

**A COMPARISON OF LONGITUDINAL STATISTICAL METHODS IN
STUDIES OF PULMONARY FUNCTION DECLINE**

by

HELEN D. DIMICH-WARD

M.Sc.(Kinesiology), S.F.U.

**A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

in

THE FACULTY OF GRADUATE STUDIES

Interdisciplinary Studies (Epidemiology/Medicine/Mathematics/Physiology)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

February 1991

© Helen D. Dimich-Ward, 1991

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

Department of Interdisciplinary Studies
The University of British Columbia
Vancouver, Canada

Date February 11, 1991

ABSTRACT

Three longitudinal pulmonary function data sets were analyzed by several statistical methods for the purposes of:

- 1) determining to what degree the conclusions of an analysis for a given data set are method dependent;
- 2) assessing the properties of each method across the different data sets;
- 3) studying the correlates of FEV_1 decline including physical, behavioral, and respiratory factors, as well as city of residence and type of work.
- 4) assessing the appropriateness of modelling the standard linear relationship of FEV_1 with time and providing alternative approaches;
- 5) describing longitudinal change in various lung function variables, apart from FEV_1 .

The three data sets were comprised of (1) yearly data on 141 veterans with mild chronic bronchitis, taken at three Canadian centres, for a maximum of 23 years of follow-up; their mean age at the start of the study was 49 years (s.d.=9) and only 10.6% were nonsmokers during the follow-up; (2) retrospective data on 384 coal workers categorized into four groups according to vital status (dead or alive) and smoking behavior, with irregular follow-up intervals ranging from 2 to 12 measurements per individual over a period of 9 to 30 years; (3) a relatively balanced data set on 269 grain workers and a control group of 58 civic workers, which consisted of 3 to 4 measurements taken over an average follow-up of 9 years. Their mean age at first measurement was 37 years (s.d.=10) and 53.2% of the subjects did not smoke.

A review of the pulmonary and statistical literature was carried out to identify methods of analysis which had been applied to calculate annual change in FEV_1 . Five methods chosen for the data analyses were variants of ordinary least squares approaches. The other four methods were based on the use of transformations, weighted least squares, or covariance

structure models using generalized least squares approaches.

For the coal workers, the groups that were alive at the time of ascertainment had significantly smaller average FEV_1 declines than the deceased groups. Post-retirement decline in FEV_1 was shown by one statistical method to significantly increase for coal workers who smoked, while a significant decrease was observed for nonsmokers. Veterans from Winnipeg consistently showed the lowest decline estimates in comparison to Halifax and Toronto; recorded air pollution measurements were found to be the lowest for Winnipeg, while no significant differences in smoking behavior were found between the veterans of each city. The data set of grain workers proved most amenable to all the different analytical techniques, which were consistent in showing no significant differences in FEV_1 decline between the grain and civic workers groups and the lowest magnitude of FEV_1 decline.

It was shown that quadratic and allometric analyses provided additional information to the linear description of FEV_1 decline, particularly for the study of pulmonary decline among older or exposed populations over an extended period of time. Whether the various initial lung function variables were each predictive of later decline was dependent on whether absolute or percentage decline was evaluated. The pattern of change in these lung function measures over time showed group differences suggestive of different physiological responses.

Although estimates of FEV_1 decline were similar between the various methods, the magnitude and relative order of the different groups and the statistical significance of the observed inter-group comparisons were method-dependent. No single method was optimal for analysis of all three data sets. The reliance on only one model, and one type of lung function measurement to describe the data, as is commonly found in the pulmonary literature, could lead to a false interpretation of the result. Thus a comparative approach, using more than one justifiable model for analysis is recommended, especially in the usual circumstances where missing data or irregular follow-up times create imbalance in the longitudinal data set.

ACKNOWLEDGEMENT

I would like to express my gratitude to the following individuals who each took the time to provide insightful commentary on specific aspects of the thesis: John Spinelli, Dr. Steve Marion, Dr. Susan Kennedy, Dr. Ben Burrows and Dr. Charles Lamb. Assistance in obtaining and running specialized software was generously provided by Stan Kita and Calvin Lai of the UBC Computing Centre; Dr. James Ware of Harvard University; Dr. Scott Zeger of Johns Hopkins University; Dr. Ronald Helms of University of North Carolina; BMDP Statistical Consulting; and the SFU Computing Centre. To members of the Division of Occupational and Environmental Health, I extend my appreciation for being accommodating of my dual responsibilities as graduate student and research associate. Dr. Brenda Morrison, Dr. Chris Van Netten, Dr. Ted Sterling and Mary Hehn have been most supportive of my academic endeavors. Many thanks to the committee members, Dr. Moira Yeung, Dr. John Ledsome, Dr. Sheps and Dr. Schulzer. My thesis supervisor, Dr. David Bates remained steadfast in his support throughout my years as a doctoral student, and it is to him that I owe my sincere gratitude for his kind and thoughtful tutoring.

The understanding and encouragement received from my friends, especially from Ray and Harriet Chamberlin, Janice Yee, Chris McKenzie, Joanne Evanoff, Dr. Ron Peterson and Dr. Fred Adrian, helped me to cope with the many challenges over the years. I am grateful to my family, especially my brother, Michael, as well as to Ed Westfall and the many teachers who felt that I had the potential. Most importantly, the achievement of my doctoral degree was possible only through the loving encouragement and help from my husband, Richard, who is always there for me.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENT	iv
List of Figures	viii
List of Tables	x
I. INTRODUCTION	1
Purpose	5
Definition of Abbreviations	9
II. THE MEASUREMENT OF PULMONARY FUNCTION	11
Physiological Basis of Lung Function Measurements	11
Variability of Lung Function Measures	17
Determining Abnormality of Decline	21
III. A HISTORICAL BACKGROUND OF LONGITUDINAL PULMONARY FUNCTION RESEARCH	23
Description of Tabulated Studies	26
Discussion of the Literature Background	53
IV. METHODOLOGICAL PROBLEMS IN DESIGN AND ANALYSIS	60
Missing Values	64
V. A SURVEY OF PARAMETRIC METHODS USED FOR LONGITUDINAL DATA ANALYSIS	67
Introduction	67
Individual Regression	72
General Linear Models	76
VI. DESCRIPTION OF THREE LONGITUDINAL STUDIES	90
Introduction	90
The Coal Miners Study	90
The Veteran's Study	93

Grain Handlers Study	99
Summary	102
VII. METHODS OF ANALYSIS	106
Introduction	106
Comparison of Methods	108
Ordinary Least Squares Methods	112
Additional Statistical Methods	115
Criteria Used for Comparison of Methods	119
Practical Applications	123
VIII. DESCRIPTIVE DATA ANALYSIS	125
Description of Data Sets	126
Conclusion	167
IX. RESULTS	169
Comparative Analysis	169
Unadjusted Decline Estimates	170
Adjusted Decline Estimates	174
Additional Statistical Methods	185
Random Effects Model	188
Nonlinearity of Decline	191
Practical Applications	198
X. DISCUSSION	217
Appropriate Models of Analysis	218
Uses of Longitudinal Pulmonary Data	223
Cross-sectional versus Longitudinal data	224
Statistical Aspects	225
Group Decline Characteristics	229

The Effects of Baseline Characteristics on FEV ₁ Decline	237
The Influence of Follow-up Events on FEV ₁ Decline	239
Variability of Lung Function	244
Other Lung Function Measures	246
Nonlinearity Aspects of Decline	249
Conclusion	251
REFERENCES.	255

LIST OF FIGURES

Figure	Page
2.1 Static Lung Volumes	15
8.1a TORONTO - Mean FEV_1 by Year	134
8.1b WINNIPEG - Mean FEV_1 by Year	134
8.1c HALIFAX - Mean FEV_1 by Year	135
8.1d MONTREAL - Mean FEV_1 by Year	135
8.1e TORONTO - Mean VC by Year	136
8.1f WINNIPEG - Mean VC by Year	136
8.1g HALIFAX - Mean VC by Year	137
8.1h MONTREAL - Mean VC by Year	137
8.2a Mean FEV_1 by Occasion for Grain and Civic Workers	142
8.2b Mean FVC by Occasion for Grain and Civic Workers	143
8.2c Mean MMF by Occasion for Grain and Civic Workers	144
8.3a COAL - Standardized Plots of Residuals Versus Standardized Predicted Values of FEV_1	149
8.3b VETERANS - Standardized Plots of Residuals Versus Standardized Predicted Values of FEV_1	149
8.3c GRAIN - Standardized Plots of Residuals Versus Standardized Predicted Values of FEV_1	150
8.4a COAL - Normal Probability Plots of the Standardized Residuals based on a Linear Regression of FEV_1 on Years of Follow-up	152
8.4b VETERANS - Normal Probability Plots of the Standardized Residuals based on a Linear Regression of FEV_1 on Years of Follow-up	153
8.4c GRAIN - Normal Probability Plots of the Standardized Residuals based on a Linear Regression of FEV_1 on Years of Follow-up	154
8.5a COAL - Histogram of FEV_1	155
8.5b VETERANS - Histogram of FEV_1	156
8.5c GRAIN - Histogram of FEV_1	157

8.6a	COAL - Standardized Plots of Residuals of FEV_1 by Standardized Years of Follow-up	160
8.6b	VETERANS - Standardized Plots of Residuals of FEV_1 by Standardized Years of Follow-up	160
8.6c	GRAIN - Standardized Plots of Residuals of FEV_1 by Standardized Years of Follow-up	161
9.1a	TORONTO - FEV_1 versus Year for Case #131	210
9.1b	TORONTO - FEV_1 versus Year for Case #156	210
9.2a	WINNIPEG - FEV_1 versus Year for Case #258	211
9.2b	WINNIPEG - FEV_1 versus Year for Case #250	211
9.3a	HALIFAX - FEV_1 versus Year for Case #426	212
9.3b	HALIFAX - FEV_1 versus Year for Case #403	212
9.4a	DEAD NONSMOKERS - FEV_1 versus Year for Case #1074	213
9.4b	DEAD NONSMOKERS - FEV_1 versus Year for Case #1029	213
9.5a	ALIVE NONSMOKERS - FEV_1 versus Year for Case #2036	214
9.5b	ALIVE NONSMOKERS - FEV_1 versus Year for Case #2028	214
9.6a	DEAD SMOKERS - FEV_1 versus Year for Case #3101	215
9.6b	DEAD SMOKERS - FEV_1 versus Year for Case #3107	215
9.7a	ALIVE SMOKERS - FEV_1 versus Year for Case #4069	216
9.7b	ALIVE SMOKERS - FEV_1 versus Year for Case #4102	216

LIST OF TABLES

Table	Page
2.1 Sources of Variation in Measurements of Lung Function	18
3.1 STUDIES OF NORMAL SUBJECTS: Endpoint Analysis	30
3.2 STUDIES OF NORMAL SUBJECTS: Endpoint 2 Step Analysis	31
3.3 STUDIES OF NORMAL SUBJECTS: Regression Analysis	32
3.4 STUDIES OF NORMAL SUBJECTS: Regression 2 Step Analysis	33
3.5 STUDIES OF OCCUPATIONALLY EXPOSED SUBJECTS: Endpoint Analysis	39
3.6 STUDIES OF OCCUPATIONALLY EXPOSED SUBJECTS: Endpoint 2 Step Analysis	41
3.7 STUDIES OF OCCUPATIONALLY EXPOSED SUBJECTS: Regression Analysis	42
3.8 STUDIES OF OCCUPATIONALLY EXPOSED SUBJECTS: Regression 2 step Analysis	43
3.9 STUDIES OF DISEASED SUBJECTS: Endpoint Analysis	50
3.10 STUDIES OF DISEASED SUBJECTS: Regression Analysis	51
3.11 STUDIES OF DISEASED SUBJECTS: Regression 2 Step Analysis	52
3.12 Range of Published Ranges of FEV ₁ Decrement For Occupational Studies	54
6.1 Measurements of Coal Miners' Pulmonary Function at First Examination	92
6.2 Veteran's Physical Characteristics and Smoking Habits	97
6.3 Measurement of Veteran's Pulmonary Function at First Examination	98
6.4 Measurements of Grain Workers and Controls at First Examination	101
6.5 Published Estimates of Lung Function Decline for the Three Thesis Studies.	105
8.1 Baseline Measurements of Coal Miners Study Group	127
8.2 Baseline measurements on the study group of Veterans	131
8.3 Correlation Coefficients for the Comparison of Lung Function Measures Between <i>VETERANS</i> Cities	132
8.4: Baseline Measurements of the Grain Workers and Controls Study Group	138
8.5 Regression Coefficients for the Comparison of Lung Function Measures in the <i>GRAIN</i> Data Set	139
8.6 Percentages of Missing FEV ₁ Data by Year of Follow-up	164
8.7 Correlation Coefficients Between Initial or Mean FEV ₁ or Slope	165
9.1 A Comparison of Mean Unadjusted FEV ₁ Change Estimates	172
9.2 A Comparison of Mean Age-Height and Fully Adjusted FEV ₁ Decline Estimates	174
9.3 Comparison of Age-Height Adjusted Prediction Equations	177
9.4 Comparison of Significant Predictors	179
9.5 Standardized Regression Coefficients by Group Using Method A Full Model	181
9.6 Comparison of Goodness of Fit Criteria for each Model	183
9.7 Weighted Averages of Unadjusted Individual Slopes	187
9.8 Weighted Least Squares Estimates of Decline	187
9.9 Results of Random Effects and Unstructured Covariance Models in the <i>GRAIN</i> data set	189
9.10 Variability Estimates using Random Effects and Unstructured Covariance Models in the <i>GRAIN</i> data set	189
9.11 Mean Quadratic and Allometric Decline Estimates (Method E)	193
9.12 A Comparison of FEV ₁ Decline Coefficients Using a Linear, Quadratic or Allometric Model	197
9.13 Smoking Characteristics of the <i>VETERANS</i> and <i>GRAIN</i> subjects during follow-up.	199
9.14 Initial Lung Function as Predictors of FEV ₁ Decline.	206
9.15 A Comparison of Unadjusted Lung Function Decline Estimates Based Upon Regressions of Individual Data.	209

CHAPTER I

INTRODUCTION

In the epidemiological context, longitudinal data sets are defined as having a minimum of two observations per individual studied over a period of time (Cook and Ware 1983). This terse definition encompasses a wide variety of study designs, with the more common ones ranging from pre-post "workshift" observations to time-series designs for measurements made over a long period of time. Data sets pertaining to this thesis are more appropriately analysed by a "growth curve" approach except that the outcome to be evaluated is not one of "growth" but "decline" in adult lung function with time. For a repeated measures design, experimental manipulation is usually introduced to set up "control" versus "treatment" conditions. By comparison, it is the observational process of documenting changes in the dependent variable with time and the antecedents of such changes that are relevant to the aims of this thesis.

What distinguishes longitudinal, as opposed to "cross-sectional" designs is the element of time. Observations collected in a cross-section of individuals represent only one point in time. Any inferences about change with time are obtained through regression analysis, with subject age being representative of time. The "cohort effect" may introduce a bias where the estimated age coefficient is assumed to be representative of inferences drawn from the present differences between younger and older persons (Glindmeyer et al., 1982). Aside from the ageing phenomenon, specific birth cohort differences in morphology, occupational habits, environmental exposures, as well as selective factors for drop-out, may be influential (Bande et al. 1980).

Due to the relative simplicity in carrying out such a study, the cross-sectional approach has been widely used, particularly as a hypothesis-generating tool. Such analyses can indicate the relative prevalences of different risk factors and outcomes among the different sample

populations studied. However, with no direct measurement of time, cause/effect associations can only remain conjectural and selection biases in particular may be operative.

Longitudinal designs which follow an individual's responses over time have particular relevance to studies of growth and development, and ageing. "Pure" longitudinal growth studies require the study participants to be of the same age at the study onset. In the more commonly used mixed-longitudinal study, a cross-section of the population is followed-up on repeated occasions, such that not all individuals are measured at the same ages.

In a longitudinally designed study where a cohort is defined by exposure, the age, time, and year of birth effects can usually be estimated separately (Cook and Ware 1983), unlike the cross-sectional designs, which rely on one dimension of time alone. Use of the longitudinal method has the additional potential of distinguishing long term trends from temporary fluctuations. A most significant attribute is that it may facilitate the identification of causal relationships, as a cause/effect time relationship is preserved (Assemota 1979).

The problems of longitudinal design can become apparent both in implementation and analysis of the results. Conducting a longitudinal study can be both expensive and cumbersome, as there are practical difficulties encountered in maintaining the consistent efforts of the study team, as well as funding the project throughout the lengthy study period. Using the same trained personnel throughout the study period can minimise inter-observer and survey bias; this provides an ideal yet often impractical setting. Maintenance of standard performance of equipment may also present a problem.

Lung function testing has much potential for error of measurement, due in part, to improper calibration and equipment error (Becklake et al., 1985). Submaximal efforts by study subjects will increase the variability of spirometric measurements (Kellie et al. 1987), which may be additionally affected by circadian and other physiological variation. The lung function measurement with a comparatively low coefficient of variation is the Forced Expiratory

Volume in one second (FEV_1). Within-subject variability of FEV_1 among normal subjects has generally been reported to be within 3% (Dockery et al., 1985).

Whether the study results are generalizeable usually requires that a random sampling strategy be used at the outset of the study. Logistic problems of follow-up often limit the sample group to volunteers from defined units, such as those at a work site, or hospital. The longer the period of follow-up the less representative, that is biased towards survivors, the result becomes (Anthonisen et al., 1986). The "healthy worker effect" is a potential bias in occupational settings, and may limit whether the findings can be generalized due to the level of fitness of the workforce.

Compliance by all the participants in all measurement sessions would conform to a completely balanced design and facilitate analysis. The basic assumptions of many statistical methods require the structure of a relatively balanced data set. Attrition through noncompliance, outmigration, or death, together with measurement of each subject at unequal time intervals and occasional missing observations creates an unbalanced data structure. A common assumption for unbalanced structures is that the data are missing at random, that is, the probability of missing an observation does not depend the value of that observation or on relevant covariates. However, when the causes of missing data are related to the outcome, such as when the level of FEV_1 that would have been measured was directly related to death due to respiratory causes, the missing data should be regarded as "non-ignorable".

With repeated measurements on an individual, there tends to be positive correlation between successive determinations. The more familiar statistical tools based on ordinary least squares analysis require the assumption of independence with no correlation between random error terms. If the within-subject measurements show such "autocorrelation", the ordinary least squares regression methods may indicate greater precision of the regression coefficients than is actually the case (Neter et al., 1985).

It has been estimated that it takes approximately six years of yearly follow-up for variability between individuals to exceed that within individuals, so that it accounts for at least 50% of the total variability (Diem, 1983). Variability of FEV₁ measurements may be indicative of underlying pathology (Bates, 1973). Somewhat misguided attempts have been made to "clean" longitudinal pulmonary function data; for example, one technique was adopted to decrease total variance by excluding all FEV₁ readings more than three standard deviations from the mean (Fletcher et al., 1976). According to Whittemore (1981, p422):

"Longitudinal studies show great variability in one person's pulmonary function measurements over time, as well as in rates of decline between people. The prognostic importance of this variability has not been adequately studied. Even in long-term studies the calculated individual rates of decline are not highly reliable. Better methods of analyzing this type of data are needed."

Questions on particular aspects of longitudinal analysis are dominated by overall considerations of choice of method. A limitation of the typical analytical methods employed, whether they involve crude slope estimates or maximum likelihood derived coefficients, is their reliance on a linear relationship between the variables. This has intuitive appeal as decline is described as the change in the functional variable over time. While the coefficient of decline that is obtained by such analyses can be interpreted easily, an implicit assumption is that the decline can be modelled by a constant slope throughout the follow-up period. Apart from the study by Emergil and Sobal (1971) whose findings were limited to a select group of clinical patients, few investigators have explored the non-linear aspects of lung function decline in adults. Plots of individual data often demonstrate nonlinear curves for the FEV₁ values plotted against time of the measurement. Describing the relationship between these variables as a straight line, may mask such properties as timing and extent of accelerated decline, or variability of the aging process.

Longitudinal data sets derived from the natural observation process require careful evaluation of the degree of unbalanced structure, the cause of missing data, and the extent to which the data conform to the assumptions of the chosen analytical technique. The

pulmonary function literature provides examples of simplistic analyses applied to complex data sets with little justification of the choice of method apart from its common use. As an illustration, Diem (1983) discusses the example of a 9 survey study of toluene diisocyanate workers. If only a two point analysis had been used, and the slope obtained from the first and last values, instrument-based bias in the fifth survey would have revealed an unexpected result of a positive annual change in FEV_1 . The choice of the appropriate analytical tool is by no means an obvious process. New statistical methods which have the potential to be applied to longitudinal data are being developed, but their effectiveness on real data sets, and communication of these methods on an applied rather than a theoretical basis, is as yet poorly described in the epidemiological literature. In this connection Buist and Vollmer (1988, p16) stated that "Future work in this area must focus not merely on developing newer more sophisticated models, but also on developing models which are responsive to the nature of the data... and which are understandable to the clinicians and epidemiologists working in the field."

Purpose

It is the intention of this thesis to evaluate a variety of statistical techniques currently available through the use of common mainframe software and to identify predictors of lung function decline in the longitudinal data sets. Three separate series of longitudinal pulmonary function data form the basis of the thesis dissertation. They encompass four different populations studied over time periods extending from 9 to 25 years. Individually, the data sets differ in design and implementation; collectively, they offer a basis of comparison for the evaluation of various analytical techniques and methodological issues.

The data sets, obtained through the generosity of Dr. David Bates and Dr. Moira Yeung of the University of British Columbia Faculty of Medicine, are derived from the following three studies, and associated publications.

- (1) The coordinated study of chronic bronchitis in the Department of Veterans Affairs Canada. D.V.A. (Bates, Wolf and Paul, 1962; Bates, Gordon, Paul, Place, Snidal and Wolf, 1966; Bates, 1973; Bates, 1974; Bates, 1989).
- (2) A longitudinal study of coal miners in Lorraine, France. (Bates, Pham, Chau, Pivoteau, Dechoux, and Sadoul, 1985).
- (3) A study of grain elevator workers in the Port of Vancouver and a control group of Vancouver city hall workers (Chan-Yeung, Schulzer, Maclean, Dorken and Grzybowski, 1980; Chan-Yeung, Schulzer, Maclean, Dorken, Tan, Lam, Enarson and Grzybowski, 1981; Tabona, Chan-Yeung, Enarson, Maclean, Dorken and Schulzer, 1984; Enarson, Vedal and Chan-Yeung, 1985; Schulzer, Chan-Yeung and Tan, 1982; Schulzer, Enarson and Chan-Yeung, 1985).

These data sets will be referred to as the "*VETERANS*", "*COAL*" and "*GRAIN*" data. My primary purpose is to use these sets of real, as opposed to simulated data, to answer the following questions;

- *What difference do different methods of analysis of longitudinal data applied to the same data set, make to the conclusions that might be drawn from them?*
- *Do the different methods produce the same comparative results when applied to different data sets?*

The three thesis data sets analysed differ with respect to the selection of the original study population, demographic characteristics, the frequency and regularity of measurement, duration of the follow-up period, attrition rate, and detail of information on smoking habits. The *GRAIN* data is relatively balanced; it consists of the subset of all grain workers who were examined prospectively up to four times and whose measurements spanned a 9 year period. Those with more than one missed examination were excluded. The *COAL* data represents the opposite extreme in design. The data were collected retrospectively from existing records; examinations were sporadic with individuals possessing from 2 to 11 records over a range of 5 to 33 years of follow-up. The *VETERANS* data set is a rare example of a

long term prospective study where attempts were made to conduct annual follow-up for a minimum duration of 14 years. A complication of the data structure is the various sources of missing data, attributable to loss to follow-up, deaths and missed examinations, which contribute to imbalance in its structure. Both the flexibility (ie. is the method suitable for all the different data sets?) and specificity (does the method offer a suitable analysis of a particular aspect of the data set?) will be ascertained on each data set.

The comparison of methods will be based on the decline estimates which are 1) unadjusted, 2) age and height adjusted, 3) additionally adjusted for differing distributions in length of follow-up, smoking and other characteristic variables unique to each data set.

Each data set provides valuable information on initial and follow-up levels of spirometric as well as other lung function parameters, such as diffusing capacity and residual volume. It will be determined whether any association can be shown to exist between initial values of any of these other lung function variables with FEV₁ decline. The longitudinal decline of these measures will also be described.

The data sets offer distinct groupings, clearly separated by such independent variables as place of residence, smoking status, as well as occupational exposures and vital status. Baltes and Nesselroade (1979) pointed out that in order to move longitudinal research more in the direction of quasi-experimental design, two facets to consider are including control groups and extending the number of occasions of longitudinal observations. Two of the present data sets cover at least twenty years of follow-up, while the third includes a separate control group of non-exposed workers; there is therefore an opportunity for hypothesis testing using the chosen analytical method. The following null hypotheses will be tested on the appropriate data sets;

H_0 : The rate of FEV_1 decline is independent of:

1. Cities of Residence
2. Exposure Status
3. Smoking Group - within vital status
4. Interaction of Smoking Group with City of Residence or Exposure Status

The determination of the rate of change of pulmonary function measured on a continuous scale, is the principal outcome of interest. With the onset of the observation period being in all cases restricted to adult years, the phenomenon of interest is the eventual decline of pulmonary function with time. The American Thoracic Society (1985) has defined a statistically significant rate of decline of pulmonary function beyond ageing as an adverse respiratory health effect, thus, encouraging efforts to accurately determine the rate of pulmonary function decline.

The measurement of longitudinal decline in pulmonary function is known to have important clinical implications. Postma et al. (1979) studied patients with severe chronic obstructive pulmonary disease, and found that a decline in the spirometric measure of FEV_1 was the sole determinant of prognosis. Other studies reinforce the predictive power of FEV_1 decline (Peto et al., 1983; Bates, 1989). Chronic non-specific lung disease, has been a major cause of mortality in North America and Europe for at least two centuries (Fletcher et al., 1976) and its incidence in North America has increased over the last three decades (US Surgeon General, 1984). Determining the etiology of rapid decline in FEV_1 in this condition would have important clinical implications for its prevention (Bates, 1974).

On an individual basis, the measurement of change in pulmonary function as determined through longitudinal study of the process, may provide an early warning of impending disease. The identification and control of risk factors affecting pulmonary function decline provides the epidemiological justification for its investigation.

Definition of Abbreviations

DATA SETS

COAL - A data set on French coal miners.

GRAIN - A data set on Port of Vancouver grain elevator workers and a control group of city hall workers.

VETERANS - A data set based on the follow-up of Canadian veterans with chronic bronchitis.

LUNG FUNCTION

DLCO - Diffusing capacity of the lungs.

ERV - Expiratory reserve volume.

FCO - Fractional carbon monoxide uptake.

FEV₁ - Forced Expiratory Volume in 1 second.

FEV_{0.75x40} - Indirect maximal breathing capacity.

