6 Identification, Evaluation and Description of Effect Modification Through Stratification

Effect modification is identified by the presence of heterogeneity over and above that attributable to random error in the stratum-specific measures of effect.

Effect modification may conveniently be evaluated with the use of statistical tests for homogeneity of stratum-specific effects. Conclusions should not however be drawn from the mechanical application of these statistical tests. The more general statistical tests, having low power against specific alternatives may leave non-significant *p-values* which are difficult to interpret correctly. These issues are addressed in further detail in chapter 7, in the context of their application to the analysis of the Radium Hill data where effect modification is represented by interaction terms in statistical modelling and tests of statistical significance for the interaction effect are used to examine the heterogeneity in the exposure-response relationship with variations in the potential effect modifier. It is noted here that: the notion of statistical significance is adopted as a means of evaluating the role of chance variations rather than as strict criteria for identifying effect modification; in examining effect modification, the extent of such chance variation need not as high as the commonly used 5% significance level; even significance levels of 10% or 15% could be used as indicators to invoke interest in further examination and elaboration of potential modifying effects.

In this thesis, the examination of each effect modifier commences with the categorical examination of main effects and interactions to demonstrate main effects and effect modification; interaction effect are examined after removing the main effects of the interacting factors as a means of assessing effect modification (Pearce 1989). Thereafter, various modelling approaches suggested by Preston *et al.* using EPICURE (EPICURE 1992) as outlined in appendix B will be used to examine effect modification and the most appropriate functional form of the interactions.

C.5 Standardization

C.5.1 Standardization Methods

Standardization provides another means of summarization after controlling for the effect of potential confounders. The process of standardization aims at producing an easily interpretable summary measure that retains most of the essential information contained in the data that it represents, with a view to facilitating the comparison of populations that have been similarly standardized.

The *standardization of rates* has traditionally been the basic method of facilitating comparisons in epidemiology. In this context, standardization provides a set of techniques that attempt to remove or reduce the effect of differences present in the vital structure of the groups under comparison. Standardization therefore attempts to eliminate the distortions in summary rates arising solely by virtue of differences in the vital structure of the populations being compared. Consequently, when comparing rates that have been standardized to adjusted for a particular factor, any remaining difference observed between the groups cannot be attributed to the confounding by that factor.

Standardized rates are essentially weighted averages based on a common set of reference weights, the idea being that the use of the same weights in the study populations under cosideration, provides a fairer basis of comparison. The terms *target population* is used synonymously with *study population*, and *standard population* with, *reference population*. A standardized rate can be interpreted as the hypothetical crude rate that would exist if the target population had the same distribution as the reference population with respect to the factor being controlled.

The methods of *direct* and *indirect* standardization have been of perennial importance in epidemiology. Though new and improved computational facilities have now made more effective measures of estimation available, the use of standardized rates are still quite valid as measures of summarization in epidemiology. Direct and indirect measures of standardization are also widely applied in the comparative appraisal of current epidemiological findings with previous research findings; it is in this context that both direct and indirect methods of standardization will be used in this work. These methods of standardization are described in the remainder of this section. Details of the computation and comparison of standardized measures are included in the relevant sections of the main text of this thesis (chapter 5).

C.5.2 Standardization Techniques

I. Direct Standardization

The method of *direct standardization* refers to the application of target rates to a standard distribution. In direct standardization, stratum-specific rates in a target population are averaged using the distribution of a specified standard population as weights. A *direct standardized rate* (*DSR*, P_D) is therefore, a weighted average of target rates using standard or reference weights.

The *comparative mortality figure* (*CMF*), is another measure of summarization based on the method of direct standardization. The *CMF* provides a simple summary of rate ratios between target and standard populations after controlling for possible confounders. Defined as the ratio of the *DSR* and the standard population rate, the *CMF* is also useful in comparing target populations that are directly standardized using the same standard populations.

It should be noted that though the precision of the *DSR* decreases, the bias also decreases as the strata narrow. Direct standardization is therefore, not appropriate for broad strata.