FRC - Functional residual capacity.

FVC - Forced vital capacity.

ME - Mixing efficiency (%).

MMF or *FEF₂₅₋₇₅* - Maximal midexpiratory flow.

RV - Residual volume.

VC - Vital capacity measured by the helium closed circuit technique.

TLC - Total lung capacity.

STATISTICAL TERMS

ARI - First-order autoregressive model.

ARMA - Autoregressive moving average model.

EM - EM algorithm (Expectation and Maximization)

GEE - General Estimating Equations models.

GLS - Generalized Least Squares.

HR - Horse-racing effect.

MCAR - Data that is missing completely at random.

Method A - Regression of all data points.

Method B - Regression of all pairs of data points.

Method C - Endpoint annual slope.

Method D - Percentage change of initial lung function.

Method E - Regression of individual data.

OLS - Ordinary Least Squares.

RE - Random Effects model.

RTM - Regression to the mean.

SEE - Standard error of the estimate.

SPSSX - Statistical package for the Social Sciences.

YRDIF - Years of follow-up.

R² - Coefficient of determination.

CHAPTER II

THE MEASUREMENT OF PULMONARY FUNCTION

Physiological Basis of Lung Function Measurements

Pulmonary function tests are becoming increasingly useful as a measurement tool in epidemiological studies of lung diseases and disorders. The reasons for this include their noninvasiveness, quantitative nature and their relative ease in administration (Bates,1989).

Gas exchange between the blood and atmospheric air is the principal function of the lung. The lungs consist of twenty generations of progressively narrowing airways which terminate at the alveoli where gas exchange takes place; this provides up to 200,000 airways in parallel. The large central bronchi and bronchioles, whose diameters are greater than 2 mm., depend upon cartilaginous support, while the smaller respiratory bronchioles rely upon the elasticity of the lung to support the airway structure.

It has been recommended that spirometric tests of ventilatory capacity should be an integral part of all epidemiologic respiratory health studies (Chan-Yeung et al. 1985). The effects of increased airway resistance can be assessed by means of forced expiration into a spirometer. The forced vital capacity maneuver requires the subject to inspire to full lung inflation then exhale as rapidly, completely and forcibly as possible. The maximum volume of air that can be rapidly exhaled is the Forced Vital Capacity (FVC); while the maximum flows that are measured are Forced Expiratory Volume in one second (FEV_1); and the volume between 25% and 75% of the FVC divided by time is the FEF_{25-75} , (which is also called the Maximal Midexpiratory Flow or MMF). All three thesis data sets include tests of the Forced Vital Capacity maneuver. The Coal Miners' data set lacks FEF_{25-75} measurements, while measurements on the *VETERANS* consists of the static volume measure of Vital Capacity as opposed to FVC measures, and $FEV_{0.75}$ values rather than FEV_1 measures. A

conversion factor can be applied to $FEV_{0.75}$ values to allow comparative evaluations (Miller et al. 1959).

The minimum testing equipment required is a spirometer to measure expired air, along with an attached recording device which provides the volume versus time tracing of the forced expiration. The classic water-sealed and dry bellows spirometers measure the change of volume with a drum kymograph or a moving stylus on a sheet of paper. The more recent models are electronic based and provide an output proportional to air flow, while the volume change is determined by the integration of an electric flow signal (Dawson 1985). To obtain a valid tracing, it is recommended that the forced expiratory volume maneuver should continue to be recorded for at least six and preferably ten seconds. Generally, spirometric measurements have had the most extensive trials in the past 25 years; they require relatively low cost equipment, have a short testing time and the test procedure is simple, standardized and highly acceptable to the subjects.

It appears that airway disease causing obstruction tends to result in simultaneous changes which affect FEV_1 primarily, as well as FVC. The physiological basis for such obstructive processes can include partial occlusion of the lumen due to excessive secretions, thickening of the airway wall due to edema or hypertrophy of the mucous glands, and narrowing of the airway due to destruction of lung parenchyma. The FVC will be reduced with the premature closure of airways which traps the air at high lung volumes (West 1977). Restrictive diseases are more associated with reduced FVC, although the measurement of static volumes would assist in confirming the distinction.

The maximal flow during the first third of the FVC has generally been considered effort dependent whereas flows over the latter two-thirds of the FVC are more dependent on the mechanical properties of the lungs. (Allen and Sabin 1971). The FEF_{25-75} which is measured over the mid-half of the FVC maneuver requires relatively little patient cooperation.

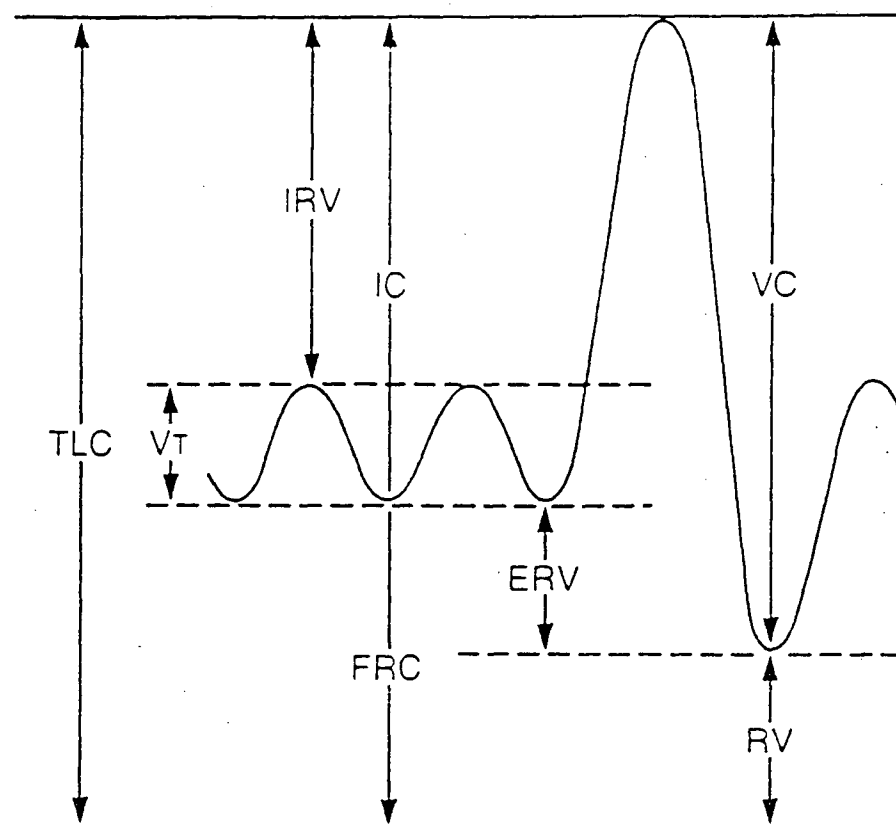
Although it is much more reproducible than tests further along the curve, published values of its coefficient of variation range from 1 to 16%, in contrast to published coefficients of variation for FEV₁ and FVC which range from 1.6 to 8.2% (Becklake and Permutt 1979). Despite its variability Leuallen and Fowler (1955) found the FEF₂₅₋₇₅ to be more sensitive as a detector of obstructive impairment than FVC or FEV₁. The physiological basis for small airway obstruction is that the small airways lack cartilaginous support and depend for their patency on the lung elastic recoil. Since the recoil pressure diminishes with decreasing lung volume, resistance of the small airways increases relative to that of the central airways in the terminal part of the forced expiration. Therefore, it is quite possible that where the FEV₁ values are normal, measurements obtained at lower lung volume such as the FEF₂₅₋₇₅ may be abnormally reduced for those with small airway disease (Pare et al., 1982; Dawson, 1985). By contrast, it has been postulated that traditional spirometric measurements may not be sensitive enough to detect dysfunction in this quiet zone of the lung (Chan-Yeung et al. 1985). Since less than 25% of the total resistance to airflow is in the peripheral airways, a large increase in their resistance would tend not to affect overall airway resistance (Miller, 1986). Further tests of small airway function include tests of gas distribution (frequency dependence of compliance, closing volume) and measurements of flow-volume relationships (Schechter, 1986).

Static Lung Volumes

Residual volume, function residual capacity and total lung capacity cannot be measured directly by spirometry because they include air that cannot be expelled from the lungs. Their measurement requires more sophisticated methods, such as body plethysmography, radiography or inert gas dilution (Shigeoka 1983). Measurements used in the *VETERANS* study were obtained through the latter method, a helium closed circuit technique. The technique involves filling the spirometer and apparatus with a mixture of helium in air, measuring the volume of gas and helium concentration therein, allowing the subject to rebreathe from this reservoir

until equilibration occurs, and measuring the final concentration in the system (Brown, 1986). Values were recorded for Tidal Volume (TV), which is the volume of each breath exhaled during quiet respiration; and the Total Lung Capacity (TLC) which is the total volume of air contained in the lungs and bronchial tree with maximal inspiration. When the subject exhaled as much air as he could Vital Capacity (VC) was obtained; and the Residual Volume (RV) was the volume of air remaining in the lungs after that full expiration. The volume of air remaining in the lung at the end of a normal tidal expiration is known as the Functional Residual Capacity (FRC) and the volume that can be exhaled from the FRC is the Expiratory Reserve Volume (ERV). Refer to Figure 2.1 for a depiction of these various lung volumes (Bates, 1989). For the determination of VC, as opposed to the FVC derived from spirometry, the patient may take as long as required to complete the maneuver. While in normal subjects there is little difference between the two measures, in certain types of airway obstruction such as emphysema, air trapping may result in an FVC reading which is less than that found for the VC (Morgan and Seaton, 1984).

Figure 2.1: Static Lung Volumes.



TLC = Total lung capacity
FRC = Functional residual capacity
VC = Vital capacity
ERV = Expiratory reserve volume
RV = Residual volume
IC = Inspiratory capacity
VT = Tidal volume
IRV = Inspiratory reserve volume

The determinants of TLC include lung and chest size, the elastic recoil of the lungs and chest wall, as well as inspiratory muscle strength. RV also is determined by most of these factors, and additionally by expiratory muscle strength and airway trapping (Ries and Clausen 1985). RV is thought to increase with advancing age due to progressive loss of lung and chest wall elasticity with ageing and also to closure of small airways at larger lung volumes. Lung volume tests are generally regarded as supplementary to spirometric measurements. For example, a low TLC helps verify a restrictive process while a large value is suggestive of the hyperinflation of emphysema (Shigoeka 1983). In obstructive conditions TLC may remain unchanged or increase only slightly while the Residual Volume will characteristically increase. In consequence, the Vital Capacity will be diminished.

Diffusing Capacity

The only noninvasive test of the gas exchange properties of the lung is the diffusing capacity. In addition to diffusion, the test may be affected by hemoglobin concentration and the reaction rate of the gas with hemoglobin (Miller, 1986). An alveolar sample is taken after a test gas mixture of carbon monoxide, oxygen, helium and nitrogen is inhaled, and the diffusing capacity is calculated as the volume of CO taken up per minute per mmHg alveolar P_{CO} (West 1977). Either single breath or steady state determinations can be used. In the *VETERANS* and *COAL* studies an additional measure used was the fractional carbon monoxide uptake. It was calculated according to the equation: $FCO = 1 - (F_{ECO}) / F_{ICO}$; where F_{ECO} is the fraction of carbon monoxide expired, and F_{ICO} is that inspired. The value is related to minute volume; an abnormal value may indicate the presence of a significant loss of diffusing surface, impairment of CO transport at membrane level or significant distributive effects (Bates et al. 1974). The diffusing capacity is reduced when the alveolar capillary membrane thickens due to such diseases as interstitial fibrosis and asbestosis and due to the destruction of the pulmonary vascular bed (surface area), such as found in emphysema (Miller, 1986). However, the tests of diffusion are less reproducible than

spirometric tests, with coefficients of variation ranging from 5 to 18% among different studies (Becklake and Permutt 1979). This can be attributed in part to the unevenness of ventilation, blood flow and diffusion properties exhibited throughout the diseased lung (West 1977).

Variability of Lung Function Measures

"Variety is the spice of life; it is also the main subject for biological study" (p132, Bates, 1989)

Sources of Variability

Variability can be categorized as arising from three sources:

- 1) Within-Individual
- 2) Between-Individuals
- 3) Between-Population

as listed in Table 2.1 (adapted from Bates 1989). A subject's level of pulmonary function is difficult to determine precisely due to the inherent variability in the measures. Intrasubject error may arise both from measurement error and physiological variation. Pulmonary function equipment that is appropriately calibrated may still have a specified accuracy of $\pm 3\%$ which gives a potential for measurement error of up to 6% per survey. For a healthy middle aged male this may represent from 200 to 300 ml in the measurement of FEV₁ (Buist and Volmer 1988). Over the duration of a longitudinal study technical factors brought about by within and between-instrument error, as well as procedural differences may come into play. There is the potential for observer errors in terms of administration of the tests or interpretation of the results; subject errors in comprehension or co-operation; effects of interactions between the subject and observer or the observer and the instrument; and other technical problems such as temperature differences and altitude effects all contribute to the variation in spirometric results (Becklake 1986). Lebowitz et al. (1982) concluded that subject

and technician training were major contributors to variability in their pulmonary function results.

Table 2.1: Sources of Variation in Measurements of Lung Function

VARIATION	SOURCE
Within-individual	
Technical:	Within and between instruments Within and between observers in administration and reading of tests Curve selection Temperature extremes
Biologic:	Comprehension and/or co-operation of the subject Circadian, weekly and seasonal effects Endocrine; other Transient environmental effects Transient disease states
Between-individual	
All the above sources +	
Subject	Size, sex, age, respiratory muscularity Race and other genetic characteristics Past and present health Habits (e.g. smoking, physical activity)
Environmental	Residence (income, ambient pollution) Indoor pollution (smoking, gas stoves etc.) Occupational exposures
Between-population	
All the above sources +	
Selection into or out of the target or study population	

Through the efforts of the scientific community, attempts have been made to improve the accuracy of pulmonary function measurements, particularly those of spirometric recordings. (See the statement from the American Thoracic Society (1983) on the screening for adult respiratory disease). Unacceptable curves could result from the subject coughing, early termination, glottis closure, equipment leak or an obstructed mouth piece. The most common source of error is failure to record a complete FVC (Townsend, 1982). It is considered essential in epidemiological surveys to monitor subject effort by inscribed curves rather than

rely on the protocol, as well routinely perform external calibration on the automated equipment used (Miller et al. 1980).

The deterioration of plastic hoses, bells and rolling seals can result in the development of spirometer leaks. The FVC is particularly liable to be in error if leaks should occur, while the FEV_1 was found to be robust enough for analysis even if recorded in the presence of small spirometric leaks. This provides further support to the choice of FEV_1 as the spirometric measure recommended for longitudinal analysis especially where equipment is not routinely leak-tested (Townsend 1984).

Townsend (1982) found that short recorded lengths of expiration led to underrecording of FVC which inflated the FEV_1/FVC % ratio as well the FEF_{25-75} values. This was particularly apparent among slightly obstructed subjects. This problem was illustrated in a longitudinal study by Graham et al. (1981). In an earlier survey, respiratory efforts were timed 2.5 to 3 seconds, in comparison to the more recent recommendation of a minimum of 6 seconds of expiratory effort. In consequence, it appeared that there was an overall increase in FVC and in FEV_1 from the earlier to the later survey.

The second factor contributing to intra-subject variation is based on biological or physiological variation. Reports of spontaneous day-to-day and diurnal variation of pulmonary function have appeared since the time of John Huston in 1846. FEV_1 for example, has been shown to have its highest flows recorded at mid-day, while for vital capacity a slight rise was shown in the morning followed by a fall in the afternoon (Hruby and Butler 1975). It is thought that the time of year may influence spirometric results through differences in levels of pollutants or incidences of respiratory infections, or due to variations in the hours of daylight (Miller and Thornton, 1980). Other factors postulated to affect biological variation and pulmonary function include posture, relationship of testing time to meals and endocrinological factors such as those that accompany the menstrual cycle (Becklake 1986).

Reducing Variability

Achieving a reduction in the variance of pulmonary function measures would allow for a more precise estimation of longitudinal decline in such measures. Selection effects which operate by emphasizing interpopulation variability are difficult to adjust for. They are a consequence of the epidemiological approach, where experimental randomization to take account of differential characteristics cannot be achieved.

Intersubject variability on the other hand, is affected by population and environmental characteristics which can to some extent be controlled in the analyses, with appropriate measurement of these characteristics, through adjustment, using analysis of covariance statistical procedures as well as through stratification and matching strategies. Variability can be reduced to some extent by considering between subject differences which can be modelled, as attempted in the thesis, in terms of height, age, cigarette smoking status, occupational exposure and residence. Other potential personal factors include gender, physical activity, muscularity, race, past and present health; environmental factors could include urban air pollution, climate and socioeconomic status.

Intrasubject variability is a much more elusive problem to deal with apart from ensuring consistency and accurate techniques of measurement. At the measurement level, attempts have been made to reduce the variability of spirometric measures by insisting that two acceptable curves may not vary by more than 200 ml during a particular test session. Using this test criterion, it was found that a nonrepeatable test may reflect not only measurement error but also pathology in the tested subjects. The exclusion of study subjects without repeatable pulmonary function tests, particularly in longitudinal follow-up, could result in selection bias (Eisen et al. 1984). The opportunity to study those who perhaps have the more pathological responses may be lost by such a procedure.

Determining Abnormality of Decline

The delineation of "abnormality" has a strong tradition in cross-sectional studies where absolute values of spirometric measurements are compared to that of a reference population. The use of predicted values to differentiate "normal" from "abnormal" requires a thoughtful application. The variability of the given measurement and its potential for a skewed rather than normal distribution must be given consideration (Knudson et al. 1983).

Comparison prediction equations are intended to minimise the variance between individuals in observed lung function, and this is employed through simple linear models accounting for age and height. Such analyses do not account for an increase in decline of FEV_1 with age, or the possible dependence of rate of decline on height-age interactions; which are observed in many longitudinal studies (Burrows et al. 1986). A quadratic relation of lung function with age has been demonstrated for the FEV_1 , FVC and MMF spirometric measures among exposed groups (Chan-Yeung et al. 1981). The use of prediction equations, which assume a linear relationship of the independent variable of age with FEV_1 decline may be inappropriate when applied to individuals showing nonlinearity in FEV_1 decline with time.

Due to the inherent sources of variation, the designation of "abnormality" onto a particular overall decline estimate is difficult to achieve with confidence. One suggestion by Bates (1973) is that a fall in FEF_{25-75} of greater than or equal to 0.6 l/sec in one year with no improvement the following year would be an indicator of progressive disease. Other possibilities include a NIOSH recommendation that workers whose annual decline in FEV_1 or FVC that is greater than 10% be recommended for a clinical evaluation (Hankinson 1986); it was suggested that a smoker who is consistently losing FEV_1 at 100 ml per year is liable to develop disease, whereas a smoker whose FEV_1 is variable without a clear downward trend greater than 60 ml per year is probably not at high risk. It has been also suggested that a

loss of more than 40 ml of FEV₁ per year is indicative of a decline that is greater than expected (Rosenstock and Cullen, 1986). An alternative to an attempt to obtain a clear distinction of "abnormal" based on FEV₁ decline, is to jointly assess the relative magnitudes of deviation from the norm for a variety of lung function tests in order to establish the direction to which the degree of abnormality is tending (Oldham 1979). However, at present the information on abnormality of change in various pulmonary function measures over time and their interrelationships, is lacking.

CHAPTER III

A HISTORICAL BACKGROUND OF LONGITUDINAL PULMONARY FUNCTION RESEARCH

The basis of this chapter is a set of tables in which the pertinent characteristics of studies published on the topic of adult pulmonary function decline are listed. Although the number of studies referred to are not all inclusive they are reasonably representative of the published literature. A description of the tabulated studies is followed by a discussion of the generic problems of study design, analysis, and interpretation. Tables are organised in chronological order according to the type of subjects studied. The three categories, "normal", "occupational" and "diseased" were chosen in accordance with the longitudinal data sets that will be evaluated in this thesis.

The "normal" classification consisted of population samples drawn from various cities, with interest being focussed on the difference between decline in pulmonary function for smokers versus nonsmokers. The separate control group of workers evaluated in the *GRAIN* study were Vancouver City Hall workers who are representative of this category, as they were chosen for being unexposed to occupational respiratory hazards.

Studies classified as "occupational" in nature, include those of such diversely defined groups as cotton operators, hemp workers, pulp and paper mill workers and firemen. Exposed workers of such occupations or industries have the potential to develop pulmonary related disease from respiratory hazards. Two thesis data sets are the results of occupational studies; the *COAL* study was based upon an assessment of coal miners in Lorraine, France, while the *GRAIN* study was on grain handlers in the Port of Vancouver.

The "diseased" category is comprised of studies on patients who are identified as having chronic obstructive pulmonary disease, including chronic bronchitis, emphysema and

asthma; as well as patients with silicosis, who typically develop their disease due to occupational exposures. The purpose of the *VETERANS* study was to describe the natural history of chronic bronchitis.

Within these categories, studies are additionally grouped by type of statistical method employed. The different analytical methods reflect both a change in sophistication in the design as well as analysis of longitudinal data. The numbers of studies employing a particular method are generally indicative of its popularity for a particular time period, although, the listing of studies is not all inclusive and may be more complete for certain types of statistical approaches. The categories are as follows:

- 1) Endpoint: The principle use was for a follow-up study but was sometimes applied when there were a few data points taken per individual. The individual slopes were calculated by dividing the number of years of follow-up into the difference in FEV_1 between the first and last measurements.
- 2) Endpoint/Two Step: In addition to the previous analysis, an attempt was made to adjust the estimate of change for various other influencing factors in the analysis and/or assess their relative contribution to decline.
- 3) Regression: Where a number of points of follow-up existed, an ordinary least squares regression approach was used on the individual's values. The coefficients obtained were averaged to obtain an overall estimate of slope.
- 4) Regression/Two Step: Upon determining the individual regression coefficients, use of multiple regression was used in the majority of instances (although in some cases generalized linear modelling was introduced), to account for the influence of other risk factors on the decline.

The studies chosen were, for the most part, published in the pulmonary literature rather than in statistical journals. Material relevant to pulmonary decline in adult males has been emphasised. The columns of Tables 3.1 to 3.12 are each set up with reference to the

following column headings:

Study: The first author and year of publication given.

Characteristics: Three terms of interest are the frequency of measurements over the duration of the study; the average initial FEV₁ of the study group; and their mean age at the start of the study. With each mean value a standard deviation is indicated by an adjacent value in parentheses.

Sample: A description of the chosen study sample is followed by the number (n) that were actually studied over the follow-up period, along with the original potential study population (given in parentheses). Subgroups of interest (along with their number) may be included.

Change in FEV₁ ml/yr: The value found for the change in FEV₁ per year may be followed by a standard deviation (given in parentheses) or by a standard error of the mean (parentheses are followed by "SE"). The published decline estimate is given for the total study group unless indicated by specific subsamples. Decline estimates, given in liters per year, or millilitres over several years, have been converted to millilitres per year; while an asterisk (*) indicates significant differences in the estimates of decline.

Other Lung Function/yr: Annual changes in levels of other specified variables measured longitudinally will be indicated, along with their units of measurement.

The other pulmonary measures tabulated were found in the three thesis data sets. The most common pulmonary function test used for longitudinal modelling is FEV₁ and it is the outcome measurement of choice for the comparison of methods section of this thesis. It is of interest to note how the declines of other respiratory measurements compare to FEV₁ decline, as the adjunct use of these measures may provide a more complete description of lung function decline.

Description of Tabulated Studies

Studies on Normal Populations – Tables 3.1–3.4

The study of "normal" populations usually included a subset of a residential population where the emphasis is placed on pulmonary function differences due to smoking status. No one category of technique dominates in this group of studies. One of the earliest studies of this type was performed by Wilhelmson et al. (1969), in which Swedish men born in 1913 were studied. Nonsmokers had a lesser decline of 170 ml over 4 years (42.5 ml/yr) compared to the smokers' average value of 280 ml over the same period (70 ml/yr). Huhti and Ikkala (1980) again found that smokers had a significantly higher decline than nonsmokers, with more conservative values of 51 versus 33 ml/yr over a 10 year span. Rode and Shephard (1984) concluded in their study of Canadian Inuit that those men aged 25 at the start of the study had a decline of 23 ml/yr compared to those who were 45 years of age who showed a greater decline of 54 ml/yr. However, these estimates are unstable as the size of the total group studied was comparatively small and numbers are unreported at each age level. Erikson et al. (1985) classified groups into an unusual category; those who had α 1-antitrypsin deficiency versus controls had a greater decline of 75 ml/yr in FEV₁ versus 53 ml/yr found in the controls.

An endpoint two-step analysis of a "normal" population who were not selected on the basis of disease or exposure, was published a little under a decade ago by Krzyznowski (1980). The overall FEV₁ decline of 2572 people tested in the Polish city was 68.4 ml/yr. A multiple classification analysis (which is based on linear regression) was applied to the change in FEV₁%, which was defined as initial minus final FEV₁ divided by the sum of the two. Significant predictors included age, initial FEV₁, obstructive syndrome, chronic bronchitis, dyspnea, sex, smoking, education and air pollution level.

As part of the Boston aging study, Bosse et al. (1981) compared smokers to nonsmokers and found declines of 78 versus 61 ml/yr respectively after adjusting for age by analysis of covariance. A stepwise multiple regression procedure was added, and initial FEV₁, age, and the log of the lifetime cigarette consumption, were found to be significant predictors of decline.

Tashkin et al. (1984) in comparing continuous smokers with those who never smoked also presented adjusted mean declines in which age, height and area of residence in Los Angeles were taken into account. The values of 70 ml/yr for the smokers and 56 ml/yr for the nonsmokers' FEV₁ decline were comparable to the Boston study.

For Beatty et al (1984) the unadjusted decline estimates of never versus heavy smokers was 30 ml/yr compared to 41 ml/yr. The authors concluded that the intrinsic factors, such as initial FEV₁, height, and time between visits were unnecessary to include in the model as, in no case did adjustment for these intrinsic covariates alter the results of the analysis.

Dontas et al. (1984) did have three measurement points available but only looked at the first and last measurement occasions for analysis. Values for FEV₇₅ rather than FEV₁, were adjusted for age, height and initial lung function. The presence of chronic obstructive lung disease was found significant according to an analysis of covariance. Olofsson et al. (1986) found a much larger decline of 73 ml/yr in heavy smokers compared to nonsmokers, whose decline was 47 ml/yr. Multiple regression analysis showed that the slope of phase 3 (a measure of gas distribution), smoking habits, age, and initial FEV₁, were all significant predictors of decline.

Fletcher and Peto (1977) published an analysis based on their book (1976) which presented their study of working West London males. A dose-response gradient of FEV₁ change with smoking was shown, ranging from -42 ml/yr for nonsmokers to -66 ml/yr for heavy smokers. While details of the analysis were not given in the paper, according to the

book, the FEV₁ slopes were corrected for observer and secular biases. In addition, after the deletion of certain outliers, 15 ml was subtracted from each slope to conform to expectations based on cross-sectional considerations.

Bosse et al. (1980) had only three different measurement occasions available on each subject, yet a regression analysis was performed; with equidistant intervals between measurements the same results would have been obtained using the endpoint technique. Nonsmokers had very similar declines to smokers over the ten year period. Although Townsend et al. (1985) had annual measurements done from 1974 to 1982 the presented data only covered the period from the third examination, due to previous quality problems. Analysis of variance was used to determine the significance of different smoking groups in a risk intervention trial. The difference between those who never smoked and the smokers was significant, after outliers had been deleted in the analysis.

Burrows et al. (1987) presented their analysis of adult pulmonary function measurements in terms of both absolute decline and percent change in FEV₁. In both instances, smokers had a significantly greater rate of decline than ex-smokers.

An early example of a two-step regression analysis was that of Fletcher (1968). Measurements were made at six month intervals over four to five years; groups were subdivided according to initial FEV₁, standardised to age forty and height of 170 cm. Those with a higher initial FEV₁ level showed less decline, with multiple regression analysis identifying the significant risk factors of mean FEV₁, smoking status, presence of sputum and eosinophilia.

Van der Lende et al. (1981) analysed data collected four times over a nine year period in two Dutch towns. A slight but significant difference was found between those residing in a polluted city, who had a decrement of FEV₁ of 28 ml/yr compared to those residing in a non-polluted city, whose decline was 23 ml/yr. A stepwise multiple regression revealed age,

smoking, male sex, and which city of residence as significant predictors of decline. Vollmer et al. (1987) using five measurements over an 11 year period, showed significant differences between their smokers and nonsmokers; the age adjusted FEV_1 values being 66 ml/yr and 43 ml/yr respectively.

Table 3.1: STUDIES OF NORMAL SUBJECTS: Endpoint Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function /yr
Wilhelmsen et al. (1969)	2x / 4yr FEV ₁ =3.62(0.6) Age=50 yrs	Swedish City n=313 Smokers n=88 Non-Smokers n=64	-58 -70 -42	VC = -80
Huhti and Ikkala (1980)	2x / 10yr FEV ₁ =3.28-3.53	Finnish City Smokers n=193 Non-Smokers n=77	-51(33) -33(30)	
Rode and Shephard (1984)	2x / 10yr FEV ₁ =2.39-4.44	Canadian Inuit n=69(198) 1971 Age = 25 Age = 45	-23 (9)SE -54 (8)SE	FVC = -13 (3)SE FVC = -70 (15)SE
Eriksson et al. (1985)	2x / 6yr FEV ₁ =3.53-4.05 Age=50 yrs	Swedish City antitrypsin deficient n=32 Controls n=31	-75 * -53 *	

See the beginning of Chapter 3 for a definition of symbols and terms

Table 3.2: STUDIES OF NORMAL SUBJECTS: Endpoint 2 Step Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function /yr
Krzyzanowski (1980)	2x / 5yr FEV ₁ =3.89(0.03) Age=42.4 yrs	Polish City n=2572(3688) Smokers n=607 Non-smokers n=206	-68.4(2) SE -56 (5)SE -73 (3)SE MULTIPLE CLASSIFICATION ANALYSIS Age, FEV ₁ , Obstruction, Bronchitis, Dyspnoea, Sex, Smoking, Education, air pollution.	
Bosse et al. (1981)	2x / 6yr FEV ₁ =3.58-4.18 Age=39.0-44.2 yrs	Boston, U.S.A. Smokers n=354 Non-Smokers n=398	-78 -61 (age adjusted) MULTIPLE REG (Stepwise) Initial FEV ₁ , Age, log of cigarettes smoked in lifetime	FVC = -91 FVC = -68
Tashkin et al. (1984)	2x / 5yr FEV ₁ =3.56-3.88 Age=43.5-45.1	Los Angeles, U.S.A. Smokers n=278 Non-Smokers n=414	-70 -56 ANCOVA (adjusted by Age, height, area of L.A.)	<u>FVC</u> <u>FEF2575</u> -64 .122 -60 -86
Beatty et al. (1984)	2x / 4.7yr FEV/FVC = 77% Age=43.6 yr	Baltimore, U.S.A. Smokers n=1019 Never n=207	-40.78 -30.07 MULTIPLE REG Age, Smoking	FVC = -43.74 FVC = -24.07
Dontas et al. (1986)	3x / 10yr	Crete Smokers n=82 Non-Smokers n=137		<u>FEV</u> <u>0.75</u> <u>FVC</u> -37 -61 -31 -62 ANCOVA (adjusted by age, ht, initial lung function, presence of cold
Olafsson et al. (1986)	2x / 7yr FEV ₁ =3.30-3.69	Swedish City Smokers n=48 Non-Smokers n=94	-73 -47 MULTIPLE REG slope of phase III, smoking, age, initial FEV ₁	

See the beginning of Chapter 3 for a definition of symbols and terms

Table 3.3: STUDIES OF NORMAL SUBJECTS: Regression Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function /yr
Fletcher and Peto (1977)	17x / 8yr	London n=792(1136) Smokers n=180 Non-Smokers n=103	-66 (4) SE -42 (6) SE	
Bosse et al. (1980)	3x / 8.2-9.8 yr FEV ₁ =3.31-3.50 Age=43.7 yr	Aging Study Smokers n=32 Non-Smokers n=132	-70 (38) -66 (48)	FVC = -91 (51) FVC = -82 (56)
Townsend et al. (1985)	9x / 8yr Age=46.9(6.0) yr	Risk Intervention n = 6123(12866) Smokers n=1845 Non-Smokers n=1039	-55.2 (83.1) -63.8 (85.3) * -53.4 (81.2) *	
Burrows et al. (1987)	5.4x / 10yrs FEV ₁ =97.1 (14)% Pre Age=45.3 (13.6)	Tucson, U.S.A. Smokers n=57 Ex-Smokers n=79	-23.4 (40.3) * -3.6 (38.7) *	
See the beginning of Chapter 3 for a definition of symbols and terms				

Table 3.4: STUDIES OF NORMAL SUBJECTS: Regression 2 Step Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function /yr
Fletcher (1968)	>4x / 4.5yr	London, England n=904 (1136)	-26.0 MULTIPLE REG Mean FEV ₁ , Smoke, Sputum, Eosinophilia	
Van der Lende et al. (1981)	4x / 9yr	Dutch Towns Polluted n=131 Nonpolluted n=28	-25.6 (24.2) * -13.9 (15.3) * ANCOVA + MULTIPLE REG (Stepwise) Age, Smoking, Male Sex, City	
Vollmer et al. (1987)	5x / 11 yr Age=47 (15) yr	Screening n=715(1024) Smokers Non-Smokers	-66 (30) * -43 (26) * MULTIPLE REG (Age adjusted)	
See the beginning of Chapter 3 for a definition of symbols and terms				

Very few studies which focus on occupational or industrial exposures, are carried out beyond a follow-up basis of two repeated measurements. The majority of studies tabulated in this category are based on endpoint analysis only. The first such study cited is that of Higgins et al. in 1968; only the decline in FEV_{75} , was tabulated according to very general occupational groupings. Over an 18 month period Peters et al. (1970) found an extremely high level of FEV_1 change of 220ml (-147 ml per annum). Another short study duration of two years was conducted by Fox et al. (1973). An extreme difference in the 2 year decline of FEV_1 of 28 ml was found for workers with grade 0 byssinosis at the start of the study compared to 163 ml found for those cotton operators having the grade 2 category. While no test of statistical significance of these results was applied an advantage to the study was the very large number of men followed (421 in all). The shortest duration of follow-up presented among all the studies was the one year study by Peters et al. (1974). A very large group (1430) of fire fighters was followed and declines were grouped in accordance to the number of fires fought. Those in the lowest exposure category had an average FEV_1 decline of 49 ml/yr in comparison to those who had fought over a hundred fires that year, whose decline estimate differed significantly at 109 ml/yr.

Over a four year period Hall et al. (1975), in addition to looking at the change in FEV_1 among only 17 coal miners studied, also looked at the change in FVC, the $FEV_1/FVC\%$ ratio, RV, TLC and RV/TLC. This scope of measurements was rarely attempted in any other occupational based study. A control group was included but it consisted of only 6 individuals.

Pulmonary function measurements were conducted three times over the course of a 7 year study by Bouhuys and Zuskin (1976). A two-point analysis revealed a small difference of 53 versus a 47 ml/yr decrement between hemp workers and controls which tested as

non-significant. Wegman et al. (1977) demonstrated a dose-response gradient of exposure for their toluene diisocyanate workers. The gradient was maintained even after attempting to standardise each decline by dividing it by the initial FEV₁. Graham et al. (1981) redid the study of granite workers analysis presented by Musk et al. (1977) and found an increment in FEV₁ of +6 ml/yr. Doubt was cast on these values due to the short expiratory time of the previous measures. Saric et al. (1982) studied a diverse number of workers who were grouped according to five categories (3 of which are tabulated). Of 95 aluminum workers tested (whose average age was 30.2 years) the FEV₁ decline of 21 ml/yr contrasted to a much larger decline of 93 ml/yr seen for ship yard workers (whose average age was 45.6 years).

Poukkula et al. (1982) compared smokers versus nonsmokers at a pulp mill and found a significant difference in their decline of 49 versus 37 ml/yr. Over a five year follow-up period Rom et al. (1983), found a discrepancy in that smokers had a decline of 17 ml/yr in a longitudinal analysis, as opposed to a cross-sectional estimate by age where a much greater decline of 48 ml/yr was predicted. Douglas et al. (1985) studied a large group (890) of London firemen found an opposite trend, a longitudinal analysis of FEV₁ decline of 90 ml/yr contrasted to a much smaller predicted decline based on age of 27 ml/yr.

A decline in FEV₁ of 55 ml/yr was found among copper smelter workers by Smith et al. (1977). Significant predictors of decline using analysis of covariance, were the level of sulphur dioxide, initial FEV₁ as percent predicted, cigarette status and height. However, the follow-up duration was only 1 year. Kauffmann et al. (1979) on the other hand, followed up 575 factory workers over a 12 year period, and the overall decline of 47 ml/yr found was further analysed with respect to such significant predictors as smoking, occupational exposure, social class and FEV₁/height³. Sparrow et al. (1982) studied a group of fire fighters and also included controls. Among nonsmokers a decline of 81 ml/yr for fire fighters contrasted to 64 ml/yr for controls. These declines already were adjusted for age, height and initial lung function; raw values were not available. According to the additional multiple regression

analysis whether or not the subject was a fire fighter further contributed 12 ml/yr of decline after the overall effects were removed.

Three measurement occasions were available in the analysis by Love and Miller (1982) of coal miners. They chose to define the change in FEV_1 as the first minus the third FEV_1 measurement divided by the number of years in between, times eleven, in order to standardise it proportionally to an average follow-up period of eleven years. Calculation of slope in ml/yr would have been a more intuitively obvious approach allowing comparability. The large number of coal miners followed, (1677 men) had an overall decline of 46 ml/yr. The sequential multiple regression analysis furthermore showed age, height, smoking status, colliery worked in and previous dust exposure to be significant predictor variables. Yet the total proportion of variance (R^2) accounted for was only 6.1%. The study by Beck et al (1983) elaborated the previous published results of cotton textile workers (presented in 1982) in which the results were given in terms of residual lung function which complicates any comparison with other analyses. Multiple regression done in a forward stepwise manner showed that a significantly larger decline was observed for men, and for persons smoking at both examinations, and for cotton textile workers in comparison to controls. A subgroup of male cotton textile workers who had no impaired pulmonary function status, showed that those who were retired by follow-up had a greater decline (at 53 ml/yr) than those who remained active at both measurement occasions (22 ml/yr). This study is one of the few that determined the pulmonary function status of retired workers. The thesis data set on *COAL* miners has a further advantage in that it not only contains the year in which the worker retired but also continues with repeated follow-up measurements.

Cotes et al. (1983) measured both changes in FVC as well as in vital capacity in a group of 79 beryllium workers. While the FVC showed a decrement of 30 ml/yr, vital capacity increased by 21 ml/yr. Ames et al. (1984) in a five year follow-up study of Western U.S. miners found no significant difference in the decline of diesel compared to

non-diesel underground workers, even when a dose-effect model of diesel exposure was attempted. A seemingly large difference of a change of -107 ml/yr versus that of -57 ml/yr for longer exposure, was found not significant. These changes were made under simultaneous adjustment for height, weight, age and cigarette smoking status.

Finally, in the study by Buist et al. (1986) there were five measurement occasions available on loggers but a pairwise analysis was chosen, due to the variability of the measurements. A dose-response gradient was observed for changes in FEV_1 between 1980 and 1984 for the different volcanic ash exposure groups. But more importantly, when the decline between 1980 and 1981 only was analysed much larger decline values of up to three times the magnitude were found. The confidence intervals that were given were adjusted for age, height, smoking status and job category. The study emphasised the value of studying groups for a relatively long duration of time and having more than two points for analysis; although the authors chose only to use pairs of measurements for analysis.

Howard (1970) published one of the few occupational studies that had more than one set of repeated measurements. Industrial workers were studied over a period of 9 to 11 years of follow-up, and analyzed data for those measured between three to five times. $FEV_{.75}$ rather than FEV_1 was measured, and emphasis was given to describing unusual stepwise decline. Flood et al. (1985) had abundant data, where measurements were taken every 6 months from 1969 until 1975, and then followed by annual follow-up. The results presented in the table are taken after 1972 as the pre-1972 values were found to be variable, and were taken when the study measurement procedures were unstandardised between the various factories. Rather than choosing to do a multivariate analysis, tables are presented with regression coefficients grouped by risk factors such as degree of exposure, skin prick sensitivity and atopy.

Berry et al. (1973) analysed 3 years worth of data on cotton mill workers, which was collected at 6 month intervals. The minimum criteria for inclusion in the analysis was having at least three data points and a duration of over 18 months of follow-up. While males showed a decline of 55 ml/yr, when adjusted for sex age, smoking, health and exposure using an additive least squares model, the values changed slightly to 58 ml/yr. Unfortunately, very little baseline information is given such as the average age and FEV₁ values of the cotton workers. Diem et al. (1982) also applied an eligibility criterion that each toluene diisocyanate worker must have three of nine possible data points to be entered in the analysis. The overall decline of the group of 24 ml/yr had both the standard deviation and the standard error of the mean given. In addition, coefficients of decline were given for FVC, MMF, DLCO, RV and TLC, with the latter two showing an increase over the interval. The multiple regression procedure used was weighted according to the precision of the slope. The age effect contributed a decrement of 6 ml/yr, while low FEV₁/Height³ contributed 20 ml/yr and the exposure effect in never smokers was 38 ml/yr. All these risk factors were significant.

Eisen et al. (1984) studied both survivors and drop-outs amongst granite workers. They offered a unique analysis in that they log transformed the FEV₁ slopes in order to stabilize the variance and to make the distribution symmetric. In addition to demonstrating a survivor effect (drop-outs having larger decline) for those whose lung function tests were rejected because of having two test failures, a coefficient of decline of 81 ml/yr was found in comparison to 46 ml/yr for those with no test failures. This was significant even after controlling for age, height, silica exposure and current smoking status, as determined by the application of a multiple regression model.

Table 3.5: STUDIES OF OCCUPATIONALLY EXPOSED SUBJECTS: Endpoint Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function /yr
Higgins et al (1968)	2x / 9yr FEV _{0.75} × 40 = 2.22-2.36 Age=55-64	n=594(756) Miners and Ex-Miners n=149 Dust or Chemical n=92		FEV _{0.75} = -52 ml FEV _{0.75} = -41 ml
Peters et al. (1970)	4x / 1.5 yr FEV ₁ = 4.12	Toluene Diisocyanate n=19(38)	-147 *	
Fox et al. (1973)	2x / 2yr	Cotton Operators n=421 Grade 2 Byssinosis Grade 0	-81.5 -14.0	
Peters et al. (1974)	2x / 1yr FEV ₁ = 3.578 Age = 43.13 yr	Firefighters n=1430(1768) 100+ fires 1-40 fires	-68 -109* -49 *	FVC = -77 ml FEV/FVC = -0.01
Hall et al. (1975)	2x / 4yr FEV ₁ = 3.26 Age = 57 yr	Coal Miners n=17(25) Controls n=6	-20 -25	Coal Miners FVC = +8 ml FEV/FVC = .5% RV = +90 ml TLC = +88 ml RV/TLC = +1%
Bouhuys and Zuskin (1976)	3x / 7yr FEV ₁ = 2.01	Hemp Workers n=58(219) Controls n=19(167)	-53.3 -47.1	
Wegman et al. (1977)	2x / 2yr Age = 30.9yr	Toluene Diisocyanate n=57(112) High n=20 Low n=20	-51 -103 -6	
Ferris et al. (1979)	2x / 10yr FEV ₁ = 100-104%	Pulp and Paper Chlorine n=48 SO ₂ n=61 Paper n=91	-30 -44 -43	FVC -25 ml -42 ml -31 ml
Graham et al. (1981)	2x / 5yr FEV ₁ = 3.51(0.7)	Granite n=487	+6	FVC +108
		>20yr exp. n=242	-7	+120

Saric et al. (1982)	2x / 3-4yr FEV ₁ =3.34-3.95 Age=30.2-51.6 yr	Aluminum n=95	-21	FVC ml
		Cement n=96	-42	-190
		Shipyard n=38	-93	-136
				-91
Poukkula et al. (1982)	2x / 10yr	Pulp Mill		
		n=816 (905)	-44 (33)	
		Smokers n=258	-49 *	
		Non-smokers n=123	-37 *	
Rom et al. (1983)	2x / 5yr FEV ₁ =4.12 Age=37.6	Trona Dust		
		n=125(230)		
		Smokers	-17	
		Non-Smokers	+14	
Siracusa et al. (1984)	2x / 7yr FEV ₁ =3.87±.09 Age=38.3±1.1 yr	Asbestos Cement		FVC ml
		n=65(77)	-49 (10) SE	-48 (10) SE
		PVC n=30(34)	-9 (16) SE	
Douglas et al. (1985)	2x / 1yr Mean=12.4 months FEV ₁ = 3.71 - 5.27	London Firemen		
		n=890(1006)	-90	FVC=-110 ml
				FEV ₁ /FVC=-0.05%

See the beginning of Chapter 3 for a definition of symbols and terms

Table 3.6: STUDIES OF OCCUPATIONALLY EXPOSED SUBJECTS: Endpoint 2 Step

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function yr
Smith et al. (1977)	2x / 1yr FEV ₁ =3.682 Age=36.0 yr	Copper Smelter n=113(268)	-54.6(5.1) ANCOVA (ht adjusted) SO ₂ , Initial FEV ₁ , Smoking, Height	FVC = -39.8(22.2)
Kauffmann et al. (1979)	2 x (12 yr) FEV ₁ =3.58 Age=41(7) yr	Factory Workers n=575(2154)	-47 MULTIPLE REG Smoking, Occupation, Exposure, Social Class, FEV ₁ /Ht ³	FVC = -40 ml FEV/VC = +4(.7)%
Sparrow et al. (1982)	2x / 5yr FEV ₁ =3.94 Age=41.4 yr	Never Smokers Firefighters n=50 Other Non-Smokers n=447	-81.2(19.2) -64.1(39) (adjusted for age, ht, initial FEV ₁) MULTIPLE REG Initial FEV ₁ , Age, Height, Current smoker, Firefighter	FVC ml -76.8(10.7) -71.3(4.1)
Love and Miller (1982)	3x / 10-12yr FEV ₁ =3.06 Age=45.3 yr	Coal Miners n=1677(6191)	-46 MULTIPLE REG (Sequential) Age, Height, Smoking, Colliery, Previous Dust Exposure	
Beck et al. (1983)	2x / 6yr Age=59yr	Cotton Textile Workers n=383 Controls n=277	-42 -27 MULTIPLE REG (Forward Stepwise) Male Sex, Smokers, Cotton Workers	
Cotes et al. (1983)	2x / 10yr FEV ₁ =3.22-3.52 Age=40.4-48.4 yr	Beryllium n=79(146)	-34 MULTIPLE REG Smoking, Exposure	FVC=-30 ml VC=+21 ml RV=+16 ml
Ames et al. (1984)	2x / 5yr FEV ₁ =3.9 Age=36.1 yr	Underground Miners Diesel n=280 Non-Diesel n=838	-24.5 -30.6 (Age adjusted) LEAST SQUARES MEANS Age, Height, Weight, Smoking	FVC ml -19.8 -43.6
Buist et al. (1986)	5x / 4yr FEV ₁ =100-109%pre Age=37 yr	Volcanic Ash Exposed Workers High n=120 Control n=179	-62(59) -45(52) ANCOVA Age, Height, Smoking Status, Job Category	

See the beginning of Chapter 3 for a definition of symbols and terms

Table 3.7: STUDIES OF OCCUPATIONALLY EXPOSED SUBJECTS: Regression Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function /yr
Howard (1970)	3-5x / 9-11yr FEV ₁ =2.78(.63) Age=42(9) yr	Industrial Workers n=159(289)		FEV _{0.75} =34(38)ml FVC=64(41) ml
Flood et al. (1985)	18x / 11yr >4x / >18mos	Detergent Industry Factory A n=789 Factory C n=646	-22(3) SE -53(1) SE (post 1972 results)	

See the beginning of Chapter 3 for a definition of symbols and terms

Table 3.8: STUDIES OF OCCUPATIONALLY EXPOSED SUBJECTS: Regression 2 Step Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function /yr
Berry et al. (1973)	6x / 3yr >2x / 18 mos	Cotton Mills n=226	-55(6) SE ADDITIVE MODEL Sex, Age, Smoking, Bronchitis, Mill Type, Job	
Diem et al. (1982)	9x / 5yr >3x / 4.1yr FEV ₁ =100% Pred	Toluene Diisocyanate n=223(277)	-24.4(25.5) (3.1) SE WEIGHTED MULTIPLE REG Age, Low FEV ₁ /Ht ³ , exposure among never smokers	FVC=-12.1 ml FEF2575=-92.8 l/s -DLCO=0.716 ml/min mmHg RV=+43.5 ml TLC=+32.2 ml
Eisen et al. (1984)	6x / 5yr FEV ₁ =95(.6)% Age=40.2(.4)	Granite Complete n=515 Drop-out n=103	-47.8(3.9) SE -47.8(3.9) SE -69.4(12.6) SE MULTIPLE REGRESSION Age, Height, Silica, Exposure, Smoking	
See the beginning of Chapter 3 for a definition of symbols and terms				

Studies of diseased populations generally provide the most sophisticated approach to the design and analysis of longitudinal studies of pulmonary function. Comparatively few were based on a follow-up limited to two points in time. Schachter et al. (1984) reported a very small annual FEV_1 decline of 6 ml/yr among his controls compared to a decline of 24 ml/yr found among asthmatics. Petty et al. (1976) likewise found that his "more normal" group of subjects who had at least a 75% FEV_1/FVC ratio, had a small annual decline of 4 ml/yr in contrast to 41 ml/yr found for those patients with less than a 60% value of FEV_1/FVC . Interestingly enough, the FVC among those with worse lung function patients in that study declined at a very high rate of 106 ml/yr. The study of Johnston et al. (1976) unlike the previous two, involved four measurement occasions; however, only the difference between the first and last measurement occasion was analysed. The chest clinic patients showed an overall decline of 34 ml/yr in FEV_1 . An increase of 5.1% for the ratio of residual volume to total lung capacity (RV/TLC) was found. Very few studies have looked at the changes in residual volume over time, which from theoretical considerations would be expected to increase as pulmonary function declined. Examples of the two step approach to two-point analysis were not found in studies of diseased individuals.

The earliest, as well as the most extensive use of regression analyses of pulmonary function decline, is observed for studies of diseased patients. One of the very first published studies regarding longitudinal regression analysis of pulmonary function was that on chronic bronchitics by Fletcher and Oldham (1966). FEV_1 decline was compared among patients given certain therapies or combinations thereof versus those given no treatment, using regression analysis on up to 25 measurements per subject. A reasonable dose/response relationship emerged where those given no therapies declined at 79 ml/yr as compared to 90 ml/yr for those with a combination of therapies. The authors recognised that averaging individual coefficients would give undue emphasis on those calculated from the smallest number or

smallest duration of the measures. An attempt was made to incorporate these variables in the estimates by weighting them accordingly. The maximum difference from the unweighted coefficients of decline however averaged to be only 5 ml/yr less. Other studies which present estimates based on regression analysis only, merely presented the averages of the regression coefficients.

The analysis presented by Howard (1967), focussed solely on $FEV_{0.75}$ and both the standard error of the estimate (SEE) as well as the standard error of the slope was presented. Both measures have been in more common use as a standard output of a regression package program. An interesting aspect of Howard's study is the separate derivation of an estimate of decline for those 13 who had died in the follow-up period from 1962 to 1965, which showed a trend of having a higher value of 93 ml/yr (versus 83 ml/yr for all of the 125 patients). In accordance to these results, the decline estimates for the coal miners data will be estimated separately for those who had died during the period of the study follow-up.

Jones et al. (1967) presented the results of a follow-up of patients with irreversible airway obstruction, over a period averaging under 3 years. For the studies of diseased patients a period of follow-up of less than five years was found for only two other investigations; that of Barter et al. (1974) and Anthonisen et al. (1986). On a theoretical basis, a duration of follow-up of less than five years will result in less precise and stable estimates of decline (Clement and Woestijnje, 1982). While Jones et al. (1967) had presented separate estimates of decline for two cities for FEV_1 and VC, no statistical significance of the comparison was given. In a separate study of chronic airway obstruction patients, Burrows and Earl (1969) studied patients who also had a small initial FEV_1 value which averaged about one liter. A decline of 56 ml/yr in FEV_1 was found, while the RV/VC ratio (as opposed to RV/TLC) was found to be increasing at 0.5%/yr.

The analysis by Emirgil and Sobol (1971) presented the most innovative approach to longitudinal data analysis. Chronic obstructive pulmonary disease patients were followed from 2 to 13 years (the average being 5 years), with an average of 6 measurement occasions being available for analysis. In an unusual form of presentation of the data, the patients were separated into categories according to initial maximum mid-expiratory flow measurement values. Regression analysis revealed FEV_1 decline estimates ranging from 109 ml/yr, for those whose original MMF were greater than 1.0 l/sec to a value of 66 ml/yr for those with under 0.5 l/sec originally. In an alternate analysis, modelling was based on a log-linear relationship which showed an opposite trend; the fall in percentage of FEV_1 change per year ranged from 6.3% to 8.7%, the latter being for the more impaired group. An informative result of the analysis is the calculation of the number of years required to fall to 50% of the initial value. For those with the better original MMF values the value was 14.6 years, compared to 9.7 years for those with the worse initial values.