II. Indirect Standardization

The method of *indirect standardization* refers to the application of standard rates to a target population.

Indirect standardardization is based on a comparison of the total number of occurrences observed in the target population with the number expected in that population if standard population rates applied. The ratio of total number of ccurrences observed (O) to those expected (E), is defined as the *standardized mortality ratio* (*SMR*).

An *indirect standardized rate* (*ISR*, P_I) is defined as the product of the *SMR* and the *crude population rate* (P_C). Indirect standardization is often used when stratum-specific target population rates are not known.

In using *SMR*s for comparative purposes, it should be noted that the indirect method of standardization does not adjust stratum specific rates in the populations under comparison to the distribution of the reference population since it is the *standard* population rates that are applied to the *target* populations. In this light it can be seen that the method of indirect standardization is analogous to direct standardization, it being simply that the target population contributing to the *SMR* is now used as the standard. The comparison of *SMRs* amounts to comparing measures with different standards and is therefore invalid, even if the same standard population is used in their computation.

Despite the comparison of SMRs being *'invalid'* such methods of comparison are widely used and relevant issues have been discussed and debated for more than 60 years (Wolfenden 1923; Yule 1934); these arguments have been comprehensively summarized by Rothman (Rothman 1986). Therefore, further discussion of this perennial issue will be avoided in this work.

C.6 Overview on Statistical Modelling

C.6.1 Introduction

The primary aim of this work is to evaluate the risk of lung cancer mortality related to *Rn* exposure amongst the cohort of uranium miners studied, after controlling for the effects of *potential confounders*, and demonstrating the role of any *effect modifiers* that may exist. Methods of risk evaluation previously outlined are based on the implicit assumption that there is no interaction between the outcome measured and levels of the covariates. They are therefore, not appropriate in the presence of effect modification.

Modelling is a tool that enables risk evaluation whilst allowing for the *simultaneous control of multiple confounders*, and the identification and demonstration of *effect modification*.

In *modelling* terminology, the outcome studied, which in this case is lung cancer mortality, is referred to as the *dependent* or *outcome variable*; potential confounders, effect modifiers, causative factors and other explanatory variables are referred to as either *covariates,* or *predictor variables*. *Modelling* deals with the problem of *confounding* through the assessment of *heterogeneity* in covariate stratum-specific estimates of risk, and addresses the role of *effect modification* by considering *interactions* between the *covariates*.

C.6.2 Statistical Modelling

The purpose of *modelling* is to study the *relationship* between variables. The relationship between variables can be defined by the *nature and the degree of their association* which are expressed by the statistical concepts of *regression* and *correlation*. *Regression* is the statistical tool for describing the nature of the association between variables and is represented in the form of *models*. *Correlation* which evaluates the degree of the association between variables, is measured by coefficients that are derived from parameter estimates obtained through regression modelling.

Statistical modelling is a quantitative technique for studying the relationship between patterns of variation observed in outcomes of interest, and known predictors of this outcome. *Statistical modelling* is based on the underlying notion of correlating patterns of variation observed in variables with a view to attributing components of the variation seen in the outcome variables, to known and unknown predictors. Statistical modelling can therefore, be described as a process of separating out the components of variation that can be explained by known predictors - *explained variation* or *signal* - from the component the total variability seen in the observed outcome; the component that remains unexplained is referred to as the *unexplained variation* or *noise*. The separation of components of variation is performed by the use of regression models comprising two components: one that represents the *signal* (the *deterministic component*), and the other the *noise* (the *stochastic component*); to present a simplified or smoothed representation of the underlying population. The *deterministic component* of a *statistical model* is represented by a mathematical depiction of the relationship between *outcome* and *predictor variables* and is therefore, referred to as a *mathematical model*. The *stochastic component* of a *statistical model* is the component of variation that cannot be explained by the predictors in the model. Being non deterministic, it is also termed the *random error* or *random variation* component of the statistical model. The aim of statistical modelling can therefore be described as extracting from the data as much information as possible about the signal as it is defined by the model.