As is often the case with chronic obstructive lung disease, bronchodilator response is evaluated as part of a therapeutic regime. In the studies by Barter et al. (1974, 1976) the FEV_1 measurements analysed were those taken after the bronchodilator treatment; the pre-bronchodilator values were considered more variable. The group of obstructive airway patients studied by Howard (1974) appeared to be very similar in initial characteristics to Jones et al. (1967). Both groups were on average 58 to 60 years of age and had initial FEV_1 measurements averaging 0.93 l. But the FEV_1 decline found for Howard's group, of 23 ml/yr was roughly half the magnitude. Some indication of the source of this difference is that the results of Jones and colleagues were of a combined estimate from patients of two cities; a 16 ml/yr average difference in decline distinguished the two.

In addition to an FEV_1 decline of 54 ml/yr and a decrement in VC of 89 ml/yr, Postma et al (1979) also looked at the change in the ratio of FEV_1/VC . An unexpected increment of 3% per year was found over the study duration of 11 years. The problem

apparent in examining an expected decline of such a ratio is that a comparatively smaller decline in FEV_1 compared to VC could result in such an estimate; even though the absolute value of this particular ratio has value, on a cross-sectional basis, for revealing abnormality.

Rather than reporting raw coefficients obtained from regression of the values of FEV_1 , Hughes et al. (1982) chose to weight the coefficients according to the squared differences between the average time and the actual time of each follow-up value. For this group of emphysematous males, declines found for FEV_1 and VC for the smoking groups were very similar. The most recent study to use regression analysis alone, was that of Burrows et al. (1987). The follow-up period covered 10 years, however information on the actual number of measurements on the airway obstruction patients is lacking. Estimates of decline were obtained according to more specific diagnoses. Thus, for example while chronic asthmatic bronchitics had a value of 5 ml/yr decline in FEV_1 the COPD emphysematic patients had a decline estimate of 70 ml/yr. The difference in decline was significant and the test of significance used for comparison of the groups was analysis of variance; such information is often lacking in other published studies.

One of the earliest investigations which used a two-step regression approach was done by Kanner et al (1979). They used stepdown linear regression to assess the significant risk factors for the coefficient of FEV_1 decline. In their study of chronic bronchitic patients it was found that the presence of lower respiratory illness in the past had the greatest magnitude of effect on the FEV_1 slope, that of -62 ml/yr. Other significant predictors included airway reactivity, years smoked, age, as well as α_1 -antitrypsin level. The finding of a mean decrement of 69 ml/yr of FEV_1 applied only to a subset of their original population, in that outliers were deleted based on the slope of the middle one third of the observations. Inclusion criteria were also adopted such that a minimum of 5 measurement occasions and 2 years of follow-up was applied to the study population. Hughes et al. (1982b) chose to remove one outlier only, in that this person demonstrated an unusually high

decline of 1470 ml/yr in his FEV₁. Silicotic sandblasters were compared on the basis of their x-ray results, as to whether they showed progression or not. A relatively larger decline in lung function for those who showed progression was found for FEV₁, FVC, the FEV₁/FVC ratio and TLC. A smaller decline or a greater positive slope was shown for FEF₂₅₋₇₅, RV, RV/TLC and diffusing capacity. Upon using multiple regression analysis on the obtained slope it was found that total silica exposure was significant after adjusting for initial level of FEV₁.

Kanner's 1984 analysis was based on the previous 1979 data. Those with both bronchitis and emphysema had a significantly increased decline for FEV₁ and FVC, as assessed by analysis of variance. Initial FEV₁ was an additional risk factor of interest, but in their study this was found not to be a significant predictor.

The silicosis patients studied by Bucca et al. (1985) had an initial FEV₁ of 2.77 l on average, which is much higher than that typically found among patient groups. An overall decrement in FEV₁ of 41 ml/yr was found to be significantly related to the initial FEV₁/Height³ ratio, silica exposure and to the presence of chronic obstructive pulmonary disease. In the study by Campbell et al. (1985) a major risk factor studied among their chronic bronchitic population was the area of residence. The rate of FEV₁ decline was found to be significantly less for the Queensland residents than for the residents of New South Wales or Victoria. The linear analysis took advantage of both multiplicative and additive relationships. The resulting regression equation (which had an excellent R² of 61%, in comparison to typical published values of less than 10%) was; The rate of decline of FEV₁ (ml/yr) = -111 - 32.37(If Q Resident) + 11.02(If Vic Resident) + 18.48x(Response to Bronchodilator) + 28.59(Exposed to Dust) + 1.712x(Current Smoking) - 0.2393x(FEV₁/VC% x Response to Bronchodilator(%)) + 2.213x(Age).

An unusual analytical approach was suggested by Anthonisen et al. (1986), in that baseline FEV_1 was not used in calculating the slope because it was one of the selection criteria for entrance in the study and might bias the slope as result of regression to the mean. The initial values and categorical groups were presented in terms of percent predicted FEV_1 in comparison to a normal population. The unusual risk factors determined to be significant by multiple regression were post/pre $FEV_1\%$, wheeze and social scale. Finally, Ng et al. (1987) compared granite quarry workers who had simple and progressive silicosis. The unadjusted means coefficients of -64 ml/yr versus -97 ml/yr are almost identical to the means that were adjusted for age, pack years smoking and initial lung function. Multiple regression of these adjusted means showed tuberculosis history and silica concentration to be further predictors of decline.

Table 3.9: STUDIES OF DISEASED SUBJECTS: Endpoint Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function yr
Johnston et al. (1976)	4x / 10yr FEV ₁ =1.65 Age=52 yr	Chest Clinic n=54(111)	-34 (33)	VC=-72 (25) RV/TLC=+.51
Petty et al. (1976)	2x / 6-7yr Age=47.1 yr	FEV ₁ /FVC<60% n=25 FEV/FVC > 75%	-41 -4 ml	FVC=-106ml FVC= +3ml
Schachter et al. (1984)	2x / 6yr	N=1303 Asthmatic n=29 Non-Asthmatic n=361	-24 -6.3	
See the beginning of Chapter 3 for a definition of symbols and terms				

Table 3.10: STUDIES OF DISEASED SUBJECTS: Regression Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function /yr
Fletcher & Oldham (1966)	25x / 5 yr FEV ₁ =2.032 Age=36-59 yr	Chronic Bronchitics Prophylaxis and Therapy n=91(120) None n=99(122)	-90 -79	Weighted -86 ml -74 ml
Howard (1967)	2-13yr Mean 7.1 yr FEV ₁ =1.2(.165) Age=60.7(9.86) yr	Clinic Patients n=125 Deaths n=13		FEV ₁ 0.75x40 =-83.4(67) ml = -93 ml
Jones et al. (1967)	4x / 3yr FEV ₁ =0.93 Age: 58.1 yrs	Airways Obstruction London n=38(50) Chicago n=38(50)	-37 -53	VC=-160 ml VC=-80 ml DLCOb (Chicago) -2.6 ml/min/mmHg
Burrows and Earle (1969)	1-7x / 1-7yr FEV ₁ =1.0(.4) Age=59.1(8.3) yr	Airways Obstruction n=171(200)	-56	VC=-86 ml MMF=-17 ml/s RV/VC=.5% DLCOb=-1.6 ml/min/mmHg
Emergil and Sobol (1971)	6x / 2-13yr FEV ₁ =1.48 Age=59 yr	COPD n=52(91)	-8.6(5.5)%/yr 11.3(8.5)yrs to fall 50% EXPONENTIAL REGRESSION	MMF=-11.4(5.9) %/yr 7.1(3.7)yrs to fall 50%
Barter et al. (1974)	5x / 2yr Age=51yr	Chronic Bronchitics n=56(110)	-50(60)	
Howard (1974)	>5/>1yr FEV ₁ =.93(.55) Age=59.6(9.0) yr	Obstructive Airway n=144(178) (20 females)	-23(53)	FVC=-69(107)
Barter and Campbell (1976)	6x / 5yr Age=56(6) yr	Mild Chronic Bronchitis n=34(110)	-46(57)	
Postma et al. (1979)	>4x / 11yr FEV ₁ =.61 Age=54(9.1) yr	COPD n=129(138) (26 females)	-54	VC=-89 ml * FEV ₁ /VC=+3%
Hughes et al. (1982)	3-13x / 5.6 yr FEV ₁ =1.17-1.44 Age=53.7-56.6 yr	Emphysema Ex-Smokers n=19 Smokers n=37	-16.4(8.8) SE * -53.5(5.4) SE * (Weighted)	VC=-14.9 ml VC=-53.1 ml
Burrows et al. (1987)	10yrs FEV ₁ <65% Age=40-74 yr	Asthmatic Bronchitics n=27 (19 female) COPD-Emphysema n=45 (16 female)	-4.6(35.8) * -70.2(85.6) *	

See the beginning of Chapter 3 for a definition of symbols and terms

Table 3.11: STUDIES OF DISEASED SUBJECTS: Regression 2 Step Analysis

Study	Characteristics	Sample	Change in FEV ₁ ml/yr	Other Lung Function /yr
Kanner et al. (1979)	>5x/>2yr FEV ₁ =1.276-2.989 Age=52.3(6.3) yr	Chronic Bronchitis n=84(190)	-69.2 STEPDOWN REG alpha ₁ antitrypsin, age, years smoked,airway reactivity, lower respiratory illness	FVC=-93.4 ml
Hughes et al. (1982b)	2-8x / 1-7yr FEV ₁ =73% Pre Age=43.8 yr	Silicosis n=61(83)	-114 MULTIPLE REG Silica exposure	FVC=-116 ml FEF2575=-139ml/s
Kanner (1984)	See Kanner (1979)	Bronchitis & Emphysema n=49 Chronic Bronchitis n=14	-90.0(79.1) * -12.2(93) * STEPDOWN REG See Kanner et al. (1979), initial FEV ₁ nonsignificant	FVC=-129.7 * FVC=-8.6 *
Bucca et al. (1985)	4x / 9yr FEV ₁ =2.77(.58) Age=60.7(26.1)	Silicosis n=90	-41 REGRESSION (GLM) FEV ₁ /Ht ³ , Silica exposure, COPD	FVC=-35 ml
Campbell et al. (1985)	>4x / 4-6yr FEV ₁ =2.65(.46) Age=53.5(6.4) yr	Chronic Bronchitis n=66(96) Queensland n=20 N.S.W. n=17 Victoria n=29	-58.8(57.4) -25.8(40.8) -65.8(64.1) -77.1(55.0) REGRESSION (GLIM) State, Bronchodilator response %, dust exposed, smoking, age, FEV ₁ /VC% x bronchodilator response.	
Anthonisen et al. (1986)	>4x/2.5-3yr FEV ₁ =36.1 %pred Age=60.9(7.7) yr	COPD, not Asthma n=985	-44(129) MULTIPLE REG Post/Pre FEV ₁ %, wheeze, psychosocial scale	
Ng et al. (1987)	2-4x / 2-10 yr FEV ₁ =2.39(.525) Age=52.9(5.2) yr	Silicosis Progressive n=24 Simple n=29	-97(9) SE * -64(9) SE * MULTIPLE REG Progression, Tuberculosis history, silica concentration	FVC=-95(13)* SE * FVC=-59(7) * SE *

See the beginning of Chapter 3 for a definition of symbols and terms

Discussion of the Literature Background

A distinguishing characteristic of the studies on "normal" populations was the large samples of subjects followed, which allows the estimates of FEV₁ decline to be more readily generalized. The tabulated estimates range from a 4 ml/year decrement in FEV₁ for Tucson nonsmokers (Burrows et al. 1987) to a high of 78 ml/year for Boston smokers (Bosse et al. 1981). Within a given study, smokers were consistently shown to have a greater FEV₁ decline over time than nonsmokers, although the significance of these differences were often not tested. For smokers the FEV₁ decline estimates varied from 23 to 78 ml/year while for nonsmokers it was from 4 to 61 ml/yr.

Studies of occupational exposures have relied in general on only two points of data collection. Any within-subject error could greatly affect the value of the decline estimated from this data. In comparison to the studies of diseased subjects, obtaining subject co-operation consistently in an occupational setting can be very difficult. This is particularly the case for those that leave employment. Often less than three-quarters of the original study population have been followed-up at least two times. In those instances where a large number of the workers have been studied, the follow-up period is usually short in duration, although one exception is the study by Poukkula et al. (1982) in which 816 of 905 original pulp mill workers were followed, from 1967 to 1977.

The estimates of FEV₁ decline have ranged from a low of 1 ml/year over a 2 year period among cotton operators (Fox et al. 1973) to a high of 147 ml/year, found over a 1.5 year period, in a study of toluene diisocyanate workers (Peters et al. 1970). Both extremes were obtained from studies where testing was conducted during two measurement occasions only, and over a relatively short time period. Two nonsmoker subgroups actually showed an increment in FEV₁ over the study interval; both were results of two point studies (Poukkula et al. 1982; Rom et al 1986). A range of published estimates are presented for specific

occupations which were represented at least twice:

Table 3.12: Range of Published Ranges of FEV₁ Decrement For Occupational Studies

Occupation	Range of Annual Declines of FEV ₁
Coal miners	20-55 ml
Toluene diisocyanate workers	6-147 ml
Cotton workers	1-82 ml
Firefighters	49-109 ml
Copper smelter workers	5-55 ml
Granite workers	6-81 ml

The explanation for the discrepancies in the FEV₁ decline estimates can be attributed both to the study design (frequency and duration of measurement; statistical technique) as well as differences in the composition of the occupational group (intensity and duration of exposure; selection characteristics). For example, the extreme values found for firefighters described a comparatively high exposure group (>100 fires fought) in contrast to a low one (1-40 fires fought) (Petersen et al. 1974). The values however, can best be regarded as unstable as the endpoint analysis was based on a study of only one year in duration.

The study of chronic obstructive disease has provided much opportunity for the application of techniques which utilize all points of measurement to describe the time related decrement in lung function. Estimates for the coefficient of decline in FEV₁ range from 16 ml/yr for emphysema patients who were ex-smokers (Hughes et al. 1982) to 114 ml/yr for silicotic patients (Hughes et al. 1982b). The focus for the majority of the pulmonary literature was FEV₁ decline. Among those studies of diseased subjects, many additional pulmonary function variables were measured, and the interrelationships of these variables with FEV₁ decline deserve further consideration.

In spite of the considerable amount of effort expended to study FEV₁ decline in various populations, inferences from the data remain unclear. Interpretation of the literature is

hindered by incomplete description of the data base, as well as questionable approaches used in the study design and analysis.

Presentation

An inevitable result of having limited space in which to publish a study is the lack of opportunity for elaboration of the protocol. However, there are certain aspects of the study procedures that require description in order to ease the task of comparative evaluation of the studies. For example, in a study by Burrows et al (1987) a regression analysis was used for a study undertaken over a ten year period. The year the study began was not given and there was no information on the mean and range of measurements available for analysis. Ng et al. (1987) on the other hand, not only provided the date on which the study was initiated and the possible range of years follow-up for the study group, but also the mean number of years of follow-up, as well the mean and the maximum number of measurement occasions used for analysis.

The average initial FEV_1 of a study group not only gives an indication of their general state of health, but also, that of expected decline. For instance, those studies of diseased patients who had initial levels less than one litre may be expected not to show dramatic declines over the yearly periods as a levelling out of lung function to a necessary minimum is expected. The use of percent predicted (Ferris et al. 1979, Buist et al. 1986) is not as informative without reference levels for comparison between studies. Declines in FEV_1 according to subgroups of initial FEV_1 have been estimated by Fletcher (1968) and Valic and Suskin (1974), for example. Alternatively, covariance adjustment for initial level of FEV_1 has also been used, although this may result in an "overadjustment". A simultaneous attempt at height adjustment has also been made by adjusting for FEV_1/height^3 (e.g. Kauffman et al. 1979) or else by stratifying by this variable (e.g. Diem et al. 1982). The cubic term is based on dimensional relationships of volume to a linear term.

The value of presenting the numbers of subjects studied is revealed by their relative completeness of follow-up and therefore the representativeness of the subgroup studied; but also small numbers in the subgroups are generally indicative of instability of the derived coefficients of decline. Larger sample sizes are indicative of greater power for the study to be able to detect a smaller difference in the mean decline estimates to be statistically significant. As age is very often a significant risk factor in any analysis of pulmonary function decline, knowledge of at least the average age for the group as a whole, and on a more informative basis, the average age of each subgroup analysed, would be of interest, particularly if no age adjustment had been performed on the decline estimates. No indication of the age of the follow-up group has been given by Flood et al. (1985) and Peters et al. (1970). The disadvantage of stratifying by risk factors such as age is that reduced numbers in each subset will decrease the power of the study to detect a significant difference in decline among the different subgroups. Adjustment, on the other hand, generally takes place through an analysis of covariance which assumes a linear relationship between the risk factor and the dependent variable.

The presentation of estimates of decline, especially when no significance testing has been applied, necessitates an indication of the variability of the estimate, preferably in terms of the standard error of the coefficient estimated. Almost universally where significance testing was applied, the results are presented in terms of p values. Along with the knowledge of standard error, a more informative estimate in this regard is the use of confidence intervals (Buist et al. 1986) as both the interval width and value of the extremes aid in interpreting the stability of the estimate, provided that the model specification is correct.

Design issues

It appears that more effort to collect longitudinal data has been directed towards the study of diseased patients. The logistics of conducting a longitudinal study favour a population that can be traced, such as found in a hospital or registry setting. As the value of longitudinal follow-up becomes more apparent, more attention paid to the follow-up of "normal" as well as occupational populations will allow a clearer understanding of normal pulmonary function decline with aging in comparison to the longitudinal response to insults to the lungs, whether due to smoking, occupational exposures or other environmental factors. A most important facet of the study design is selecting a control group that is representative of the study base.

For a large number of studies either the follow-up period was less than six years, or follow-up was limited to two points of observation. As a result the estimates of FEV_1 decline obtained either lack precision (within-subject variability dominates) or a measure of the variability of the decline estimate cannot be calculated. Results from such study designs are therefore less credible when addressing issues of inference. A more critical factor in this regard, is the evaluation of selection factors which may affect the representativeness of the population studied.

Statistical Analysis

What techniques were used to "clean" the data in preparation for analysis deserves careful consideration of the potential for introducing bias. Two major types of statistical techniques have been used for the analysis of pulmonary function decline. When limited to two-point determinations, the choice of analysis is restricted to the linear slope obtained by calculating endpoint differences. In the case of diseased subjects, regression techniques have been used most frequently to give an average coefficient of decline. One of the early

publications of the latter type was by Fletcher and Oldham (1966) where they considered number and duration of the measurements and weighted their estimates accordingly. It was not until 1982 that published estimates of decline, again were weighted, rather than simply averaged; in this instance it was by a measure of the precision of the slope estimate (Diem et al. 1982, Hughes et al. 1982).

The first two-step approach when presenting decline estimates was published by Fletcher in 1968, using a multiple regression of various risk factors applied to FEV₁ decline. Taking account of known factors which affect FEV₁ decline through statistical adjustment, yields a more appropriate comparison of the differences found between groups. Berry et al. (1973) used the first two-step approach for an occupational study, applying least squares in an additive model to adjust the estimates by the significant risk factors found. However, it is with the studies of diseased subjects that the most sophisticated analyses have been accomplished, in that multiple regression relationships which take account of many covariates as well as the analyses of multiple dependent variables have been used. Newer developments in the statistical literature, such as the use of generalized linear and random effects models, have yet to be fully implemented in studies of adult pulmonary function decline conducted by epidemiologists and chest physicians. Taking account of the unique characteristics of longitudinal data, such as autocorrelation and regression to the mean, through covariance structure modelling is a statistical development that holds promise for future applications.

Another aspect which has been overlooked is the modelling of change of decline over time. This effect is borne out by findings of different slopes calculated by age stratification. One can interpret Emergil and Sobol's work in 1971, in which log-linear modelling was used to obtain estimates of percent fall per year, as well as the number of years to fall to 50% of the original value, as a demonstration of a non-linear estimation of the decline. Schulzer et al. (1985) in a subanalysis of the grain handlers longitudinal data demonstrated a significant quadratic age effect, particularly among smokers and the exposed grain handlers. In

none of the three types of studies tabulated was there any attempt made to look at more than a linear effect of age on decline.

The longitudinal decline of other pulmonary function measurements apart from FEV_1 and FVC, was almost exclusively the subject of studies of diseased subjects. The spirometric measures of FVC and MMF exhibited declines that were usually similar or larger than that of FEV_1 , however, the estimates of variability, particularly in the case of MMF, were also much larger. The RV, and its ratio with TLC usually showed an increase over the period; an exception was the study by Madison et al. (1981). However, the use of estimating change in ratios as the response variable, is generally of less value, because either a change in the numerator, denominator or both could be affecting any change in that ratio.

A literature review of the statistical issues and analytical options for longitudinal data analysis is presented in Chapters 4 and 5. The concluding discussion of Chapter 10 will provide recommendations for the optimal design and analysis of a longitudinal study of pulmonary function decline.

CHAPTER IV

METHODOLOGICAL PROBLEMS IN DESIGN AND ANALYSIS

Introduction

Variability in pulmonary function measurements depends on the within- and between-subject variability of the pulmonary function measurements (discussed in Chapter 2), while the precision of the decline estimate, estimated by its standard error is also dependent upon the study design and the method of analysis. Within-subject variability due to the nonreproducibility of the FEV₁ measurements on the same subject, far exceeds its expected annual decline (Burrows et al. 1986). However repeated hourly, daily and weekly measurement of pulmonary function has shown that within-subject variability is small compared to between-subject variability (Cochrane et al. 1977). Study design affects standard error by virtue of the number, the frequency and the spacing of measurements.

Design Aspects

Random measurement error, which affects the standard error of the observed rate of decline, is reflected by a wider confidence interval about the estimate. This reduces the power of the study to detect the true effect of an independent exposure variable, thus diminishing the chances of detecting differences between study groups (Dales et al. 1987). Systematic errors on the other hand have a greater effect on the external validity of the study.

Within-subject standard error may be reduced both by increasing the length of follow-up, and the frequency of observation, while between-subject standard error can only be attenuated by increasing the number of subjects. Berry (1974) observed that taking more measurements for the same duration had a small effect on the standard error, while increasing the duration reduces it considerably. The relationship was dependent upon the

calculation of the standard error of the linear decline which was formulated as the product of a ratio of the number of measurements with the between-subject standard deviation divided by the duration of follow-up. He found that for a duration of follow-up of five years, the optimal frequency of measurement to detect a difference of 30 ml per year at a significance level of 0.05 and power of 0.80, was biannual rather than annual. Because of loss to follow-up, subjects who were not available at the end of the survey would not contribute to the calculation of the rate of decline, affecting the representativeness of the sample and potentially reducing the power of the study. Dales et al. (1987) confirmed that a smoking effect on FEV₁ decline could be detected consistently only over a six year observation period. Shorter observation periods were likely lead to a false negative conclusion.

Burrows et al. (1986) assessed the effect of duration of follow-up and the number of measurement occasions, upon the variability of of computed rates of FEV₁ decline, using the standard errors of the decline estimates as the measure of variability about the individual regression lines. While additional data points did reduce the variability in the coefficient of FEV₁ decline, the effect was very small compared to the duration of follow-up.

The importance of a multisurvey longitudinal study (of at least 3 measurement occasions) as contrasted to a follow-up study, is illustrated by Diem (1982). After 2.5 years of follow-up of TDI exposed workers, an apparent increase in the annual FEV₁ of 55 ml per year was observed by the fifth survey. Extending the follow-up to 5 years with a total of 9 repeated measurements resulted in a calculated FEV₁ annual change of -21 ml per year. Higher values were found for this fifth survey, and the bias was significant. The investigators chose, however, not to exclude this survey data as they could find no explanation for the bias despite careful investigation; it was close to the mid-point of this present study and therefore had little influence on slope estimations; as well, deleting this data resulted in no change in the FEV₁ analysis. An important conclusion was that in longitudinal studies with only two surveys, where data from the first and last surveys are

analysed to compute annual change, it is not possible to detect or assess the effects of such survey biases.

Statistical Considerations

Errors of measurement and other factors which can affect different subjects to different extents, and hence the resultant value for the correlation coefficient leads to the occurrence of regression to the mean (Healy and Goldstein 1978). The occurrence of "regression to the mean" describes the phenomenon whereby values which are extreme relative to some normative population will, upon subsequent measurement, tend to be less extreme. Galton in 1886 (Stigler 1986) first coined this expression when he noted that tall fathers tend to have tall sons, except that the sons, while taller than average, tended to be less extreme in this respect than their fathers were. In the pulmonary function setting, subjects who perform especially well in their first test are likely to show a later decline, whereas those with a poor initial effort would appear to improve upon subsequent tests.

The phenomenon called the "horse-racing effect" by Fletcher and his associates (1976) results in an association in which subjects with low initial levels of lung function are likely to experience a more severe decrement in lung function over time. The analogy for this occurrence is that "in a horse race the fastest horses are out in front because their speed has put them there" (p.73, Fletcher et al, 1976). The horse-racing effect, unlike the regression to the mean phenomenon, may have a biological basis. For example, Heedrik and Brunekreef (1985) found this effect to be more usually associated with strong environmental factors such as cigarette smoking.

To counteract regression to the mean, adjustment to take into account the initial measurement occasion FEV_1 value is often used. But this can lead to biased results, particularly in longitudinal studies in which assignment to comparison groups is not made at random. This pertains to the situations where the distribution of true initial values differs in

the two groups being compared, or the true relationships between change and the initial value may differ. For example, on a cross-sectional basis smokers, on average would have reduced initial lung function compared to non-smokers, and these differences would be expected to persist over the course of a longitudinal study. According to Vollmer (1988) adjusting for the initial value would have the effect of exaggerating differences between smokers and non-smokers, by producing divergent slopes of their rate of decline over time.

Apart from the possibility of controlling pulmonary function variability through the manipulation of the measurements themselves, or through study design modifications, attempts have been made to reduce the variability through statistical adjustment of the pulmonary function data. An example of an analytical technique that was introduced to reduce variance is the method of "winsorizing" FEV₁ slopes, that was applied by Fletcher et al. (1976), in a study of 1136 males in London over an 8 year period. Any single observation that was greater than 3 standard deviations (in this study ± 450 ml) from the mean value of FEV₁, was deleted. The total variance of the FEV₁ slopes so calculated were found to be 20% less than if no values had been deleted, and moreover the correlation of the "winsorized" slopes with the FEV₁ measurements became considerably stronger. While much can be done to improve the reliability of lung function values, such as by assuring competent and consistent technique in measurement by the technicians involved, deleting suspect values without ascertaining the cause of the extreme value is a misrepresentation of the data.

Burrows et al. (1986) chose to adjust for apparent "survey" differences in the distributions of FEV₁. Two of the seven surveys had shown deviations of -36.8 and +31.1 ml from a smoothed regression slope of all seven survey measurements. The procedure was justified as having the property of completely eliminating between-survey variability in mean Δ FEV₁ values, yet it would not affect the shape of the distribution or the variability of the FEV₁ values within a given survey. However the procedure could affect the individual estimates of annual decline, particularly where the values in the extreme surveys were

influential in the regression analysis. In this case, the adjusted levels were middle visits which would not be expected to have much impact.