The art of modelling may be described in three stages viz., *model specification, parameter estimation* and *model evaluation.* The modelling process commences with model specification: postulating a functional relationship between the outcome variables and the predictor variables which is based on unknown parameters i.e., specifying the *mathematical model,* which can be based on postulated etiological mechanisms of exposure; and postulating a functional form or distribution for the outcome variable which corresponds to postulating the *error function.* This is followed by parameter estimation: the estimation of the unknown parameters in the postulated model and obtaining risk estimates; parameter estimation is generally based on methods of *least squares estimation* or *maximum likelihood estimation.* Finally, the appropriateness of the model is examined in *model evaluation* by methods of assessing the *goodness-of-fit* of the model; goodness-of-fit basically implies the comparison of components explained and unexplained by the specified model and methods of drawing inferences on how substantial a component has been explained; these techniques include the *analyses of variance* and *analyses of deviance.* Model evaluation also includes *model criticism,* the evaluation of the appropriateness of the model and examination of the assumptions governing its use for the existence of violations. The modelling process can be regarded as one of *conjecture* and *refutation*, and is often iterative with the process being repeated until a suitable *'fit'* is obtained.

Models most commonly used in statistical evaluation belong to the family of *linear models.* The simplest linear models are those based on linear combinations of predictors which reflect the situation where the effect is proportional to the cause. Models that are nonlinear in their predictors include quadratic and linear-quadratic models which are commonly used in etiological assessment. In a strict statistical sense, linear models refer to the family of models that are linear in its parameters. Linear models may therefore, include functions of predictors that are not necessarily linear. Being intrinsically linear in their parameters, these models too are handled with linear modelling techniques.

The most comprehensive methodology for statistical evaluation of linear models is based on the theory of *generalized linear models*. Generalized linear models are characterized by *link* and *error functions*. The *link function represents the deterministic component of the model* and the *error function* describes the stochastic component. General linear models are a special case of generalized linear models where the *error* function is assumed to be *normal* and the *link function* is the identity function; whereas, in generalized linear models the error function may be from an exponential family of distributions (not necessarily from a normal distribution) and the link function may be any monotonic differentiable function (McCullagh and Nelder 1989).

Apart from its uses in describing the patterns of mortality as postulated by the model, mathematical modelling also provides summary estimates of risk which are vital for comparative purposes. *Mathematical modelling* provides a technique for obtaining *estimates of summary relative risks* whilst addressing the issues of *confounding* and *effect-modification*. This is unlike the summary measures of risk obtained from summarization techniques proposed in previous sections of this appendix which are based on the implicit assumption that there is no effect modification present; i.e., that there is no interaction between the outcome variable and the levels of the covariates.

In addition to providing summary estimates of risk, *modelling* enables the identification and demonstration of effect modification by incorporating a mathematical description of modifying effects as interactions in the estimation process; modelling enables the evaluation of confounding by examining the heterogeneity of covariate stratum specific estimates, once other prerequisites for confounding have been satisified.

A scientific approach to statistical modelling in epidemiology is an iterative approach of fitting statistical models inferred from postulated biological mechanisms, examining how well they fit the data (Murphy 1978) and using both inductive and deductive reasoning (Jacobsen 1976) in forming plausible conclusions (Siemiatycki and Thomas 1981).

Statistical modelling techniques most commonly used for the evaluation of risk in epidemiological studies include *Poisson* and *Cox regression* techniques. Both these techniques are based on models that belong to the family of *generalized linear models* (Dobson 1983; McClaugh and Nelder 1989; Aitken *et al.* 1989). The application of Poisson and Cox regression techniques have been widely discussed in the literature in the past decade (Frome 1983; Frome and Checkoway 1985; Pearce and Checkoway 1987; Cox 1972; Breslow and Day 1980 and 1987; Breslow *et al.* 1983); therefore, detailed description is avoided in this work.

Since Poisson regression techniques will be widely used in this work (chapters 5 and 6), a simple introduction to *Poisson regression methods* within the context of statistical modelling is provided in this appendix. Relevant introductions to Cox regression techniques are included in the main text of this work in chapter 7, in conjunction with its use.

I. An Overview on Poisson Regression Modelling

The *Poisson distribution* is known to be suitable for describing the *probabilities of occurrence of rare events*. The stratum-specific probalities of lung cancer mortality in this study cohort may be considered *small* because of the small number of lung deaths in each stratum. Therefore, it is reasonable to assume that lung cancer mortality in the Radium Hill study cohort follows a *Poisson distribution.*

Poisson regression methods are based on the assumption that the *outcome variable* follows a *Poisson distribution*, and that the functional relationship between the predictors and the outcome being observed is known. In Poisson regression modelling a *rate function* that postulates the relationship between the outcome and predictors variables comprises the *deterministic component* of the statistical model. The stochastic component of the statistical model is represented by a *Poisson error* term.

Parameter estimation under the Poisson model depends on the specific functional form postulated to describe the relationship between the outcome and predictor variables. The specific functional forms between outcome and predictor variables can be based on postulated etiological mechanisms of exposure. Estimation is then performed by using generalized linear modelling techniques that specify a *Poisson error* and a suitable *link function* (generally a *logarithmic link function)* that is defined by the functional relationship postulated between the outcome and predictor variables. Parameter estimates are obtained from the generalized linear model through maximum likelihood estimation techniques based on the iterative solution of the likelihood function.

II. Addressing Temporal Effects of Lung Cancer Risk Through Statistical Modelling

Occupational exposure to radon daughters amongst underground uranium miners generally occurs over an extended period of time, with varying intensity. In the absence of a detailed understanding of the biological repair mechanisms concerning exposures experienced over a period of time it is important that the characteristics of exposure delivery be considered in analyses. If *fractionation* allows adequate time for damage repair, then cumulative exposure is not an adequate or proper measure of exposure.

Furthermore, since individual susceptibility to the nature and amount of exposure may also vary considerably between persons, estimates of risk based on the compounded risk/survival experience of individuals would provide more precise estimates of risk than those averaged over individuals at particular points in time.

Statistical methods for the analyses of data based on these suggested approaches are also different. Traditional approaches such as standardization generally follow the former approach. The advent of computing facilities makes it possible to assess the survival experience of each individual over time using extensions to Cox's approach (Cox 1972). Temporal effects can be addressed by using alternative time scales reflecting elapsed time from different origins or by addressing changing exposures with time. Though these two types of variables pose similar analytical problems their interpretations differ.

Age at first exposure, time since last exposure, the duration of exposure and the intensity of exposure are *temporal factors* - potential determinants or modifiers of disease risk which vary as a person ages - known to influence the risk of radon related lung cancer mortality. The modifying effects of these temporal factors are examined in chapters 6 and 7 of this thesis.

C.7 Summary

Statistical modelling provides summary estimates of risk which are vital for comparative purposes. *Mathematical modelling* provides a technique for obtaining *estimates of summary relative risks* whilst addressing the issues of *confounding* and *effect-modification*. This is unlike the summary measures of risk obtained from summarization techniques proposed in previous chapters of this work which are based on the implicit assumption that there is no effect modification present; i.e., that there is no interaction between the outcome variable and the levels of the covariates.

Analytical methods introduced in chapter 5 are based on stratification and standardization as means of identifying and controlling confounding; stratum-specific estimates of risk thus obtained are regarded as *risk estimates adjusted for the potentially confounding effects of the stratifiers*. In comparative epidemiological analyses, the interpretation of several covariate stratum-specific risk estimates obtained from stratified analyses, is often found to be cumbersome. Comparisions are therefore based on overall estimates of relative risks or risk ratios that are obtained by summarizing the covariate stratum-specific risk estimates, a procedure that is only appropriate if the component measures (stratum-specific risk estimates) are homogeneous across strata.