Missing Values

Where the data has been analysed without consideration of the mechanism that has led to certain values being missing, there is potential for bias and imprecision of the estimates. If the non-response pattern does not depend on the lung function values or on covariates, the data can be considered to be "Missing Completely At Random" (MCAR). This missing data pattern is typically assumed when standard regression packages are applied to the data, and naive analyses in this case are valid (Laird 1988).

According to Laird (1988) "ignorable" missing data patterns occur when it is known why the data are missing, as the process leading to the censoring depends upon what have been observed already. The absence may be related to observed characteristics of the individuals, or to past history, or group membership. "Ignorable" missing data patterns are frequently suspected in longitudinal studies as subject attrition might well be related to previous performance.

With "non-ignorable" missing data, the probability of non-response is defined as depending on unobserved outcomes. If attrition depends upon values that would have been measured had the subject been available, then this would also be the case of a "non-ignorable" process (Louis 1988). An example given is that if subjects were dropping out of a placebo group because the treatment was not effective, and they were not getting better, then this response would be related to unobserved values. In such an event a model for the missing data process must be explicitly introduced. Due to the complexity of handling "non-ignorable" missing data, and since it is not even possible to detect by examination of the data, the occurrence of incomplete data is often arbitrarily classified as showing "ignorable

non-response" patterns. The distinction between the two is vague.

A number of different analytical techniques have been suggested for use in incomplete longitudinal data sets. Standard statistical packages such as SPSSX, BMDP and SAS use a case-wise deletion approach in which the analysis is based only on those individuals having complete data. While this is generally easy to carry out and could be satisfactory with small amounts of missing data, it can lead to biased estimates and is not very efficient as the study power is reduced. Complete case analysis is potentially wasteful as values for particular variables are discarded when they belong to cases that are missing other variables. The natural alternative is to include all cases where the variable of interest is present. While this pairwise available case method has advantages in that more values are used, the disadvantage is that the sample base changes from variable to variable.

Imputation-based procedures use estimates of missing values so that the resultant completed data set can be analysed by standard methods. Common procedures include "hot deck imputation" in which recorded sample values are substituted; "mean imputation", in which the means from sets of recorded values are substituted; and "regression imputation", where the missing variables for a unit are estimated by predicted values from the regression on the known variables for that unit. However, imputing missing values can create problems as it can force collinearity and may underestimate the variability (Little 1987).

A general method of incomplete data estimation has been described by Dempster, Laird and Rubin (1977) and referred to as the "EM algorithm" (Expectation and Maximization algorithm). The EM algorithm uses maximum likelihood techniques and essentially proceeds by:

- (1) replacing missing values by estimated values;
- (2) estimating parameters;
- (3) re-estimating the missing values assuming the new parameter estimates are correct;

(4) re-estimating the parameters, and so on, until iteration achieves convergence.

The advantages are its flexibility and also it allows the maximum amount of available data to be utilized. Problems encountered in using the EM algorithm include the slowness of convergence and dependence on the choice of starting values (Laird et al.,1987).

The identification of sources of variability in a longitudinal data set allows for the specification and possible control of these factors in the analysis. A naive statistical approach may easily introduce bias in the resulting estimates. Since the collection of longitudinal data is usually a major undertaking both in time and expense, thoughtful planning in the design and analytical approaches to the data will permit its maximum utilization. As the presence of missing data is a major complication in the analysis, the collection of data at at the same time for all individuals at each measurement occasion and the adoption of aggressive follow-up procedures are the best methods of avoiding subsequent complicated analyses or imprecise results.

CHAPTER V

A SURVEY OF PARAMETRIC METHODS USED FOR LONGITUDINAL DATA ANALYSIS

Introduction

Longitudinal data analysis can be viewed as a hierarchical process in which the units of analysis are the population, groups, individuals and measurements. It is the individual's measurements based on time that are considered the longitudinal component while characteristics that relate to population and group differences are more cross-sectional in nature. While any parameter that determines a cross-sectional mean can be estimated by repeated cross-sectional studies (in the absence of selection or drop-out bias), subject-specific rates of change can only be estimated by a longitudinal study (Louis et al, 1986). Modelling of longitudinal data invariably consists of two stages: initially, the relationship of each individual's lung function values with time is usually treated as a linear relationship. Population comparisons, which are the second stage, are derived by various methodologies. Averaging is the most commonly used method of combining individual coefficients, but other methods include a weighted least-squares or generalized least-squares approach and modelling random effects.

The majority of methods are based on linear models which assume a normal (Gaussian) distribution for the outcome variable of interest. To obtain population estimates, it is also assumed that estimates from different subjects are independent.

Endpoint Analysis

A conceptually simple method of longitudinal analysis is to use only the first and last observations for each subject to determine the differences in lung function over the time interval. Differences in the mean level of the outcome variable can be used as an estimate of the mean change in a group, or alternatively, differences between the first and last

measurements for each individual can then be averaged to give the mean rate of change for the group (Wu and Bailey 1988, Bryant and Gillings 1985). The typical application is for longitudinal studies in which only two measurements on each subject were available, giving little alternative for analysis. Johnston et al. (1976) justified the endpoint approach when three data points were available by intimating that this constituted insufficient data for calculation of regression coefficients. Certainly, if the measurements were equally spaced, the two methods would result in the same estimates. Another reason given for calculations of endpoint changes in lung function where more data points were available was the presence of "noise" in the spirometry data collected annually from 1980 to 1984 (Vollmer et al. 1986).

By relying on endpoint analysis only, the characteristics of the full data set are not taken advantage of, in that there is undue reliance on the accuracy of the two tests, and the average linear rate of decline is the only estimate of change. While the distribution of FEV_1 decline is typically normal in adult populations, with ageing, skewness in the data can occur in which a "tail" of clinically significant abnormalities are present. Therefore as pointed out by Miller and Thornton (1980) the average linear rate of decline of such a distribution is not the best index, since the effect on the susceptible minority tends to be overwhelmed by the unaffected majority.

Ordinary Least Squares Analyses

Linear models typically arise when a response variable Y is assumed to follow a normal distribution and have an expected value that depends linearly on exposure variables and individual characteristics. Based on a random sample of n units, the population mean, μ , is estimated by the sample mean, while the population variance σ^2 , is estimated by the sample variance, s^2 .

The Gaussian distribution is unimodal, bell-shaped and symmetrical about $x = \mu$, the mean of the distribution. This distribution was a mathematical construct introduced by Gauss

in the 18th century to describe the variability in the observations of astronomical data. Later in the nineteenth century, the normal distribution was used to describe the variation of individuals in a biological population with respect to height, and led to numerous biological applications of linear models (Stigler, 1986).

One of the simplest ways to characterise change in lung function for an individual is to describe the relationship by a straight line: $y_t = \beta_0 + \beta_1 t + \epsilon_t$. Time (t) is a variable assumed to be measured without error, β_0 and β_1 are unknown parameters representing the intercept and the slope of the line, respectively. The dependent variable y , deviates from the expected mean value given by the regression line by an amount ϵ , which is an unobservable "random" error term, whose values are unknown but assumed to have a mean value of zero. In a longitudinal study of pulmonary function "errors" depend on a number of factors, among them being inherent biological variation, errors of measurement, and departures of the "true" relationship from linearity (Schlesselman 1973).

The assumptions underlying the statistical analysis of the method of fitting straight lines by ordinary least-squares are as follows:

- 1) The population (true) relation between lung function y , with time, the independent variable, and other predictors is assumed to be a straight line.
- 2) The standard deviation of y is assumed constant, that is homoscedastic. The observations are thus considered to have the same scatter at all points along the curve, such that equal weight can be attached to all observations.
- 3) The errors in the observations of y , the dependent variable, at each of the selected time values will follow Gaussian distributions, and there is independence between the errors (Berenson et al. 1983).

A potential violation of these assumptions which arise in applying the OLS procedure to longitudinal data is the correlations which may be present between the values of the dependent variable for the same individual at different points in time. Where assumptions are violated such that error terms are heteroscedastic and/or autocorrelated, the estimate of the regression coefficient (β_1) will remain unbiased; however, the estimates of its standard error will be affected such that a nonconservative bias may be introduced, increasing the probability of incorrectly rejecting the null hypothesis that the regression coefficient is zero (Neter et al. 1985).

Ordinary least squares (OLS) methodology is used to estimate the unknown parameters, β_0 and β_1 . In order to calculate the best straight line by the method of least-squares it is necessary to find the line that will minimise the sum of squares of deviations of the dependent variable from the regression line of predicted values, based on the vertical distances between the two.

Cross-Sectional Analyses

Cross-sectional analysis is usually applied to a sample of individuals having one observation each taken during one point in time only. Alternatively, repeated cross-sectional observations can be taken on different samples of subjects (of the same population) at different points in time. Conventional analyses for this type of data incorporates the analysis of variance approach, with errors of observations being independent between individuals.

A cross-sectional type of analysis can be performed on longitudinal data in a number of different ways:

(1) Each survey year's data can be analysed singly; for example, Brinkman et al. (1972) used analysis of covariance to adjust the occupation-smoking-bronchitis group means for differences in age. Linear plots of maximum mid-expiratory flow by occupation group were presented separately for each of the measurement years of 1959, 1964 and 1970. Lawther et al. (1978)

also treated each measurement occasion as separate regressions analyzed for each of the three years of follow-up.

(2) The second, more crude method, lumps all the data together. Not only does this violate the assumption of independence of observations required in ordinary least squares analysis, it tends to give greater weight to subjects with multiple tests. After stating these problems, Burrows et al. (1983) chose to use this method to provide descriptive equations on the growth and decline of FEV_1 over a lifespan.

(3) Cross-sectional analysis has been performed (on the middle visit for instance) for each individual in the study. A linear slope of FEV_1 calculated by linear regression using each subjects' age midway through the follow-up, has been justified as being equivalent to fitting each subject's data to a model of accelerating decline as the tangent of a curve at mid-point can be represented by a straight line. As well, where data points of a given subject are extremely variable, it was felt justified to use this OLS type of analysis as opposed to complex modelling (Burrows et al. 1986). Age-related cross-sectional estimates of decline have been found to differ from time-related estimates of decline using longitudinal analysis (Louis et al. 1986). It was only where the age distribution is Gaussian or symmetrical, and where the age effect is modelled as linear and quadratic, that the expectation of the estimation of age effects will agree. Other discrepancies could result from cohort and drop-out effects.

Rather than performing a cross-sectional analysis on the FEV_1 values, an alternative approach is the use of the differences in FEV_1 level between adjacent measurements as the dependent variable and the time between those measurements as the independent variable. Neter et al. (1985) recommend using this approach to remove first order autocorrelation in the data sets. However the calculation of successive decrements in lung function relies too heavily on individual values which are prone to measurement error. Successive decrements are negatively correlated due to the fact that x_2 , an adjacent lung function measurement, is involved in both $\Delta x_1 (= x_2 - x_1)$ and $\Delta x_2 (= x_3 - x_2)$. If x_2 is measured with error ($\epsilon >$

0) then Δx_1 is "too small" and Δx_2 is "too big" (Van't Hof et al. 1977). Although this method somewhat lessens the impact of the independence assumption violation, the standard error of the estimates will be inflated, having a conservative effect on significance tests.

Individual Regression

The use of a least-squares analysis for each individual's data points to calculate the slope as an outcome measure, is the most commonly used analytical approach to longitudinal pulmonary function data. It is readily understood and accepted by the medical community; all available data points are used; and missing data and variable follow-up times present no computational problems (Buist and Vollmer 1988). The relatively short follow-up intervals encountered in most epidemiological studies are considered to be sufficiently modelled by straight line techniques (Hui and Berger 1983). A follow-up period of four to six years for individual lung function data and for subjects under sixty years of age is felt to be adequately described by a linear relationship (Buist and Vollmer 1988).

When comparing population slopes, the method almost universally used is to average the slopes. The problem with this approach is that each slope is measured with a different precision, having a different standard error. Thus the more imprecise slope of a subject who was only measured twice would be given equal weight to that measured ten times in another individual over the same study period. Typically the slope of the individual lines is treated as the dependent variable and a multiple regression or analysis of covariance is performed to assess the impact of various predictor variables on this dependent variable of change in pulmonary function. Modelling a linear relationship is inappropriate where the underlying change in pulmonary function is not linear in many subjects. In an attempt to overcome this shortcoming, interaction terms and/or variable transformations in the model have been attempted, but this may lead to difficulties in the interpretation of the results.

Polynomial Curves

Linear models need not necessarily imply a linear relationship of the independent variable with the dependent one; they are defined as models in which the parameters (the quantities to be estimated), appear linearly. For example, the equation $y = \alpha + \beta x + \gamma x^2 + \epsilon$ represents a parabola, even though the relationship between y and x is non-linear, the parameters α , β , and γ appear linearly. Increasingly complex shapes can be described by including higher powers of x .

The basis for estimating the unknown parameters of the polynomial model is the criterion of minimizing the sum of squares. If the errors are considered to be independent and identically distributed normal random variables then the OLS estimates of the unknown parameters have the property of being minimum variance estimators (Ratkowsky 1983).

The quadratic parameter γ denotes the "accelerative" component of ageing (Schlesselman 1973). While quadratic functions have convenient mathematical properties the biological interpretation of the results obtained is unclear. Furthermore, the estimates of the unknown parameters can be statistically dependent, (for example, when time and time² are used in the model) making subsequent statistical analysis difficult, both technically and interpretatively (Van't Hof et al. 1976). Certainly, for higher order polynomial models the fitting of a model becomes an empirical process as the parameters have no real physical meaning. A p^{th} degree polynomial can always be found that will exactly pass through any specified $p + 1$ points.

It has been demonstrated that FEV₁ decline in normal adults can be well approximated by a quadratic function of age (Louis, 1988). An accelerated decline over time, as represented by a second degree polynomial, has been noted particularly among industrial workers exposed to such irritants as cement dust (Siracusa et al. 1986) or grain dust (Schulzer 1985).

The exponential function is an example of a non-linear process and is considered to be perhaps the most important of all basic functions used in the biological sciences (Eason, Coles and Gettinby 1980). We find many applications in analysis of growth, an example being the logistic model: $y^\theta(t) = k/[1+\exp(a + \theta t)]$. As in the case of linear models, least-squares can be used to estimate θ by minimising the expression $S(\theta) = \sum (y_t - y_t^\theta)^2$. However, one does not derive an explicit expression for the estimate of θ (Ratkowsky, 1983). In almost all non-linear problems the solution involves successive approximations or iterations.

The least-squares estimators of the non-linear parameters can be biased, non-normally distributed, and have variances exceeding the minimum possible variance (Ratkowsky 1983). It is with larger sample sizes that the non-linear estimators tend to become more and more unbiased, normally distributed and approaching a minimum possible variance (provided that the error term ϵ_t is independent and normally distributed) (Ratkowsky 1983). Unified analysis of serial measurements on a group of individuals, each described by non-linear growth curves, with coefficients randomly distributed in the population is technically difficult and requires simplifying assumptions and iterative methods (Cook and Ware 1983). Unlike polynomial curve fitting, the character of individual non-linear curves is subject to distortion through averaging (Guire and Kowalski 1979).

Nonlinear curves do not necessarily require such complicated considerations. A typical power relationship, $y = ax^b$ can be log transformed such that the relationship between $\log y$ and $\log x$ becomes linear. An application is that the logarithms of the FEV₁ measurements against \log_{10} time would be linearly related (Causton 1983). Provided the necessary assumptions hold for the transformed variables, this allometric model therefore allows ordinary least squares principles to be applied. The allometric coefficient relates a percentage change in Y with that of X (Morris and Rolph, 1981).

The only published study to date which has attempted to apply a non-linear analysis to pulmonary function data was that of Emergil and Sobol (1971). The observation of an exponential decline of percent fall of FEV_1 or MMF per year was estimated by a logarithmic-linear regression equation. An advantage of this analysis is that the decline in pulmonary function could be represented by the "half-life" which is defined as the years taken to fall fifty percent in value. Attfield (in a discussion of a paper by Cole (1975)) alluded to this method when he questioned the tradition of forcing a linear decline on FEV_1 with age, and suggested that FEV_1 may be reduced by a constant proportion per year rather than by a constant decrement.

Weighted Least-Squares

A violation of the assumption of constant error variance which underlies ordinary least-squares analysis, results in some of the observations being less "reliable" than others. To carry out weighted least squares (WLS), the deviation between the observed and expected values of y_i is multiplied by a weight, ideally chosen to be inversely proportional to the variance of y_i . Typically these weights are unknown initially and have to be estimated based on results of an ordinary least-squares fit applied iteratively (Montgomery and Peck 1982).

For data in which the relationships are not truly linear or where there is imbalance or non-randomly missing data, there is potential for subtle biases to be introduced in a weighted regression approach (Palta and Cook 1987). Examples of the use of weighted regression for estimating pulmonary function decline include the studies by Hughes et al. (1982), Diem et al (1982) and Berkey et al. (1986).

General Linear Models

Generalised least squares methods attempt to fit a line that not only minimises the sum of squared distances of each point from the regression line, but also provides a solution that is compatible with the correlation of the errors (Fabsitz et al. 1985). The generalised least-squares estimator takes into account the variances and covariances of the error terms, and therefore any evidence of heteroscedasticity or autocorrelation. The key to applying GLS is the knowledge of the error term variance-covariance matrix, Σ , which contains any violations of the OLS assumptions about the structure of the variances and covariance of the error terms (Hanushek and Jackson 1977). Using the proper covariance structure can increase efficiency, protect from the biasing effect of drop-outs, and produce valid standard errors (Louis 1988). Where error terms have had unequal variance originally, GLS gives greater weight to those observations whose error terms have smaller variances. In the autocorrelated case GLS transforms the variables, such that the error terms implicit in the transformed variables are uncorrelated (Hanushek and Jackson 1977). The ordinary least-squares estimator is a special case of the GLS estimator if the covariance structure = $\sigma^2 I$, where I denotes the identity matrix (Hanushek and Jackson 1977).

Within-subject covariances have been traditionally analysed using two major types of models. The univariate mixed model (Winer 1971) assumes that the observations from the same subject have a constant variance and a common correlation. However the response-over-time data often involves observations made on the individual at two closely spaced time points, being more highly correlated than two observations made at longer intervals. It is for the general multivariate repeated measures (Cole and Grizzle 1966) and growth curve (Grizzle and Allen 1969) models that no assumptions about the structure of the within-subject covariance matrix are made.

A general linear model which represents both the fixed mean vector component and the random covariance parameters is represented by the following, derived from Louis (1988):

$$Y_i = X_i\alpha + Z_i\beta_i + \epsilon_i(t_i)$$

Associated with the vector of responses over time for each individual are covariates X_i for fixed effects and Z_i for random effects. The Gaussian model assumes that the random vectors β_i and ϵ_i are independent, each with multivariate Gaussian distributions with mean 0 and covariance matrices D and $R(t_i)$, which allows for different times of follow-up for each individual. The mean vector and the covariance matrix are functionally independent of one another. The covariance matrix of Y_i is:

$$\text{Cov}(Y_i) = \Sigma_i = Z_i D Z_i^T + R(t_i)$$

Special cases of this model which will be described are:

- 1) The general multivariate model in which the unconstrained covariance matrix depends on i only through the time of observations.
- 2) The time series autoregressive moving average model (ARMA) which usually assumes a first order Markov covariance structure, where $D = 0$ and R is stationary.
- 3) The random effects model, which assumes that $R = \sigma^2 I$, with I being the identity matrix. The expected values of responses in this model are modelled in a marginal or cross-sectional form. Arbitrary linear models derived for the expected values can be used to model the response at time t in terms of covariates measured at the same time t or previous times. The true longitudinal nature of the data is captured in the model through the correlations between measurements on the same individual as determined by the covariance.

Inference for the general linear model can be based either on least-squares or maximum likelihood methods or on empirical Bayes methodology. (Laird and Ware 1982). Upon selecting a particular model it is necessary to estimate the unknown parameters and obtain some measure of the accuracy with which they have been estimated. Estimation proceeds by designing a measure of goodness of fit between the data and the corresponding set of fitted values generated by the model and choosing the parameter estimates as those that minimise the chosen goodness-of-fit criterion. For general linear models, the usual form of estimation is through maximising the log-likelihood of the parameters given the data. Maximising the log-likelihood is equivalent to minimising the deviance, which is defined as $-2(\log \text{ likelihood of the parameters minus the log-likelihood of the fitted values equivalent to the data})$. For a normal distribution given by;

$$f(y;\mu) = 1/\sqrt{(2\pi\sigma^2)} \exp(- (y-\mu)^2 / 2\sigma^2),$$

the log-likelihood becomes

$$l(\mu;y) = -1/2 \ln(2\pi\sigma^2) - (y-\mu)^2/2\sigma^2,$$

and the deviance becomes

$$-2[l(\mu;y) - l(y;y)] = (y-\mu)^2/\sigma^2$$

When the underlying distribution is normal, generalized least-squares can give results identical to the method of maximum likelihood. The problem in the use of maximum likelihood for incomplete data is that it does not take into account the loss in degrees of freedom in estimating the fixed effects. Restricted maximum likelihood (REML), which maximises the likelihood for the residuals of the fixed effects and thereby takes into account the degrees of freedom can be substituted (Berk 1987). Unfortunately maximum likelihood estimates are often not available in closed form.

For unbalanced data, the EM algorithm has been recommended for computing maximum likelihood estimates based on iterative re-estimation of the missing values (Laird and Ware 1982). Although the EM algorithm is usually not the most efficient algorithm in this particular setting as it may be very slow to converge, it does offer a general approach which can be broadly applied to the longitudinal setting (Laird 1988).

The General Multivariate Unstructured Model

This model applies to the situation where all individuals are observed at the same occasions, and where there is no basis to assume a special covariance structure. Thus one can assume that:

$$\text{COV}(\epsilon_i) = \Sigma$$

(Ware 1985). This unstructured or fully parameterised form for the covariance structure allows a generalisation of several standard multivariate analyses to incomplete data.

The approach can be implemented even when the design is moderately imbalanced or data are missing completely at random or at least are ignorable. The covariance matrix may be estimated either from complete cases or from pairwise present cases (Berk 1987). When the design matrix, X_i , includes values of time varying covariates, the algorithmic approach to maximum likelihood estimation depends upon the availability of the values of the time varying covariates, when the response is missing (Ware 1985). When they are not available, an iterative approach is required using the EM algorithm applied to the residuals $y_i - X_i\alpha$. Estimates for α and the covariance matrix Σ are solved for two sets of equations based on initial use and application of ordinary least squares. (Ware and De Gruttola 1985). The estimator derived in this way is consistent but may be severely biased in small samples.

The general multivariate model is not flexible and should not be applied to situations where individuals are measured at arbitrary or unique times, and when the set of observation times becomes large relative to the number of individuals. A full multivariate model with an unrestricted covariance structure requires a proliferation of variance parameters many of which will be poorly estimated (Laird and Ware 1982). Additionally when the covariance of successive measurements depends on the time of measurement the general multivariate model will either be inestimable or very inefficient (Ware 1985). For prediction purposes, the general multivariate model performs poorly as it is not possible to extrapolate beyond the range of the data (Louis 1988).

Time Series Approach

The autoregressive moving average (ARMA) models are particularly suitable for designs in which there are relatively few subjects, a relatively large number of observations per subject and measurements are equally spaced (Louis 1988). The typical setting for such methods has been the single subject time series. Autoregressive models may be more appropriate for studies of longer duration as short term declines are often more linear. The first-order autoregressive (AR) model is an example of the time series model usually employed. With the response variable being the FEV₁ measurements, the covariance structure between different error terms for each individual is a function of the correlation;

$$\text{cov}(\epsilon_{ij}, \epsilon_{ik}) = \sigma^2 \rho^{|t_{ij} - t_{ik}|}, \quad 0 \leq \rho < 1.$$

As long as the ratio of the smallest to largest intervals measured is not too great an iterative fitting method has been developed to analyse this type of data with positively correlated error structure. This iterative method developed by Louis and Spiro (1986) analyses the data first by OLS, and the cross-sectional residuals are then used to find initial estimates for ρ and σ^2 . The next estimate for α is then found by applying

OLS estimation procedures to the transformed data which is then analysed in an iterative manner. The advantage of this method is that the correlation structure of the residuals can account for regression to the mean (Louis 1988). This method of estimation has been applied to a longitudinal analysis of pulmonary function data as reported by Diem and Liukkonen (1988). The authors concluded that they were unable to obtain useful results as the error correlations (ρ) began to cycle periodically and failed to converge.

The time series approach assumes an autoregressive (AR) structure for the error terms e_{ij} , not the observed values Y_{ij} (Ware 1985). An alternative autoregressive approach by Rosner et al. (1985, 1988) includes previous Y values as regressors. The fixed effects α then depend on the covariance structure. A model which incorporates missing and or unequally spaced examinations has been developed. The application of this method has been shown to be successful in a study of the pulmonary function of children over years spaced 1, 4, 5, 6, and 7 years apart in East Boston Mass. (Rosner and Munoz 1988). In this example the effects of previous FEV_1 , height, sex of the child and current smoking were all significant predictors of FEV_1 .

This approach involves a much simpler form of estimation with all available pairs of consecutive examinations being used for the application of a weighted non-linear regression. However the regression coefficients must be interpreted conditionally (Ware 1985) and this can cause difficulty in interpretation. An example given by Louis (1988) is that if a less precise measurement device was used the fixed effects would change as would the estimated effect of covariates on lung function. The AR model described by Louis (1988) is robust to such error misspecification and adding noise would only add to the variance and leave the mean vector unchanged.

Louis (1988) emphasized the most important problem with applying an ARMA approach to longitudinal data:

"We looked around and we could not find any data in a biological setting where autoregressive moving average covariance structure fit very nicely..... does anybody have real data that fits these models?". (p.359).

The Random Effects Model

The random effects models are based on a hierarchical structure for data, imposing correlations among subunits. For example, repeated observations on individuals are nested within the individual, and individuals may be nested within larger units such as occupational or smoking groups (Louis 1988). This model has the properties of:

- (1) accounting for the longitudinal structure and the correlation of repeated measures.
- (2) accommodating missing longitudinal data and
- (3) allowing for covariates which change over time and where the covariate change can differ for each individual.

The model allows for fixed effects, which stem from demographic characteristics such as race, and sex, and separately allows for within-subject and between subject variability (Fairclough and Helms 1984).

The random effects model consists of two stages; in the first stage of the model the distribution of the serial measurements for each subject are assumed to have the same form but the parameters of the individual distributions vary from person to person. The distribution of these parameters or "random effects", in a population constitute the second stage of the model. For example, it can be assumed that the relationship between lung volume and the cube of height among children is linear, but with linear regression parameters that vary among the children. The two-stage models have no requirement for balancing the data and they facilitate the study of the effects

of background variables on this response. They are based on explicit identification of individual and population characteristics (Laird and Ware 1982). Therefore this approach has great potential for application to the study of chronic disease, where changes in patient status are usually of primary interest and data sets are often highly unbalanced (Vacek et al. 1987).