Furthermore, the summarization techniques for obtaining overall risk estimates presented in chapter 5 of this work are based on the implicit assumption that there is no effect modification present; i.e., that there is no interaction between the outcome variable and the levels of the covariates. therefore, the comparison of summary measures obtained through these methods is strictly valid, only in the absence of effect modification.

APPENDIX D

D. Background on Modelling With EPICURE

D. Background on Modelling With EPICURE

This appendix on modelling with EPICURE has been prepared from material presented in the EPICURE Users Guide (EPICURE 1992). The purpose of this appendix is to enable readers who may be unfamiliar with this particular package to obtain adequate familiarity with modelling in EPICURE so that the reading of the text of this thesis may be independent of reference to external material.

D.1 General Representation of Models In EPICURE

All EPICURE models have a general form known as the Product Additive Excess Risk Model, in which mortality rate λ can be defined as:

$$\lambda = \lambda_s T_0 \left[1 + \sum T_i\right]$$

Where:
- λ - Lung cancer mortality rate
- λ_s - Represents a *stratum parameter*
- T_0, T_i - Represent *terms*,
- T_0 - Represents the *background risk* term, and
- T_i - Represents the *excess risk* term.

Stratified models include a separate multiplicative parameter λ_s for each combination of values of the stratification variables.

Each *term* may contain *subterms* which can be of three types viz., *linear* ($\beta' x$), *log-linear* ($\alpha' y$) and *product-linear* ($\gamma' z$).

Each term may therefore be defined more completely as:

$$T_i = (\beta' x)\, e^{\alpha' y}\, (1 + \gamma' z).$$

A term can have only 1 subterm of each type.

The terminology *product additive excess risk model* is coined from the components of the model - product, referring to T_0 the background risk term which multiplies all other terms, and additive excess referring to 1 which is added to the excess risk terms ΣT_i.

Special cases of this general model include the product additive model, the additive model and the geometric mixture model. In the absence of the constant term 1, the general model reduces to the Product Additive Model. In the further absence of the background term, the general model reduces to the Additive Model. The geometric mixture model is a more complex model where the user defines a *mixing parameter* that cannot be estimated by the program. These models are defined as follows:

Product Additive Model: $\lambda(t,d) = \lambda_s T_0 \Sigma T_i$

Additive Model: $\lambda(t,d) = \lambda_s \Sigma T_i$

Geometric Mixture Model: $\lambda(t,d) = \lambda_s T_0 (\Pi T_i^*)^\theta (1 + \Sigma (1 - T_i^*)^{1-\theta}$

Where, T_i^* is a modified term specified by the user.

T_i^* is either a *Relative Risk* (RR) term (T_i) or an *Excess Risk* (ERR) term ($1+T_i$).

Unless otherwise specified, T_i^* are considered as RR terms by default.

θ is a *mixing parameter* which must be specified by the user,

and cannot be estimated by the program.

D.2 Regression Models for Evaluating Radon Progeny Associated Risk of Lung Cancer Using EPICURE

In analyses based on estimates of cumulative exposure, a RR model for lung cancer mortality rate $\lambda(x,z,w)$ may be defined as:

$$\lambda(x,z,w) = \lambda_0(x)RR(z,w)$$

where λ - lung cancer mortality rate

w - cumulative exposure in WLM

x - vector of covariates describing the background disease rate

z - vector of covariates affecting the dose response relationship.

Under the assumption of a RR model, λ can be written as a product of the background disease rate among non-exposed, denoted by $\lambda_0(x)$, and an exposure function $RR(z,w)$. The background rate depends on x and the exposure-response function depending z, which may include one or more components of x as z.

The background rate can be defined as and exponential function of covariates x, denoted by:

$$\lambda_0(x) = e^{\alpha x}$$

In our analyses x may comprise indicator variables for age group, calendar year of observation and other potential confounders such as other occupational exposure to radiation or asbestos.