The two-stage random effects model describes population parameters, individual effects and within-person variation in the first stage, and between-person variation at stage two (Laird and Ware 1982). The model supposes a population trend over time with variation and individual trends with their variances distributed around the population mean. Components of variance allows for the contrast of the amount of variability over time within an individual compared to the average amount of variability across individuals.

The model, which corresponds to the general linear model described earlier (Louis 1988) is:

$$Y_i = X_i\alpha + Z_i\beta_i + \epsilon_i$$

where Y_i is the vector of observations for the i^{th} subject;

α is a vector of unknown constants which includes fixed population parameters;

X_i is a known constant design matrix corresponding to the fixed effects, α ;

β_i denotes a vector of unknown individual parameters;

Z_i is a known constant design matrix corresponding to the random effects β_i ;

The ϵ_i are assumed to be distributed as $N(0, R_i)$ (normal with mean zero and covariance matrix R_i). (Louis, 1988; Laird and Ware, 1982, Fairclough and Helms, 1984).

Fixed effects are attributable to a finite set of levels of a factor that occur in the data while random effects are attributable typically to an infinite number of levels of a factor of which only a random sample are deemed to occur in the data. This "within-individual" regression model has the marginal expectation

$$E(Y_i) = X_i\alpha$$

where α can be interpreted as the rate of change in the population averaged Y with X , the population curve (Zeger et al. 1988). For example, when fitting separate linear curves for two groups α would contain four parameters – a slope and an intercept for each of the two groups. (Schluchter 1988).

The random effects, β_i are treated as $N(0,D)$ independent of each other and of the ϵ_i . D is a positive definite symmetric covariance matrix of the random effects β_i . The covariance of the y_i is:

$$\text{cov}(y_i) = \Sigma_i = R_i + Z_i D Z_i^T$$

Typically, R_i is taken to be $\sigma^2 I$, where " I " denotes an identity matrix (Laird and Ware 1982). This restricts the within-subject variance to be uncorrelated and homoscedastic. Setting the covariance matrix as $R = \sigma^2 I$ implies that the covariances remain the same with increasing separation of the measurements with time, an assumption that may not be valid. While it is possible to incorporate other covariance patterns the estimation processes become further complicated (Berk 1987). A less restrictive covariance structure would allow Z_i to contain only the dependence on time and would be a function only of the pattern of observations and not of individual characteristics (Laird et al. 1987). The term $Z_i \beta_i$ models the deviation of the i^{th} individual mean curve from the overall mean, and ϵ_i models the "within-subject" variation (Waterneaux et al, 1989).

When all the covariance parameters are known, and α is treated as a fixed effect, estimates for the population and individual effects are straightforward to derive (Laird and Ware 1982). For most longitudinal epidemiological studies the covariance matrix must be estimated. The classical approach is based on maximum likelihood estimation of the fixed effects α as well as the variance and covariance parameters from the marginal distribution of the outcome variables Y_i (Laird and Ware 1982). An alternative form is the use of restricted maximum likelihood estimates (REML) which are estimates of the variance and covariance parameters, for those values that maximise the likelihood of the least-squares residuals;

$$e_i = y_i - X_i \alpha \text{ (Jennrich and Schluchter 1986).}$$

REML estimates have been found not to be biased under balanced ANOVA models; the estimates can be derived both by maximum likelihood procedures as well as by using Bayesian approaches (Laird and Ware 1982).

The random effects model has been applied successfully to incomplete pulmonary function data on children (Fairclough and Helmes 1984), grain workers (Vacek et al. 1987) and TDI exposed industrial workers (Diem and Liukkonen 1988). In the latter study the data consisted of up to 9 spirometric tests, which were administered to each of 277 workers. The model adopted was the following:

$$y_{ij} = \alpha_0 + \alpha_1 x_{ij1} + \dots + \alpha_k x_{ijk} + \beta_{0i} + \beta_{1i}(t_{ij} - t_{i1}) + e_{ij};$$

where $j = 1, \dots, n_i$ and $i = 1, \dots, N$

and y_{ij} is the FEV₁ at time t_{ij} and the x_{ij} are the value of various covariates for worker i in addition to time, represented by $t_{ij} - t_{i1}$. The α terms are unknown parameters with fixed global mean effects; and β_{0i} terms are individual deviations from mean levels and slopes and are regarded as random effects. In total 100 iterations were

attained to obtain convergence to the final estimates using the EM algorithm. (Dien and Liukkonen 1988).

The maximum likelihood ratio estimates that are needed for likelihood ratio tests to allow for significance testing must be obtained by iterative techniques. The EM method described by Dempster, Laird and Rubin (1977) is recommended for estimation of the random effects model parameters (Ware and Wu 1981). When applied to the two-stage random effects model, the use of the EM algorithm can result in problems due to slow convergence or sensitivity to starting values (Laird and Ware 1982). A restriction is that the number of time dependent covariates considered for an individual must be smaller than the number of measurement occasions (Rosner et al. 1985).

In the case where data to be analysed is affected by censoring (for example, when patients have died during the course of the study) the censoring times and the slopes may be related such that steeper slopes may be associated with shorter periods of observations. In this situation the random effects model may create biased estimates of the coefficients of decline as in effect, it "shrinks" the individual estimates of the slopes toward a common mean. The greater the sampling error of the coefficient of decline the greater the degree of shrinkage. Typically, steeper slopes have fewer times of observation which are measured with less precision and hence are shrunk more than others and as a result the combined slope is underestimated. (Little 1988) Wu and Bailey (1988) suggested an approach based on the random effects model to overcome this problem, which is based on regressing the slopes on the censoring times.

Since the main computational burden to applying random effects model is the iterative computation of the variance and covariance parameters, Laird and Ware (1982) suggest the investigation of the properties of non-iterative alternatives. At the present time there has not been consensus as to the most appropriate method of non-iterative

estimation. Vonesh and Carter (1987) published a non-iterative procedure for estimating and comparing the population parameters under a random effects approach for unbalanced data. Although these estimators are demonstrated to be both computationally and statistically efficient for large samples, for small samples their properties remain to be investigated. Vacek et al. (1987) applied a non-iterative procedure to compare maximum likelihood estimates and Empirical Bayes estimates to a random effects model of their data. Problems were encountered using the non-iterative method, as negative variance estimates were found for the covariance matrix D ; as well, more programming and data manipulation was required to use this alternative approach.

A number of authors, including Hui (1984), Goldstein (1986) and Diem and Liukkonen (1988) have presented approaches for fitting repeated measurements based on variants of the two-stage random effects model. The estimation approaches for such models are based on GLS analyses performed iteratively until convergence is achieved. Each method relies on the use of ordinary least-squares to obtain initial estimates of the coefficients of decline, followed by regression of the slopes on various covariates calculated using a maximum likelihood weighted regression method. The likelihood function accommodates differing variances of the individual slopes due to the data having varying chronology and time (Diem and Liukkonen 1988).

The Robust Variance Approach

The models proposed so far have all been parametric approaches to analysing longitudinal data which is continuous in outcome. As long as repeated observations are independent, whether the outcome is continuous or discrete, regression methods can be unified under the class of generalised linear models or quasi-likelihood models. These models can be used to fit linear, logistic, log-linear or survival models (McCullagh and Nelder 1983). While the degree of dependence in repeated observations is small enough

that often crude regression estimates which ignore the correlation are nearly optimal, the assessment of precision of these regression coefficients must take the dependence into account (Zeger et al. 1988). Zeger and Liang (1986) presented a multivariate analogue of quasi-likelihood that can be applied to longitudinal and other dependent data. The Generalised Estimating Equations (GEE) approach is based on the use of standard regression models for independent observations to estimate coefficients but using robust rather than standard variance estimates (Zeger et al. 1988).

As an example of the GEE approach, a linear model such as the random effects model can be adopted, where the least-squares estimating equation for α would be defined by the inverse of a matrix of weights (V^{-1}) which would have the form

$$\alpha = \Sigma(X'V^{-1}X)^{-1} (X'V^{-1}Y)$$

The robust variance estimate is given by the following equation:

$$\text{var}(\alpha) = (X'V^{-1}X)^{-1} [X'V^{-1}[\Sigma(Y_i - X_i\alpha)(Y_i - X_i\alpha)]V^{-1}X](X'V^{-1}X)^{-1}$$

For a large number of subjects in the study, consistent inferences require the correct specification of only the first moment, which is the expected value of $E(Y_i)$ which for the linear models is given by $E(Y_i) = X_i\alpha$. Thus the inferences about the regression coefficients are robust to misspecification of the model for time dependence or random effects, (Zeger et al. 1988).

Standard ordinary least-squares regression programs have been shown to give estimates for the variance of the regression coefficients which are too small by a factor of two or more (Zeger et al. 1988). The GEE method on the other hand relies on independence across subjects to estimate consistently the variance of the coefficients even when the assumed correlation is incorrect, as it often could be (Zeger and Liang, 1986). While the strategy has been shown to be effective under balanced conditions and

where data is missing at random, it performs poorly in the presence of non-ignorable missing data, or where there is severe missing data (Zeger et al. 1988).

Summary

Methods based on ordinary least squares assumptions have been a relatively long history for longitudinal data applications. Two point 'follow-up' studies must be limited to endpoint-based approaches with no possibility of estimating individual variability. With at least three points of follow-up, the most common approach for estimating the time-related change is by averaging individual-based regression coefficients. This technique has value in that an estimate of within-individual variability can be determined. With violations of assumptions underlying OLS procedures, which can be affected by such characteristics of longitudinal data as heterogeneity of variance and autocorrelation, more appropriate modelling of the covariance structure may be warranted.

The generalized linear model-based approaches offer sophistication in the analysis of longitudinal data. However, each method is limited by assumptions of the covariance structure or by moderate or severe imbalance in the data, particularly in the case of "non-ignorable" missingness. Future developments in non-linear modelling may allow more flexibility in the analysis of long term decline of adult pulmonary function. The use of bootstrapping techniques and other nonparametric approaches to longitudinal data are currently under development (Segal et al. 1988; Statistics in Medicine, 1988). Categorical outcomes could make use of marginal and transitional models, but for the immediate question of modelling decline in pulmonary function with time, parametric approaches are pertinent.

CHAPTER VI

DESCRIPTION OF THREE LONGITUDINAL STUDIES

Introduction

The data provided by the three longitudinal studies to be described in this chapter, form the basis for the subsequent analyses to be presented in this thesis. Each data set is the product of a concerted effort by the investigators to extend their repeated follow-up to a lengthy duration while maintaining consistent quality of their measurement process. For each longitudinal study a description of the planned protocol and measurements will be followed by a summary of the initial characteristics of the study population and an overview of the published results. A table of decline estimates similar to that used for the literature review (Chapter 3) will conclude the chapter.

The Coal Miners Study

Objectives

Information was collected periodically on coalminers who were referred to the Pneumoconiosis Centre at Crehang either on account of radiological change or because of a complaint of breathlessness. The purpose of this retrospective study was to record pulmonary function changes over the course of employment and after retirement from mining (Bates et al., 1985).

Methods

To be included in the longitudinal study records on each patient had to have an indication of smoking history and subjects were chosen on the basis of not having tuberculosis or complication of pneumoconiosis. Of 698 original records, 64 were rejected

as the work exposure was under 20 years; another 47 were rejected for being defective in their original lung function tests, while others were rejected for not fulfilling the following criteria:

- 1) they had been followed for at least 5 years;
- 2) at least two sets of pulmonary function tests had been available;
- 3) and a note on their smoking history was available.

Data on 396 coal miners were thus available for analysis. All observations were made in the same laboratory, and the same chief technician was involved over an 18 year period. Records since 1950 were used and the separation of subjects into smokers and non-smokers was made on the basis of consistency in their habit, such that no ex-smokers or persons who changed their smoking habits were included in the study population.

Measurements

A water sealed spirometer was used to record the forced vital capacity maneuver measurements of FVC and forced expiratory volume during one second (FEV_1); the best one of the three was noted. Measures of fractional uptake of carbon monoxide ($FuCO$) and diffusing capacity (DLCO) required use of the Dechoux/Pivotou apparatus for steady state measurement of carbon monoxide uptake. Also noted in the records was the subject's age at the first test, height, and the year of retirement, when this was applicable.

Results

The smoker and non-smoker groups were each divided into living and deceased sub-groups, according to their status at the time of ascertainment. Baseline characteristics of the 1984 study group are given in Table 6.1, as published by Bates et al. (1985). Alive smokers on average were slightly younger in age at first examination than the

men in the other three groups. Deceased nonsmokers had lower initial FVC and FEV₁ values, averaging about 91.6% and 81.1% respectively. The rates of fractional CO uptake however, were within normal levels, the lowest being 91.7% of predicted for the group of deceased smokers.

Table 6.1: Measurements of Coal Miners' Pulmonary Function at First Examination

No.	G R O U P			
	NONSMOKERS (deceased) 103	NONSMOKERS (alive) 38	SMOKERS (deceased) 110	SMOKERS (alive) 146
Height (cm)	168.2 (5.5)	167.9 (5.7)	168.6 (6.4)	168.9 (6.0)
Age at first Examination (yr)	50.9 (8.3)	49.1 (6.9)	49.8 (8.9)	46.6 (7.3)
Age at Retirement (yr)	56.5 (5.2)	55.4 (5.2)	55.7 (6.9)	54.8 (6.8)
FVC (ml)	4.249 (.638)	4.467 (.590)	4.485 (.710)	4.635 (.722)
FVC (% predicted)	91.6 (12.3)	96.3 (11.2)	95.7 (12.7)	97.1 (11.7)
FEV (ml)	2.769 (.677)	3.084 (.623)	2.887 (.596)	3.186 (.617)
FEV (% predicted)	81.1 (16.3)	89.6 (16.7)	83.6 (16.0)	89.6 (14.8)
F _{Co}	42.6 (5.3)	47.5 (6.3)	39.9 (6.1)	42.8 (5.8)
F _{Co} (%predicted)	98.6 (10.6)	107.8 (12.3)	101.7 (13.7)	96.8 (13.1)
() = standard deviation				

The published decline estimates were based on a 2 point slope approach, that was not adopted for the thesis. Where the 2 point FEV₁ differences are weighted by the number of years of follow-up, the estimate obtained is exactly the same as if the endpoints only were used for each individual. The pre- versus post-retirement decline estimates were similar overall for vital capacity and FEV₁, but the alive nonsmokers and smokers groups showed larger post-retirement declines.

As it was not based either on the total population or on a randomly selected sample, the study group cannot properly represent the total coal miner population at risk in the Lorraine area of France. However, along with the relatively large numbers of miners studied and a long duration of follow-up, the relative normality of pulmonary function found at the first examination, as well as comparability of initial data reported by Dechoux et al. (1983) on a study of 685 coal miners from the same area of France; it can be concluded that the subjects of this longitudinal study are reasonably representative of this population, so that the results may be generalized.

The Veteran's Study

Objectives

The overall purpose of the study was to evaluate the natural history of chronic bronchitis and to allow comparisons between cities with differing levels of atmospheric pollution.

Methods

The first participants to be evaluated were from Toronto in 1958. By 1962, centres at Montreal, Halifax and Winnipeg were established. Over the course of the first year of the study, monthly measurements were attempted on each subject, followed

by annual or bi-annual (winter/summer) visits over a minimum ten year follow-up periods.

In order to avoid systematic error or bias due to some chance selection of the cases, the subjects were studied in different cities with different laboratories. The longitudinal study was initiated through the cooperation of the Department of Veteran Affairs in Canada, where veterans who complained of chronic bronchitis were in receipt of disease-related pension. The patients selected were not a simple random sample and may not be representative of the population at large. Nevertheless this administrative structure could permit careful follow-up of patients necessary for a prospective study. Patients fulfilling the following selection criteria were deemed suitable for inclusion in the project: a man had to have clinical criteria of chronic bronchitis as determined by answers to the British Medical Association Research Council questionnaire (ie. a history of cough productive of sputum for two years or more; complaints of shortness of breath sometime during the year); there had to be no other known cause of the chest disease; the blood pressure and electrocardiogram had to be normal; and the subject had to be earning his own living, which was a criterion used to help exclude anyone already disabled by the disease. The selection criteria were adopted to provide more comparable subpopulations to be selected from the four cities.

Measurements

To further ensure comparability between study centres, centralised ordering and calibration of equipment was performed. As well, the technicians were trained in one laboratory, to ensure uniformity of method and procedure. Measurements of the subdivisions of lung volume were performed in the same sequence using the helium closed circuit technique. The apparatus consisted of a six litre spirometer with an electrically driven kymograph, an external CO₂ absorption canister and a blower. With

the blower adding oxygen to the circuit, the patient was seated with nose clip on, and instructed to carry out one vital capacity, by first fully inflating and then immediately deflating his lungs as far as possible; three repeated trials were performed. From this manoeuvre, readings of total lung capacity (TLC), vital capacity (VC), residual volume (RV) and expiratory reserve volume (ERV) were obtained.

The apparatus used for the steady state determination of the diffusing capacity (DLCO) consisted of an infrared carbon monoxide meter, a nine litre spirometer and a switch box controlling the gas sampling. Expired gas was sampled as the patient first breathed room air through the circuit and the carbon monoxide meter deflection was noted. The patient was then connected to the carbon monoxide mixture and samples were drawn from the inspired and mixed expired air as well as the end tidal gas. DLCO was calculated from measurements of inspired minute volume, inspired gas percent, expired minute volume and mixed expired gas percent. The fraction of carbon monoxide uptake (FCO) was calculated as the difference between the fraction of CO inspired and expired, divided by that inspired.

As suitable commercial apparatus were not available for measurement of the spirometric values of forced expiratory volume (FEV₇₅) and maximal mid-expiratory flow, the construction of a hand made spirometer was necessary. The patient was measured while standing, the nose clip was applied and false teeth if present, were removed. After taking a deep breath the patient blew into the tube as quickly and completely as possible. The procedure was repeated until four technically satisfactory tracings were obtained. Using a conversion scale the volume expired during the first three quarters of a second was computed, and then multiplied by forty to indicate a maximal breathing capacity. This obtained value can be readily converted into the more traditional FEV₁ measure by multiplying it by a fraction, as derived from Miller et al. (1959). The MMF was computed from the same curve, based on measurement of the

mid-half of the volume of the total expiration.

At the outset of the study an interview-based questionnaire on respiratory symptoms based on that of the British Medical Research Council was administered. Along with questions on symptoms of chronic bronchitis, such as cough, phlegm production and dyspnea, smoking and occupational histories were recorded in detail in addition to questions concerning climatic influences on health. Attempts were made to characterise the climate differences and pollution level of each city studied, using available reports on measured levels of hydrocarbon (COH) units as a measure of smoke and haze concentration, as well as gathering data on dust fall and sulphur dioxide. A general conclusion was that Winnipeg was least polluted, Montreal and Toronto had the highest pollution levels, with Halifax being intermediate in pollution levels.

The follow-up protocol included pulmonary function measurements undertaken at each laboratory session, as well as a categorical smoking scale. The smoking scale consisted of seven levels:

- 1 - Never Smoked*
- 2 - Smoking 1 - 10 cigarettes a day*
- 3 - Smoking 11 - 20 cigarettes a day*
- 4 - Smoking 20 + cigarettes a day*
- 5 - Less than 2 ounces of pipe tobacco a day*
- 6 - More than 2 ounces of pipe tobacco a day*
- 7 - Stopped Smoking*

Results

Average values for physical characteristics and smoking habits measured at the start of the study are shown in Table 6.2.

Table 6.2: Veteran's Physical Characteristics and Smoking Habits

CENTER	No.	Mean Age	Mean Ht(in)	Mean Wt(lbs)	Age Started Smoking	# Cigs /week
TORONTO 1958/62	70	48.4	68.1	170	18.0	107
WINNIPEG 1960/62	67	49.2	68.2	167	18.5	105
MONTREAL 1960/62	39	45.0	67.2	157	18.5	115
HALIFAX 1960/62	40	51.0	68.0	166	18.0	110

The published values for the physical characteristics are notably similar in their means between the cities. Only two of two hundred and sixteen men who began the study were lifelong non-smokers; a further ten had smoked only pipes in their adult life. The age at which the men first began to cough all averaged within one year of the overall mean of 30 years of age. As an indication of haemoptysis (whether sputum had ever been blood streaked) positive responses varied from 29.8% for Winnipeg to 45% for Halifax. In summary it may be concluded that the men in the four cities showed remarkable constancy in clinical history, physical characteristics and smoking history, even though none of the selection criteria specifically relied on any of these personal attributes. This suggests that they have representative clinical syndromes which may be commonly encountered in the population from which they were drawn.

Published values of the baseline pulmonary function tests are listed in Table 6.3 (Bates, 1966). Below the actual values of the various pulmonary function variables are

the published predicted values based on age, sex and height.

Table 6.3: Measurement of Veteran's Pulmonary Function at First Examination

CITY	TORONTO 1958/62	WINNIPEG 1960/62	MONTREAL 1960/62	HALIFAX 1960/62
n	70	67	39	40
TLC	6.22 (6.30)	6.32 (6.24)	5.64 (5.89)	6.17 (6.07)
VC	3.66 (4.24)	3.40 (4.11)	2.77 (3.98)	3.03 (3.96)
ERV	0.68 (1.28)	0.83 (1.37)	0.56 (1.37)	0.70 (1.33)
FRC	3.15 (3.21)	3.74 (3.37)	3.41 (3.17)	3.83 (3.27)
RV	2.54 (2.00)	2.91 (1.99)	2.90 (1.85)	3.14 (1.98)
RV%TLC	40.9 (32.1)	45.8 (32.1)	51.3 (30.6)	50.7 (32.7)
FEV .75x40	83.5 (110.6)	90.4 (109.6)	80.9 (111.3)	79.0 (105.7)
MMF	1.98 (3.49)	2.42 (3.48)	2.51 (3.59)	2.23 (3.40)
DLCO	18.1 (17.1)	19.7 (16.7)	13.3 (18.0)	15.1 (16.5)
F _{Co}	0.408 (0.436)	0.415 (0.433)	0.384 (0.452)	0.405 (0.438)

() = predicted values

The TLC for all centres was not significantly different from the predicted value. However, for the VC the opposite case was shown, in that the values were statistically significantly lower. The FEV_{0.75} x 40 variable and the maximum mid-expiratory flow rate also were significantly lower than predicted. The pattern of pulmonary function

impairment for the groups as a whole included evidence of overinflation as shown by an elevated residual volume, and its percent of TLC; ventilatory impairment was somewhat less in degree than might have been expected; there was only slight evidence of gas distribution impairment, as indicated by lower FCO values, particularly for Montreal; and apart from the Montreal group, a normal level of carbon monoxide exchange. The fact that the DLCO and FCO values were not much lower than predicted indicate that the population did not include many individuals who already had emphysema. It may be concluded that the degree of impairment of initial pulmonary function was greater for the Montreal and Halifax groups, in comparison to those from Toronto and Winnipeg.

Annual changes only were calculated for a few successive years of follow-up. A general impression was that a change in FEV appeared to be greater in Toronto and Halifax than in Montreal or Winnipeg. In fact for the latter two cities it appeared not to change in the four year period covered by the follow-up. It was concluded that a longer period of follow-up would be necessary before definite trends can be established and before the rate of change in individual function tests could be compared in a meaningful way. A 10 year follow-up (Bates, 1973) showed endpoint-calculated slopes to be lowest for Winnipeg which yielded an unexpected increase in the FEV₁ measurement. All the lung function change estimates were lower than those found for the original follow-up survey.

Grain Handlers Study

Objectives

The purpose of the original cross-sectional study was to determine the prevalence of respiratory abnormalities among grain elevator workers in British Columbia, and to compare this to a control group consisting of white collar workers employed at Vancouver city hall. Longitudinal follow-ups have since been conducted on three additional occasions, approximately three years apart. Subsequent publications have addressed the issue as to what host factors affected decline and how to characterise this decline.

Methods

The original study involved 642 workers who were employed in four grain elevator terminals in the port of Vancouver, and in one located in Prince Rupert. As controls, 206 civic workers employed at Vancouver city hall were used. Non-white subjects and women were excluded from the analysis.

Measurements

Spirometric measurements of FVC, FEV₁ and FEF₂₅₋₇₅ were conducted at the work site using the same 13.5 litre Collins spirometer and the same respiratory technicians on each measurement occasion. The best FEV₁ and FVC obtained from the minimum of five technically satisfactory forced expiratory manoeuvres were used for analysis. Forced expiratory flow during the middle half of the FVC (FEF₂₅₋₇₅) was obtained from the spirogram with the highest sum of FEV₁ and FVC.

A medical and occupational questionnaire which included a modified version of the British Medical Research Council questionnaire on chronic bronchitis was administered by trained interviewers at each measurement session. Allergy skin tests, as well as pre/post-shift work testing were carried out.

Results

Table 6.4 lists the physical characteristics, smoking habits, occupational information and pulmonary function results for 610 grain workers and 136 civic workers who were tested on the initial survey (Chan-Yeung et al., 1980). The results for females and non-Caucasians were excluded from the analysis, because of differences in predicted lung function values found for caucasians. In general, the grain workers as a group were younger and showed a trend of being slightly shorter. As well the civic workers had a greater proportion of non-smokers and ex-smokers while a greater percentage of current smokers were found among the grain workers. The initial FEV₁ and FVC values were lower for grain workers in comparison to civic workers. Both age and height were significant predictors of initial FEV₁.

Table 6.4: Measurements of Grain workers and Controls at First Examination.

	Grain Handlers	Civic Workers
n	610	136
Age	37.8 (12.6)	44.3 (11.2)
Height	177.2 (6.9)	178.4 (6.5)
% Current Smokers	49.2	29.4
FEV ₁ %		
Nonsmokers	100.0 (14.8)	104.0 (12.6)
Ex-smokers	100.2 (15.4)	106.1 (15.2)
Current smokers	97.8 (14.5)	99.6 (13.0)
FVC %		
Nonsmokers	99.8 (13.7)	104.7 (11.2)
Ex-smokers	100.8 (13.5)	108.8 (13.3)
Current smokers	99.4 (12.8)	101.8 (12.1)
FEF ₂₅₋₇₅		
Nonsmokers	99.0 (27.7)	94.4 (25.2)
Ex-smokers	97.2 (26.7)	94.1 (32.3)
Current smokers	93.0 (33.0)	87.6 (22.3)

The results of the second survey in which 396 grain workers and 111 civic workers had repeat measurements, showed an annual decline in lung function greater for grain workers than civic workers, particularly among workers over the age of 50 years

(Chan-Yeung et al., 1985). This decline in lung function was significantly correlated with age and smoking for both work groups. By the third survey 340 grain workers had taken part in all the surveys, and subsequent analysis was performed on 267 workers who did not change their smoking habits over the period of the study (Tabona et al., 1985). The spirometric measures were found to decline more rapidly in older grain handlers as compared to younger ones. Whilst smokers in this group had a slightly greater decline, differences failed to reach statistical significance.

Summary

In summary, the initial lung function measurements on all three study groups did not deviate substantially from the normal expected values. The civic workers, who were chosen intentionally as controls, had average respiratory values close to, or even exceeding predicted levels. It was the veterans, who were chosen for the clinical presence of chronic bronchitis, that had the widest differences between the measured and predicted lung function values. Their average initial values, approximating 75% of the predicted, indicated that their impairment was at an early stage at the start of the study. Both the *VETERANS* and *COAL* workers were about 49 years of age on average; the *GRAIN* group were the youngest, being 38 years old on average.

Table 6.5 describes what published information is available about the thesis data sets. No assessment of the significance of differences between the groups was given in any of the papers. In both the veterans publications the original $FEV_{0.75 \times 40}$ value (in l/min) was given with no attempt at converting the values to an FEV_1 equivalent. Results for the *COAL* data publication were presented for both pre- and post-retirement periods and showed relatively high yearly decline estimates for all groups with VC results similar to those of FEV_1 .

The publications on the *GRAIN* data were based on subgroups of 2 and 3 occasions of follow-up. In the study by Chan-Yeung et al. (1981) the subgroup results are presented, which show FEV_1 , FVC and MMF decline differences according to group, age and smoking category. Tabona et al. analyzed a subset of the original grain workers who were measured all three times, and stayed consistent in their smoking habits. Smokers showed the highest declines. Schulzer et al. (1985) did use a regression analysis of individual slopes which showed the civic nonsmokers, unlike the other three groups, to have an overall increase in FEV_1 over the six years. A quadratic decline in smokers and grain workers was also noted.

Table 6.5: Published Estimates of Lung Function Decline for the Three Thesis Data Sets

STUDY	CHARACTERISTICS	SAMPLE	CHANGES IN FEV ₁ ml/yr	OTHER LUNG FUNCTION CHANGES
Bates et al. (1966)	2-6x/2-6yr	<u>VETERANS</u>		FEV _{0.75x40}
	FEV _{0.75x40}	Toronto n=70		-4.9 to +5.3 l/min
	=79-92.8 l/min	Winnipeg n=67		+0.8 to +3.5 l/min
	Age - 48.4 yr	Montreal n=39		-1.5 to +1.8 l/min
		Halifax n=40	Two point	-3.0 to +5.8 l/min
Bates (1973)	10 yrs	<u>VETERANS</u> n=149		RV = +0.13 DLCO = -2.62 FCO=-0.02
	FEV _{0.75x40} = 79 - 92.8 l/min			FEV _{0.75x40} VCMMF
	Age = 48.4 yr	Toronto n=36		l/m ml ml/m
		Winnipeg n=50		-1.9 -59 -33
		Montreal n=29		+0.7 -2 -16
		Halifax n=30	Endpoint	-1.3 -9 -47 -1.4 -62 -36
Bates et al. (1985)	Mean=15.3-18.7x/6-8 yrs	<u>COAL</u>	Before	After After Retirement
	FEV ₁ =2769-3186 ml	Dead Nonsmokers n=103	-67.8	-68.9 (41.2)VC=-69.7 FCO=-2.9
	Age = 46.6-50.9yrs	Alive Nonsmokers n=38	-46.2	-47.7 (27.7)VC=-40.5 FCO=-3.3
		Dead Smokers n=110	-73.5	-78.4 (60.4)VC=-81.0 FCO=-3.4
		Alive Smokers n=146	-57.6	-63.7 (45.3)VC=-59.0 FCO=-0.5
			Two point slopes	
Chan-Yeung et al. (1981)	et 2/2.5 yrs	<u>GRAIN</u>	Age <30	Nonsmokers
	Age = 38.2-44.3 yrs	Grain n=34	+21.2 (96.9)	FVC=-7.4 MMF=-47.8
		Civic n=11	+56.9 (72.5)	FVC=+46.7 MMF=+16.2
			Age 34-49	Exsmokers
		Grain n=52	-14.8 (80.7)	FVC=-47.2 MMF=-80.7
		Civic n=18	-21.0 (89.7)	FVC=-49.5 MMF=-24.6
			Age >50	Smokers
		Grain n=31	-78.1 (90.3)	FVC=-76.9 MMF=-206.8
		Civic n=12	-20.1 (63.1)	FVC=-5.2 MMF=-94.1
			Endpoint	

Tabona et al. 3x/6yrs 1985)		<u>GRAIN</u> only		
	FEV ₁ = 3777 ml	Smokers n = 113	-37.2 (57.8)	MMF = -115.8
	Age = 38.7 (10.9) yr	Exsmokers n = 85	-17.2 (46.8)	MMF = -74.0
		Nonsmokers n = 69	-33.9 (82.8)	MMF = -89.4
			Endpoint (Age Adjusted)	
Schulzer M. 3x/6 yr et al. (1985)		<u>GRAIN</u>		
	FEV ₁ = 3723-3990 ml	Grain Nonsmokers n = 129	-31	FVC = -14
	Age = 27-43 yr	Grain Smokers n = 285	-37	FVC = -20
		Civic Nonsmokers n = 42	+4	FVC = +0.7
		Civic Smokers n = 40	-31	FVC = -26
			Regression	

CHAPTER VII

METHODS OF ANALYSIS

Introduction

It is intended in this thesis, to draw upon statistical methods that are commonly available in mainframe computer statistical packages. The majority of the analyses are based on ordinary least squares techniques, using the third edition of the SPSSX statistical package (SPSS Inc. 1988). The maximum likelihood approach to modelling was based on programs from the BMDP statistical package (BMDP, 1988). BMDP is unique among the more common statistical packages in that an iterative mixed effects analysis of variance procedure for modelling unbalanced longitudinal data has been developed (BMDP 5V, 1989). Software for more specific covariance modelling has not been freely available to the academic community. Such modelling can be developed through the use of specialized UNIX-based software (e.g. Cook 1982). Unfortunately, the memory limitations of the personal computer prohibit analysis of large longitudinal data sets.

All of the descriptive methods used will be dealt with in Chapter 8. The statistical techniques used for the main research hypotheses are described in this chapter. A linear relationship between FEV_1 and its slope, with years of follow-up and other covariates was used in order to accomodate the two point methods and to provide an intuitive basis for describing decline as a linear change of FEV_1 over time. Of particular concern is the validity of the coefficient of decline obtained. It is the accuracy of the regression coefficient and the inferences derived that is emphasized (Kleinbaum, Kupper and Muller 1988). The prediction capabilities of the model are also of interest when it is the significance of the model variables that are ascertained. For all analyses, the level of significance (α) is established by the probability being less than ".05" of incorrectly rejecting the null hypothesis when it is actually true.

The analysis undertaken on the longitudinal data sets are described in two sections. The first part addresses the comparability of results derived from using different analytical techniques to estimate the decline of FEV_1 with time. The appropriateness of the method used will be judged, using a number of different strategies. Within each data set the decline estimates for each group will be compared as will the significance of any differences between the groups. Precision will be indicated by the standard error of the decline estimate as an indicator of the confidence intervals of the decline estimate. The standard error of the estimate and the coefficient of determination (R^2) will be used as comparative indicators of the fit of the linear model. The appropriateness of the model will also be evaluated using *a priori* knowledge of expected characteristics of behaviour. For example, it is most commonly observed that the decline of FEV_1 with time occurs more rapidly among smokers than non-smokers. The magnitude of the decline estimates will be compared in terms of their raw values before and their adjusted values after controlling for predictors of decline.

The second part of the set of analyses will rely on the more suitable methods of analysis to address the particular questions of interest. The issues to be addressed include:

- 1) Modelling nonlinear decline;
- 2) The interrelationships of pulmonary function variables, including the prediction of initial pulmonary variables on FEV_1 decline, as well as the decline of other pulmonary variables apart from FEV_1 ;
- 3) A further examination of the effects of smoking on FEV_1 decline, including the evaluation of its behaviour as a time varying covariate;
- 4) Determining the predictability of variability of decline;
- 5) Evaluating pre- versus post-retirement decline in FEV_1 among the *COAL* workers.

Comparison of Methods

The methods chosen for the comparative analysis are taken from those most commonly cited in the pulmonary literature, supplemented by modifications of ordinary least squares methods as devised by the author, and also including more recent statistical techniques specific for longitudinal data analysis. The following methods have been evaluated for the comparability of the derived decline estimates:

Simple Ordinary Least Squares Methods

- A. A "cross-sectional" regression of all available points;
- B. Regression of all adjacent two-point change estimates;
- C. Calculating slopes derived from the subtraction of the first from the last measurement;
- D. Calculating percentage declines based on the difference of the first and the last measurement as a percentage of the first measurement, divided by the years of follow-up;
- E. Linear regression of each individual's measurements upon time of follow-up;

Additional Statistical Approaches

- Transformations, weighted averages and weighted least squares estimates of the regression coefficients
- Application of the random effects and unstructured covariance generalized least squares models.

For each method, the estimate of decline is modelled for the simplest relationship, that between the dependent variable of lung function with the independent variable of time elapsed; for the adjusted estimate of decline obtained by adjusting for age and height differences between groups; and finally for the most complete model in which all the independent variables of interest are adjusted for. The simpler models can be used for all methods of analysis and among all the different data sets. The more

complete models differ in the number and type of adjustment variables used. As well, full use of these available covariates was dependent on the type of method used. Despite the availability of diverse pulmonary function measures no adjustment for any of these was used. In accordance with the conclusions expressed by Vollmer (1988) the decision not to adjust for initial lung function was based on observations that bias may result if either the distributions of the true initial values differ between comparison groups, or the true relationships between change and initial value differ.

The age- and height-adjusted coefficients of decline were obtained by forced entry of the coefficients for age at first measurement and height directly in the model. Apart from being included in all three data sets, they are generally used as the most important predictors of the FEV_1 measurements. For the methods comparison exercise, variables in the expanded models were also forced into the model directly, rather than including interaction and higher order terms to allow ease in interpretation. Age at the start of the study was used as a covariate rather than age at each measurement occasion; if this was additionally used, the high correlation with years of follow-up would introduce bias to the time coefficient in the model due to the collinearity.

The fullest model obtained for each data set was derived upon forced entry of the following variables, in addition to age at first measurement and height:

- 1) The *COAL* data: Unlike the *GRAIN* data, where the intention was to include only those workers measured over the entire follow-up period, the *COAL* data was extremely imbalanced. As individual follow-up periods range from 9 to 30 years, and decline estimates may be affected by the length of follow-up, YRDIF was considered to be an important variable to include. The variable RETYRS, which is the number of years from retirement to the last measurement occasion was also included. It was felt that retirement age, which was available in the data set gave very little differential information, since most coal workers were required to retire from coal face working at

the age of 50. Also, coal workers could retire at any point in time before or after follow-up. Therefore, the length of follow-up since retirement was used.

2) The *VETERANS* data: The continuous variables relating to smoking behavior included the age at which they started smoking (AGESMK) and smoke level (SMKLVL), an ordinal variable which indicates the type of smoking and the intensity of the cigarette smoking at each measurement occasion. SMKLVL was reclassified so that the level 0 indicated no smoking for the period of interest; 1 = pipe or cigar smoking; 2 = cigarette smoking at a level of 1-10 cigarettes a day; 3 = cigarettes smoked at an average of 11-24 a day; while 4 = greater than 25 cigarettes were smoked a day on average. The variable SMK4 described smoking behavior throughout the period of follow-up, with 0 = Never smokers; 1 = No smoking during the follow-up period; 2 = Stopped or intermittent smoker; 3 = continuous smoking. An independent variable of interest that was possible only in the *VETERANS* data, was a measure of the variability of decline (FEVCV) which is the coefficient of variation, (the standard deviation divided by the mean) for the first three measurements of FEV₁. The first three measurement points for each individual in these data sets was uniformly taken within a period of two years; Only the first three points were chosen because there was little evidence of a consistent decrement in the level of FEV₁ over that short period of time. Finally, YRDIF was included as the total length of follow-up differed between cities, in addition to between-individual differences.

3) The *GRAIN* Data: Two smoking variables were considered. Information on the actual number of cigarettes smoked during the preceding follow-up period was not consistently available. Therefore, the level of cigarette smoking (CIGLVL) was treated as a continuous scale covariate according to the ordered smoking categories, where 0 = never smoked or did not smoke during the time period of interest, 1 = smoked an average of 1 - 9 cigarettes a day, 2 = smoked an average of 10 - 20 cigarettes a

day and 3 = smoked at a level of 20+ cigarettes a day. Only a small proportion of cigarette smokers in the *GRAIN* data set (3%) also smoked pipes or cigars; the majority smoked cigarettes only. Information on age at which a subject started smoking was available for two testing periods, but the agreement was so poor, as well as inconsistent with smoking status that this variable was not used. Another variable, SMK4 categorized the type of smoker into continuous, intermittent, ex-smokers and never smokers. The high correlation between SMK4 and CIGLVL at the first measurement occasion ($r = 0.78$) had a potential of introducing collinearity which could bias the estimates, therefore only the variable with greater explanatory power, that is CIGLVL, was retained in the model according to its value at the first measurement occasion. The relatively high numbers of never and non-smokers in this data set (53.2%) contributed to this inter-correlation. YRDIF, the total number of years of follow-up, was also included, as despite the balanced structure of the data set, the total years of follow-up ranged from 8 to 10 years.

Ordinary Least Squares Methods

Cross-Sectional Approach

Model A: Regression of All Data Points

The "cross-sectional" approach to longitudinal data analysis is a contradiction in terms, in that the longitudinal time dependent characteristics of the data are ignored, and each individual measurement is treated independently regardless of whether it was obtained from the same individual. A linear regression was conducted, where the dependent variable (Y) was FEV₁, and the independent variable of interest (X) was number of years of follow-up from when the first measurement was taken. The equation with an unadjusted regression coefficient (β_1) is:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

To allow for comparison of group regression coefficients, a dummy model approach was used. The simple regression model then becomes

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \epsilon$$

where for a comparative analysis of two groups the dummy variable Z takes on the value of 1 or 0. This allows for the ascertainment of parallelism, coincidence or interaction. A significant interaction term indicates slope differences, while a significant $\beta_2 Z$ term in this example, would indicate differences in the level of the dependent variable for the two groups compared. For comparison with the longitudinal estimates, the more traditional cross-sectional approach was also used, in which the initial values of FEV₁ were regressed on height and age at first measurement.

Model B: Regression of Adjacent Two Point Change Estimates.

For the previous "cross-sectional" model, the relationship of interest was the effect of years of follow-up on the levels of the dependent values of FEV₁, with no assumption of dependence. An alternative model describes the relationship of the linear regression of time elapsed between every two measurement points with the change in FEV₁ over these two measurement times, assuming independence between pairs of values. This model takes the form:

$$(Y_t - Y_{t-1}) = \beta_1(X_t - X_{t-1}) + \epsilon$$

This approach allows pairing of each individual's data points so that they are considered together rather than each being an independent observation. According to Neter et al. (1985) this technique can be used to account for first order autocorrelation in the data. However with the measurement intervals for the *VETERANS* and *GRAIN* data sets being mostly equidistant, this method would be most appropriately applied to the *COAL* data set, where the measurement intervals are highly irregular. Only the simple form of this model will be applied as adjustment for fixed covariates would result in regression coefficients that could not be meaningfully interpreted.

End Point Determinations

Model C: Endpoint Slope Determinations

The simplest way of obtaining a slope estimate on longitudinal data is to take the difference in outcome between the first and last times of measurement and divide this by the time elapsed between the first and final measurement. This method of slope estimation has been primarily used on follow-up data sets in which all individuals were measured at the same time on both measurement occasions.

To compare the decline estimates, an independent t-test was used, where the *GRAIN* workers were compared to civic workers. For the other three and four group comparisons, analysis of variance was used. To account for the effect of height and age on the slope determinations, an analysis of covariance was applied to adjust for the differing distributions of these covariates. The analysis of covariance approach was adopted after first checking whether any interaction between the covariate and grouping factor was significant. This was not found to be the case. Therefore, for example, the relationship of initial age and height to the dependent variable of slope within one group was found to parallel that of another, that is, the coefficients of the relationship did not differ significantly.

An alternative two point estimation procedure is to divide the sum of the differences in FEV_1 between every two measurements by the total period elapsed, after weighting each difference by the length of follow-up for that interval. This results in an equivalent slope to the previous two point procedure, and hence was not calculated.

Model D: Calculating percentage declines.

The slope obtained by the end-point determinations represents absolute change. For example, a 1 litre FEV_1 decrement over 10 years in an individual whose initial value of FEV_1 was 4 litres, indicates a dramatic change by 25% of the initial value; but in another individual whose initial FEV_1 was 2 litres, the same decline corresponds to a 50% fall from the initial level. The endpoint derived differences in FEV_1 were divided by the initial FEV_1 value and then the number of years (and multiplied by 100%) to obtain the percentage decline estimate. The use of the percentage change in FEV_1 in this model therefore indicates relative as opposed to absolute change.

Individual Slopes

Model E: Linear regression of individual slopes.

An intuitive way to take the autoregressive characteristics of longitudinal data into account is to group the values separately for each individual. A linear regression of the individual's FEV₁ values over the time elapsed produces a regression coefficient (β_1) to describe the change in FEV₁ with time. Only the unadjusted regression coefficient can be obtained in this way, as any between-subject characteristics, for instance, height, will remain the same throughout the period of follow-up for that individual. Once each individual's linear coefficient of decline has been calculated, it remains to be decided how to combine these estimates to result in a group estimate of slope. The usual method of comparing group estimates of decline from slopes derived by individual regression analysis is to average them. An analysis of covariance can then be used to adjust for the effects of specific covariates of interest on the dependent variable, which is the individual slope estimate.

Additional Statistical Methods

In order to improve the fit of a model, a suggested procedure is to transform the dependent variable and/or the independent variable in order to achieve a stronger linear relationship between the variables, normality and/or homogeneity of variance. The most common transformations used for the dependent variables in order to satisfy OLS assumptions are the square root of Y, $\log_{10} Y$ or the inverse of Y (Neter et al. 1985). These were attempted in order to normalize the distributions, which showed evidence primarily of kurtosis. A difficulty with using any such transform of the dependent variable is that the coefficients obtained are difficult to interpret.

The ordinary least squares approach weights each observation equally. When the individual regression coefficients obtained are combined, their average is based on equivalent weights given to each individual coefficient. A weighted average procedure can be used to give more emphasis to more precise slopes, or those that have certain desirable characteristics. This can be achieved by weighting the slopes according to the reciprocal of the SEE obtained by regression of each individual's data. Clement and Van de Woestijne (1982) used such a method of weighting each individual slope by the inverse of its sample variance. Another possible weighting scheme to use, which is opposite to that of the previous one, is to weight by the actual SEE, rather than its reciprocal. The rationale for this is that individuals who do not follow a straight line decline closely, but rather are more variable in the decline, or who follow a non-linear decline, may be experiencing a greater decrement in lung function than indicated by a straight line with time. The length of follow-up provides another weighting scheme. Those regression estimates obtained for individuals who are followed up for a longer period of time may justifiably deserve greater weighting. Squaring the time variable further emphasizes the weighting and provides a fourth scaling scheme.

Weighted Least Squares estimates

The least squares procedure weights each observation equally. Weighted least squares is most often used to attempt to correct for non-constant error term variance. The weight given to an observation is then the reciprocal of the observations error term variance which is typically unknown. However where the error term variance varies with the level of an independent variable in a systematic fashion this relationship can be exploited (Neter and Wasserman 1985).

Residual analysis, as conducted in Chapter 8, and the analysis of the standard deviations of the year of follow-up means showed a general pattern, where for the

COAL workers, variance of the error terms appeared to diminish with the longer periods of follow-up (YRDIF); in the *VETERANS* there appeared to be a gradual increase in variance by year of follow-up; in the *GRAIN* workers, a more pronounced increase in the standard deviation of the values was observed over time. Given these different patterns, three different weight factors were applied in a weighted least squares analysis based on the overall regression of all data points, in conformity with the residual analysis. The reciprocal of YRDIF, and the square of that estimate, were used to provide more weight for the measurements taken closer in time to the the initial measurement. The use of YRDIF alone as the weighting factor would serve to emphasize the values taken later than the first measurement. A weighted regression analysis of the data was invoked by using the REGWT option in the SPSSX regression package.

Random Effects model

The random effects approach can more accurately estimate the precision of the decline estimate by estimating the components of variability found between- and within-individuals. The option available through BMDP 5V to analyse this data is the random coefficient growth curve model (BMDP 1988). The inputted design matrix takes the form where each first column is a series of ones while in the second column the number of years of follow-up of the successive times of observation are listed. Thus, the follow-up sequence using this program must be the same among the individuals of the study group. A desirable property of the BMDP 5V program is its ability to handle missing responses. It does so by imputing values in which an estimated conditional mean of the missing response, given the value of responses that are present for that subject, is estimated. Where all the responses are missing for a given subject the imputed values are estimates of the unconditional means of the missing responses. The final estimate of the coefficient of decline resulting from a restricted maximum

likelihood estimation procedure is the generalised least squares estimate. An estimate of the variation of the individual slopes about the regression line is available, as is the log likelihood estimate. The asymptotic standard errors provide a measure of precision of the estimate and a Wald test of significance is used to test whether the fixed effects or covariates are significant according to a chi-squared test. Because each of the data sets was much larger than it was possible to analyse using the PC version of the software, and no MTS version of the program is available, a 1988 version of UNIX BMDP was run using the S.F.U. UNIX system (this software was unavailable at U.B.C.); as well BMDP Statistical Software Inc. was approached to run some models on the largest data sets, using their 1990 Beta version, which is, as yet, unreleased to the scientific community. These software limitations have prevented a complete comparison of the random effects model with the other methods described.

Using this approach, including an interaction term between the groups and time elapsed allows one to estimate not only if the generalised least squares coefficient for time is significantly different from zero but also whether the group coefficients of decline differ significantly from each other. The separation of variability of values about the individuals' regression line from the variability of the individual regression lines about each group estimate of decline, allows for greater accuracy in the estimation of the standard error of the coefficient of decline.

An unstructured, fully parameterized covariance structure was an alternative means of modelling the data which also provided a comparison for the goodness of fit of the random effects model using Akaike's Information Criterion (AIC). The model yielding the less negative value of AIC is considered to have the most appropriate covariance structure.

Criteria Used for Comparison of Methods

To address the question of the suitability of the chosen method of measurement of FEV_1 decline, statistical methods used to assess goodness of fit are considered with regard to the biological plausibility of the significant relationships. A statistical criterion adopted was the residual sum of squares, that is the squared relationship between the difference between the actual values measured and their fitted line. Due to the different degrees of freedom used among the models, the standard error of estimate (SEE) was evaluated as an indication of the average difference between the actual and predicted FEV_1 values. It is equivalent to the square root of the mean square error which is the sum of squared errors divided by the degrees of freedom. Associated with this measure is R^2 , which is the coefficient of determination. For a simple linear model, this statistic measures the strength of the linear relationship between the independent variable and the dependent variable of FEV_1 . For instance, it measures the proportionate reduction in the sum of squares of vertical deviations obtained using the least squares line compared to an attempt to predict FEV_1 if time was ignored, ie. using the mean of FEV_1 as a prediction of FEV_1 values. While R^2 does measure the strength of the linear relationship it does not measure the appropriateness of the straight line model in that a curvilinear relationship of points can still result in a high R^2 (Kleinbaum, Kupper and Muller 1988).

Another means of assessing the appropriateness of the obtained measure of FEV_1 decline is the precision of the estimate. For ordinary least squares and generalised least squares models the precision of the regression coefficient for time can be evaluated by looking at the 95% confidence interval of that estimate. For large samples, the confidence intervals are obtained by multiplying the standard error of the time coefficient by 1.96, the 2-sided 95%ile Z statistic.

Within each data set, the comparative values of the estimates of decline by each method for each group, can be evaluated in terms of the magnitude of the unadjusted coefficients, the age-height adjusted coefficients, and the relative values of the coefficients after adjusting for the independent variables that were available for each data set. Additionally, the statistical significance of the decline estimates along with each covariate will be compared between methods. It was decided not to choose the best predictor models for each method because the differing number of significant predictors in each model would invalidate goodness of fit comparisons.

Expectations of the behaviour of the estimate of decline of FEV_1 have been based on substantiated findings in the literature, as noted in Chapter 3. For instance, a coefficient of decline for a group of smokers would be expected to be greater than that found for a comparable group of non-smokers. Among the *COAL* data, those who had died by the ascertainment date would be expected to have steeper declines in FEV_1 than those who remained alive at follow-up. Also, *GRAIN* workers were expected to show greater declines in FEV_1 function compared to civic workers unexposed to respiratory hazards in their work place. The comparative order of groups will be evaluated with respect to such evidence of biological plausibility.

The particular technique used to estimate decline in FEV_1 will be assessed both for its applicability to one data set of interest, as well as on the ability of the method to handle the varying degrees of imbalance within the three data sets. The *GRAIN* handlers data set is relatively balanced but the follow-up period is short, and therefore very few missing values were found within the rectangular structure. The *VETERANS* data set is based on measurements made on a yearly basis. Thus the time intervals were equal but duration of follow-up was unequal. Drop-outs in the data set resulted in differing periods of follow-up for each individual. The *COAL* data set has the most irregular imbalanced structure, in that not only were the intervals between

the measurements different for each individual, but the duration of follow-up varied also.

Non-linearity of Decline

For the regression methods in which FEV_1 levels themselves are the dependent value, any effect of accelerating decline can be modelled by including a quadratic coefficient for time as well as a linear term in the regression model. Including the linear term in the model allows for a different time for the maximum level reached in the quadratic curve for each individual. The random effects model similarly can be used to test for acceleration of decline by determining the significance of a quadratic effect in the relationship of FEV_1 with time. Whether the quadratic relationship with time varies between groups can be tested for significance by incorporating dummy variables for the group factor in the model. While a quadratic relationship may not be the most biologically plausible explanation of how the measured lung function variables change over time its significance in the model would indicate that the usual straight line relationship is not suitable for the affected study groups.

Describing a nonlinear relationship of FEV_1 with time through a quadratic model limits the analysis to an imposed shape of decline which steepens with time. Because there is some limit in the level of FEV_1 that a person must maintain in order to be alive, describing a quadratic relationship of decline may be pertinent only to a certain stage of the individual's time of follow-up. An exponential decline, on the other hand, allows for a gradual decrease in the rate of change of the dependent variable with time as the years advance. Allometric decline can be modelled by regressing the logarithm of FEV_1 by the logarithm of the year of follow-up, after shifting the time scale by one year to avoid calculation the logarithm of zero. The resulting coefficient relates the percentage of change of the FEV_1 variable with the percentage change of

the time variable. The logarithm to the base ten was used in these analyses.