A general model for the assessment for a broad range of exposure-response relationships may be defined as:

General Relative Risk Model: $RR = [1+\beta w^{\kappa}]\, e^{\theta}$

where, β estimates the ERR per WLM,

θ denotes a "*cell killing*" parameter, and

κ denotes a "*non-linear*" parameter.

Particular cases of this general RR model which are commonly used in risk estimation include the following:

The Linear* ERR *model: $RR = 1 + \beta w$

when, $\kappa = 1$ and $\theta = 0$ in the General RR Model.

RR is modelled as a linear function of dose.

β measures the ERR per unit dose increase, i.e. the ERR/WLM.

The Linear-Exponential Cell Killing Model: $RR = [1+\beta w]\, e^{\theta}$

when, $\kappa = 1$ in the General Model.

The Non-Linear Model: $RR = [1+\beta w^{\kappa}]$

when, $\theta = 0$ in the General Model.

Tests for goodness of fit in relation to the parameters θ and κ are based on likelihood ratio tests.

D.3 Identification and Control of Confounders Using EPICURE

Confounders may be identified and controlled through stratification in EPICURE. Potential confounders may be controlled for, by being specified as STRATA which then correspond to stratum variables - the T_0 term - in the general form of the epicure model. Estimates of rates and risk obtained after confounders have been declared as stratum variables may be considered as estimates adjusted for the effect of confounders. Any difference that may be seen between crude estimates and estimates adjusted for confounders, is another indication of confounding.

D.4 Hypotheses Testing in EPICURE

Analysis of Deviance provide a means of testing goodness of fit for all models in generalized linear modelling. Analysis of deviance is the corollary to analyses of variance (ANOVA) in general linear modelling. Generalized linear modelling is based on maximum likelihood methods where all models are evaluated against the fully saturated model. The deviance is defined as twice the difference in the log-likelihoods of the model being tested and the fully saturated model.

Analyses of Deviance also provide comparative tests of goodness-of-fit in nested models. Nested models may be defined as follows: "*two models are said to be nested if the free parameters in the more restricted model are a subset of the free parameters in the less restricted model*" (EPICURE 1992).

In comparing nested models, LRT provide tests of effect and homogeneity of stratum specific estimates. Likelihood Ratio Tests are based on LR statistics which are computed as the difference between the deviances of the models being compared. Likelihood Ratio statistics are asymptotically distributed as $\chi 2$ distributions.

Tests for hypotheses of trends are based on two types of tests, viz., score tests and Wald tests. In EPICURE score tests are performed by constraining the variable being tested for trend to a null value and the score statistic is tested as a standard normal deviate.

Wald tests are performed by modelling the variable being tested for trend as an unconstrained continuous variable in EPICURE and obtaining estimates of the slope parameter and its standard error. The Wald test statistic is defined as the ratio of the parameter estimate to its standard error is tested as a normal deviate.

D.5 Models for Identifying Effect Modification

The role of potential effect modifiers may be examined by evaluating their interaction with the exposure-response relationship. When effect modification exists, β is seen to vary within categories of the modifying factor. Effect modifiers of interest in the study of Rn induced lung cancer include age at first exposure, duration of exposure and rate of exposure. Variations in risk with attained age and time since last exposure are also of particular interest.

In the general RR model defined above the term z comprises potential effect modifiers. Effect modification is tested using a LRT based on the comparison of a model that fits a common parameter and separate parameters for each category of the potential effect modifier. i.e. by testing the difference in deviances between the following model and the model given previously.

Modelling Modifiers of Effect: $RR = 1 + \beta_j w z_j.$

Where, β_j represents the ERR/WLM within category z_j.

Under the null hypothesis of no effect modification the difference in the model deviances had a X^2 distribution with $j-1$ degrees of freedom. A significant p-value indicated that the effect of Rn exposure on lung cancer mortality was not homogeneous across levels of z.

Variations of the ERR/WLM over z_j were modeled using a smooth parametric function based on dose as a continuous variable. Variations of the RR estimates over z_j were examined with exposure treated as a categorical variable. Categorical estimates of RR of exposure were also estimated after accounting for the effect of z_j.