The Relationships of Pulmonary Function Variables

As the standard error of the measurement of FEV_1 is relatively small compared to that with other pulmonary function tests, FEV_1 is most often adopted as the decline estimate of interest; if any other lung function measures are recorded they are not usually analysed or published. Due to the collinearity of the initial lung function variables (Burrows et al., 1965) each was assessed separately in the models in order to avoid distortion of the coefficients.

Prediction was ascertained in part by correlating the initial lung function values with the decline estimate, and through partial correlation, after adjusting for other important variables. Whether initial lung function as a covariate was significant was based on regression techniques. All three data sets allow for the analysis of other initial spirometric measures on FEV_1 decline. The *VETERANS* data set in particular is unique in that a variety of pulmonary function measurements in addition to FEV_1 were recorded throughout the course of the study. Of particular interest are the hypotheses on whether the initial value of residual volume (RV) or steady state diffusing capacity (DLCO) or the fractional carbon monoxide uptake value (FCO), might be predictive of FEV_1 decline. As vital capacity for the *VETERANS* data set was determined separately from the FEV measurement on a closed helium circuit, it provides an independent measure from the calculated value of FEV_1 .

While the latter two measurements were recorded in the *COAL* miners data, their measurement was less frequent than the FEV_1 measures. The question posed for this data set is whether a measurement of a particular lung function variable at some point in the follow-up was predictive of the overall decline in FEV_1 .

As a substitute to modelling FEV_1 decline, FVC decline (VC in the *VETERANS* data) and MMF decline (which was not available for the *COAL* data) were modelled as was RV, FCO and DLCO from the *VETERANS* data, in order to determine if any significant group differences in decline estimates could be detected.

Practical Applications

Using the most suitable methods of analysis chosen in the previous exercise, each data set has been analysed with respect to specific characteristics of interest.

For the *COAL* data, the only smoking information available was a dichotomy by smoking status. Both the *VETERANS* and *GRAIN* data sets had smoking information at each measurement, which allowed a 4 level smoking type as well as intensity classification. At the minimum, each data set can be categorized into 2 smoking status groupings;

- 1) *COAL* - smokers and nonsmokers
- 2) *VETERANS* - non, ex-smokers and intermittent smokers versus continuous smokers.
- 3) *GRAIN* - non smokers during the follow-up versus intermittent and continuous smokers

The categories were based on the distribution of each type of smoking activity. The *VETERANS* were dominated by continuous smokers, while in the *GRAIN* data set a high proportion of the workers were nonsmokers. A two-way analysis of variance has been applied to ascertain the significance of the main effects as well as the smoking/group interaction. The amount of smoking categorized on an ordinal scale at each of the measurement occasions and treated as a continuous variable is an additional feature of the *GRAIN* and *VETERANS* data sets. The random effects method can be applied directly to model the time-varying covariates describing the relationship of change in smoking behavior with change in FEV_1 .

A measurement specific to the *COAL* data set is the age of retirement of the individual. What is most valuable about this data set is that measurements were taken both before and after retirement on the majority of individuals. A question of central interest is whether the decline in FEV_1 was of similar magnitude before and after retirement from dust exposure. The majority of coal workers retired at some point during their follow-up; for just over half of the group there were at least two lung function sessions before retirement and two lung function measurements after retirement, available for analysis. It is this subset of 173 coal workers that were analysed for changes in FEV_1 decline before and after retirement. The regression coefficients of FEV_1 values with time have been calculated so that pre- and post-retirement stages are represented by a dummy variable, in a trend analysis whose relationships with time determines if the slopes are parallel, coincident or show interaction (Veney and Kaluzny, 1984).

An assumption of all the data analysis used is that the imbalanced data is due to a process of being missing at random regardless of the reasons why the data is missing. In the *VETERANS* study, extensive follow-up was conducted to determine which individuals had dropped out up to that point, and to identify what the causes of drop out were. Whether the coefficients of decline are found to differ by these groupings and the extent of such differences have been assessed, using dummy regression techniques. The findings have much pertinence, as it is often concluded that having deaths occur in the set may bias the overall estimates of change. An imbalanced data set is an inevitable consequence of long term longitudinal studies; whether those lost to follow-up or died show different characteristics from the survivors has important implications concerning their analysis as part of the groups initially studied.

CHAPTER VIII

DESCRIPTIVE DATA ANALYSIS

The data sets used for this thesis are the result of the work of investigators who had collected the data over previous decades. There was therefore, no choice but to accept the data as "clean" in the sense that it is assumed that any lung function tests which were the result of a poor effort or equipment or procedural failure, were discarded and not included in the available data sets. The data set on the Grain Handlers and their controls was available on an SPSSX system file, while the other data sets were obtained as hard copy, which were then input into the U.B.C. main frame computer by the author. In these latter data sets, transcription errors could have occurred, therefore inputted data elements were each visually checked for accuracy.

After an initial check of the accuracy of the data, representativeness and comparability of the data was ascertained by evaluating the similarity of each compared group for such characteristics as;

- 1) Absolute values, such as initial lung function, personal characteristics, and symptoms; and
- 2) Proportional characteristics, exemplified by the correlation of FEV_1 with forced vital capacity or maximal mid-expiratory flow both overall, and for separate time periods.

The rationale behind this latter procedure is that all three lung function measures are calculated from the same forced expiratory flow procedures. Any deviation in these correlations from a relatively constant value could indicate procedural differences over the specified measurement interval.

Regression diagnostics were then applied to the data to assess whether they conformed to the assumptions used for statistical analysis. For the purposes of this

thesis, a minimum of three measurements at different times were required in order for a subject to be included in the analysis. Two points can only necessarily represent linear change; any curvature requires at minimum, a third point. Any subsequent analysis in this thesis will adhere to this criterion of including cases having a minimum of three measurement occasions.

Description of Data Sets

Coal Workers Data

The data were stratified in groups according to smoking status, and whether the subject was known to be alive or deceased at the date of ascertainment. It is possible that intermittent smokers were included in either of the two categories, since details on the amount of smoking could not be obtained, and there was no information taken during prolonged periods where no measurements were taken. The exact date of follow-up to ascertain the vital status of the coal miners was not available. Unfortunately, it is not known how long a patient lived after their last pulmonary function test, or how successful the follow up was for the groups examined.

The median period of follow-up for the entire study group was 17 years, ranging from 9 to 30 years. To conform to the criteria of being tested a minimum of three times the data on eleven of the original 396 subjects were deleted from the analysis. The number of tests for each subject ranged from 3 to 12. To assess the similarity of groups being compared, initial values for the characteristics and pulmonary function measurements of the groups were compared. Table 8.1 contains the means and standard deviations of the initial measurements of these selected groups. The statistical significance of group differences between measurements has been evaluated through the application of a one-way analysis of variance procedure (SPSSX3 1988). The characteristics

examined included age at first examination, height in cm, age at retirement, and the initial lung function measures FEV₁, forced vital capacity (FVC) and a record of steady state diffusing capacity (DLCO) and fractional carbon monoxide uptake (FCO%).

Table 8.1: Baseline Measurements of Coal Miners Study Group @

	Deceased Nonsmokers n=101	Alive Nonsmokers n=35	Deceased Smokers n=104	Alive Smokers n=144
Age (yrs)*	51.1 (9.0)	48.5 (6.9)	49.3 (8.0)	46.7 (7.8)
Height (cm)	168.3 (5.5)	168.2 (4.6)	168.7 (6.5)	169.1 (6.1)
Retirement * Age (yrs)	56.1 (5.3)	54.3 (3.6)	54.0 (4.7)	53.5 (4.3)
FEV ₁ (l) *	2.76 (.72)	3.12 (.65)	2.91 (.61)	3.20 (.65)
Predicted FEV ₁ (l) *	3.19 (.48)	3.26 (.31)	3.26 (.55)	3.37 (.52)
FVC (l) *	4.26 (.64)	4.53 (.58)	4.49 (.68)	4.63 (.74)
DLCO _{ss}	15.01 (4.1)	18.58 (5.1)	14.86 (3.9)	16.0 (4.0)
FCO (%)	40.1 (6.3)	44.67 (6.9)	39.1 (5.3)	41.7 (6.0)

@ = Average values are given for those with at least 3 measurements

* = p< 0.05 by one-way ANOVA

() = standard deviation

Prediction equation based on Knudson et al. 1976.

The Scheffe procedure was applied as a multiple range test to determine whether any group was significantly different from the others. The deceased non-smokers were significantly older on average than the living smokers when the initial measurements were taken and also retired at a later age than the smokers. For this same group, the initial FEV₁ and FVC were also found to be lower in comparison to the living smokers. The DLCO and FCO% values were found to be significantly higher for living nonsmokers compared to the other groups: both of these indices have been shown to

be lower in smokers than in nonsmokers in the general population (Bates 1989). When predicted FEV₁ levels based on age and height relationships are applied (Knudson et al. 1976), an overall significant F test is observed. A lower percentage of predicted FEV₁ would be expected in coal miners compared to a nonexposed population.

In comparison to the published data on baseline characteristics on the more complete data set (Chapter 6), the only noticeable difference was found of the average age at retirement, which was slightly younger (averaging up to 1.5 years) for the study data. No attempt was made to proceed with a cross-sectional analysis of the baseline data as the coal miners differed in the time of first measurement.

The Veteran's Study

Data on veterans measured in the four Canadian cities of Toronto, Winnipeg, Montreal and Halifax, were obtained on computer print-outs. In previous publications no direct description was given of the quality control procedures used prior to any data analysis, thus it must be assumed that the data recorded are representations of adequate pulmonary function tracings. Data on the personal and smoking characteristics of the subjects were available on written sheets which were subsequently converted to separate computer files, and later merged with the pulmonary function data. Actual birth dates were not available; the age recorded was equivalent to the age of the subject in 1961. A number of descriptors of smoking behaviour were available for the first measurement occasion, such as the average number of cigarettes smoked, whether they had smoked pipe or hand rolled cigarettes, the age of stopping smoking and the number cigarettes they used to smoke. For each subsequent measurement occasion only an ordinal variable which indicated quantity and type of cigarettes being smoked was available. Initial values of this ordinal level coding scheme were checked for consistency with the available information on the first measurement occasion, and were altered to reflect that

condition in less than 10% of cases.

During the period of this study the measure of forced expiratory volume used was the $FEV_{0.75}$ which was then multiplied by 40 to derive the indirect maximal breathing capacity. The relationship of this derived measure with FEV_1 is so close that many European investigators have reported that forced expiratory volume in terms of an indirect maximum breathing capacity (Burrows et al. 1965). FEV_1 is now the most commonly used measure of forced expiratory volume. A simple conversion was used on the $FEV_{0.75 \times 40}$ measure to convert it to FEV_1 : the equation used was $FEV_1 = 0.0281 \times FEV_{0.75 \times 40}$ (Miller et al. 1959). While these measures have been found repeatedly to be highly correlated, they are not exact duplicates of one another. $FEV_{0.75}$ and FEV_1 correspond to different points on the expiratory curve that relates expired volume to time. The slope of this expiratory curve can be affected by differences in age, sitting height and different exposure histories, and the relationship may not always be truly linear (Pearson et al. 1966). Despite this limitation the derived FEV_1 measure will be used for the Veteran's study data, as it allows comparison to other study results and is a more familiar measure. A point in the procedure used was that the VC recorded in each centre was not the usual FVC from the fast expiration but a slow maximal VC measured three times on the helium closed circuit.

To ensure comparability of results between centres much effort was put into the planning of this multicentre study to ensure consistency in subject selection and in the pulmonary function testing. The equipment used was built in one location and calibration and all technicians were trained in one laboratory to ensure uniformity of method and procedure. An effort to reach consistency was recognised to be a difficult task.

"To reach the position where different laboratories will produce the same results on the same patient a number of conditions have to be met. First,

the respiratory manoeuvres requested must be identical. Secondly, the apparatus must respond in the same way to same stimulus. Thirdly, the results have to be calculated in a standard way expressed in conventional units." (Laszlo 1984)

In all centres, the same laboratory was used for all measurements. In addition in both Winnipeg and Halifax the same technicians who had been initially trained, made all measurements for the entire span of the study. In both Toronto and Montreal, changes in the technician staff occurred. An initial check of the quality of the data was made by comparing the baseline characteristics of the veterans studied at each city. Table 8.2 displays the mean initial values and their standard deviations for those subjects who had three or more measurements over the course of the study. In total the data on 11 subjects with two measurements were deleted from subsequent analysis. There were no significant differences according to an analysis of variance test applied to the smoking characteristics data in each city group. However, Montreal stood out as having data from veterans who were significantly younger, shorter and lower in weight than the participants in the other centres. All the mean baseline lung function values are averages of up to twelve monthly measurements made on each subject. The ventilatory function values of FEV_1 , MMF, VC and FEV_1/VC from Winnipeg were consistently higher. For the lung volume measures and diffusing capacity, Montreal stood out as having lower lung function than the other cities. The Toronto data also showed a few lower lung volume measures. The discrepancy between predicted and actual initial FEV_1 is noteworthy as is the fact that the between-city differences in predicted FEV_1 levels were not significant.

When compared to the baseline characteristics of the original veterans studied (Chapter 6), very few differences between any of the values were noted. While the FEV_1 measures were not directly comparable, the maximal discrepancy was for initial MMF, where the average of 2.85 l/sec for the Winnipeg group studied was .4 l higher

than that of the original population.

A cross-sectional averaging of all Winnipeg ventilatory function values comparing the pre-'69 data to the post-'69 data showed an increase in the average values with time. The source of this discrepancy was investigated and it was discovered that a substudy had been undertaken in Winnipeg where both pre- and post-bronchodilator tests were conducted. The values listed after 1969 were mistakenly given as the best of four lung function attempts over both conditions while those values listed before 1969 were the best of two pre-bronchodilator attempts. The incorrect FEV_1 and MMF data for Winnipeg were therefore re-entered according to hand written sheets obtained from the original laboratory.

Table 8.2: Baseline measurements on the Study Group of Veterans @

	Toronto	Winnipeg	Montreal	Halifax
	n=51	n=56	n=34	n=34
Age(yrs)	47.9 (8.9)	49.3 (7.3)	45.3 (7.9)	50.7 (10.8)
Height (cm)*	67.9 (2.9)	68.8 (2.6)	66.7 (2.7)	67.9 (2.6)
Weight (lbs)	169.6 (29.2)	166.2 (25.0)	154.2 (23.2)	167.4 (28.0)
% Current Smokers	89.4	84.7	91.2	80.5
Age started smoking (yrs)	18.4 (3.5)	18.5 (3.6)	18.7 (7.2)	18.5 (6.2)
# cig/wk	119.6 (79.1)	102.1 (67.0)	99.6 (126.0)	101.4 (70.2)
Cough Age (yrs)	30.1 (8.1)	29.9 (7.3)	29.3 (7.7)	29.8 (9.0)
FEV ₁ (l) *	2.53 (.76)	2.94 (.77)	2.24 (.76)	2.19 (.71)
Pred.FEV ₁ (l) *	3.58 (.65)	3.66 (.52)	3.48 (.55)	3.46 (.49)
VC (l) *	3.75 (.80)	3.41 (.68)	2.68 (.79)	2.99 (.65)
MMF (l/sec) *	2.08 (1.17)	2.85 (1.25)	2.47 (1.22)	2.23 (1.17)
FEV ₁ /VC (%) *	67.35 (13.2)	86.32 (16.3)	85.56 (24.7)	72.80 (15.5)
ERV (l) *	.71 (.43)	.84 (.42)	.53 (.28)	.69 (.36)
FRC (l) *	3.19 (.72)	3.75 (.82)	3.40 (.98)	3.82 (1.01)
RV (l) *	2.48 (.68)	2.92 (.76)	2.93 (.97)	3.13 (.88)
TLC (l) *	6.23 (.91)	6.31 (.81)	5.60 (1.11)	6.12 (1.14)
RV/TLC (%) *	39.78 (9.0)	45.83 (9.2)	52.15 (12.1)	50.68 (8.4)
ME (%) +	47.0 (10.2)	50.3 (10.1)	52.8 (10.0)	48.0 (10.4)
DCO *	18.4 (4.4)	20.0 (5.2)	12.9 (3.2)	14.9 (3.7)
FCO (%)	.413 (.06)	.414 (.06)	.374 (.07)	.400 (.06)

@ = Average values are given for those with at least 3 measurements

* = p<.05 by one-way ANOVA

() = standard deviation

Prediction equation based on Knudson et al.(1976).

A further check of the quality of data between different centres was a linear regression of FEV₁ both with MMF and VC, which was compared between cities and from the first to second half of the study. The correlations found were of comparable magnitude to that found in cross-sectional data by Burrows et al. (1965). The results are presented in table 8.3 and show no clear differences in the correlation of these pulmonary variables between cities and within time periods.

Table 8.3: Correlation Coefficients for the Comparison of Lung Function Measures Between *VETERANS* Cities

CITY	FEV ₁ /VC			FEV ₁ /MMF		
	Overall	1st half	2nd half	Overall	1st half	2nd half
TORONTO	.75	.78	.67	.86	.89	.82
WINNIPEG	.75	.74	.76	.88	.87	.89
MONTREAL	.71	.72	.73	.91	.92	.90
HALIFAX	.74	.68	.73	.89	.88	.90

To further evaluate the FEV₁ and VC data, plots based on averages for each year for those subjects who had the majority of study measurements completed were made. As shown in Figure 8.1(a-d), unlike the other cities the average FEV₁ for Montreal was more variable, with an outlier found in the middle of the surveys. Outliers found for the other cities only appeared at the last follow-up occasion where the numbers were small. For vital capacity (Figures 8.1e-h), the Montreal data appeared to show an overall increase over the years. In view of the irregularities in the data and the fact that it was not possible to determine if they were due to technical problems, the Montreal data was excluded from any further analysis. Ensuring consistency in data quality over the course of a long term longitudinal study is a difficult task, but to also ensure comparability across different centres, creates additional problems. As noted earlier the Winnipeg and Halifax centers each had the advantage of employing the same technician throughout the entire study period. The subjects of the *VETERANS* data were followed up from 2 to 23 years in total (median = 12 years); with the number of yearly measurements ranging from 3 to 22. Veterans from Winnipeg were followed for the longest period; while in Toronto, the maximum length of follow-up was 14 years.

Figure 8.1a: TORONTO - Mean FEV₁ by Year

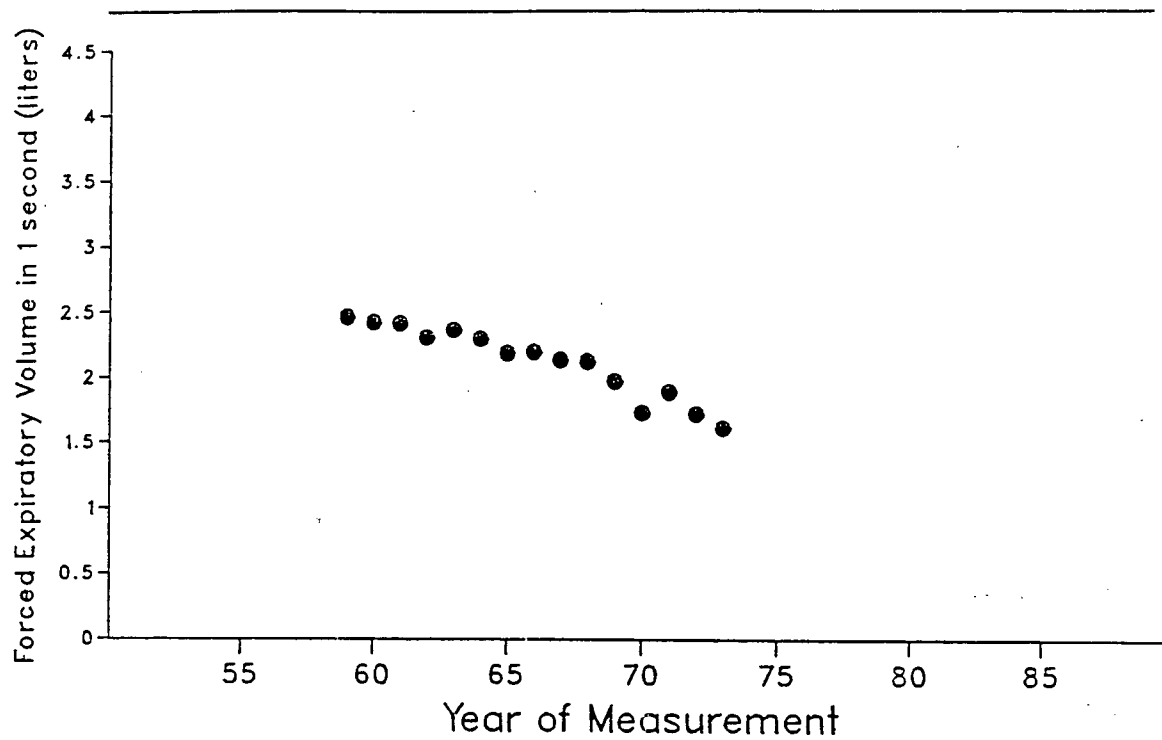


Figure 8.1b: WINNIPEG - Mean FEV₁ by Year

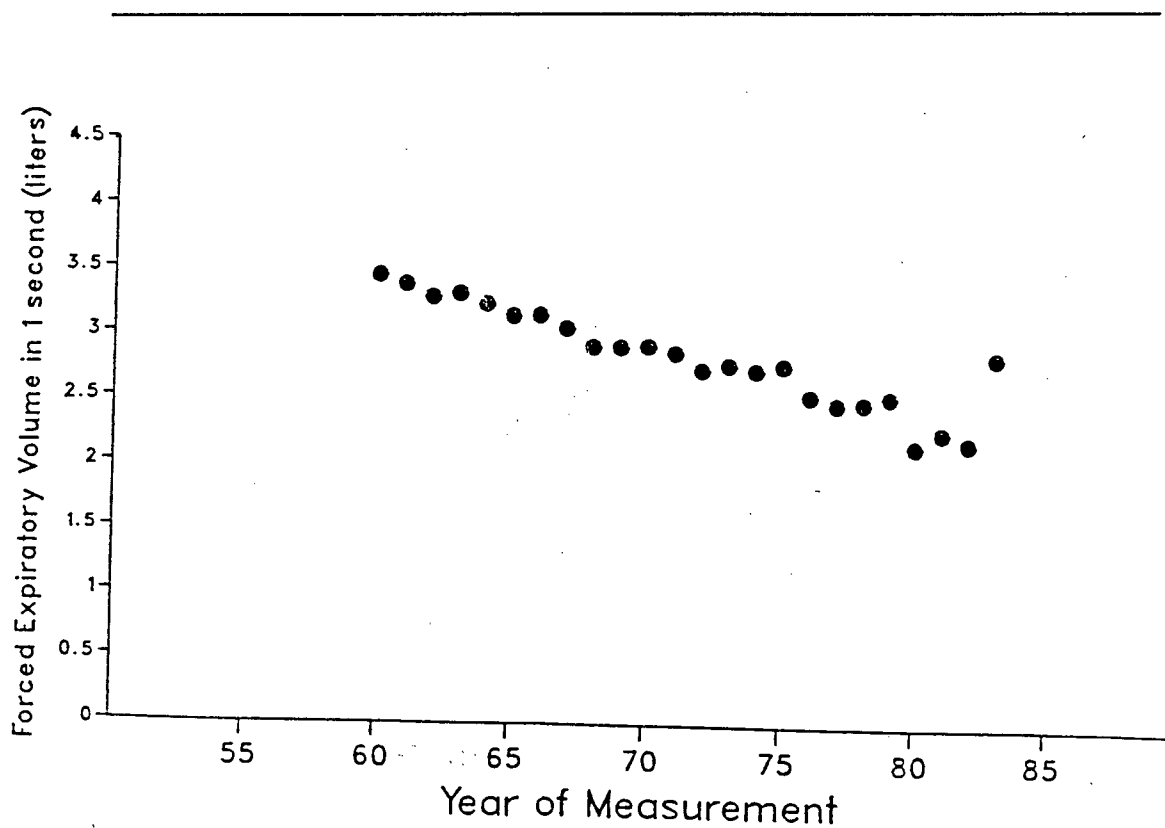


Figure 8.1c: HALIFAX - Mean FEV₁ by Year

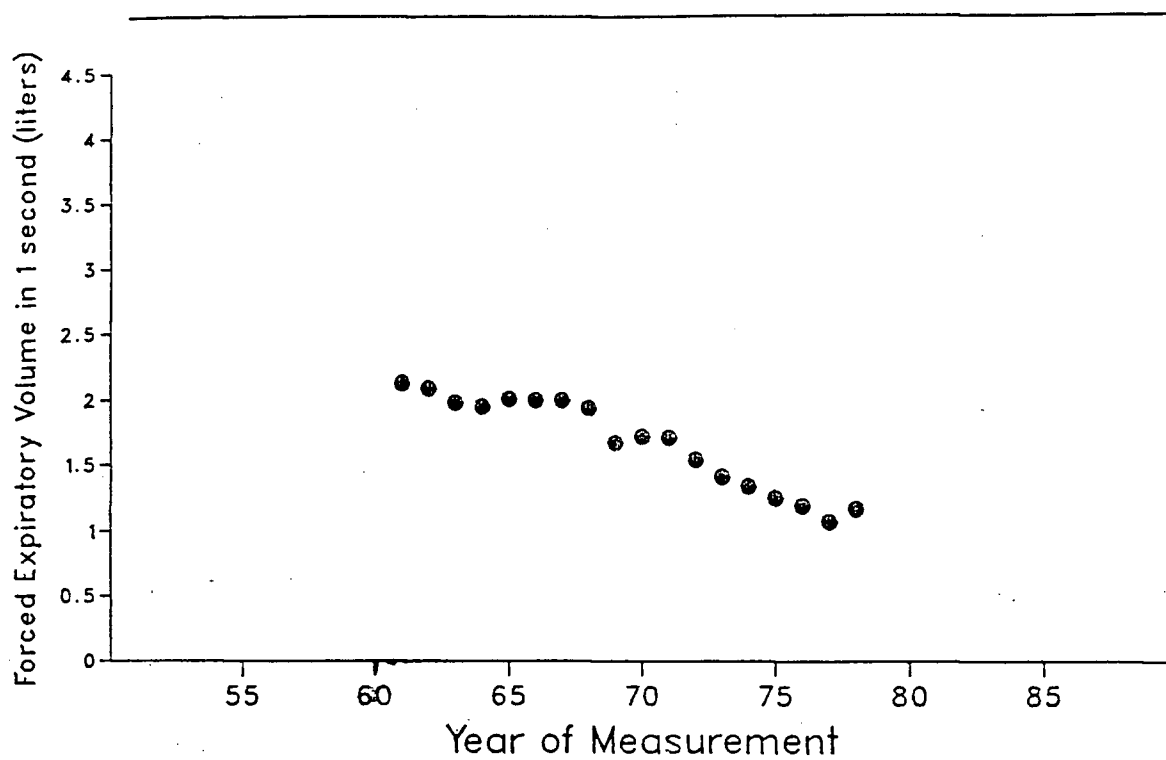


Figure 8.1d: MONTREAL - Mean FEV₁ by Year

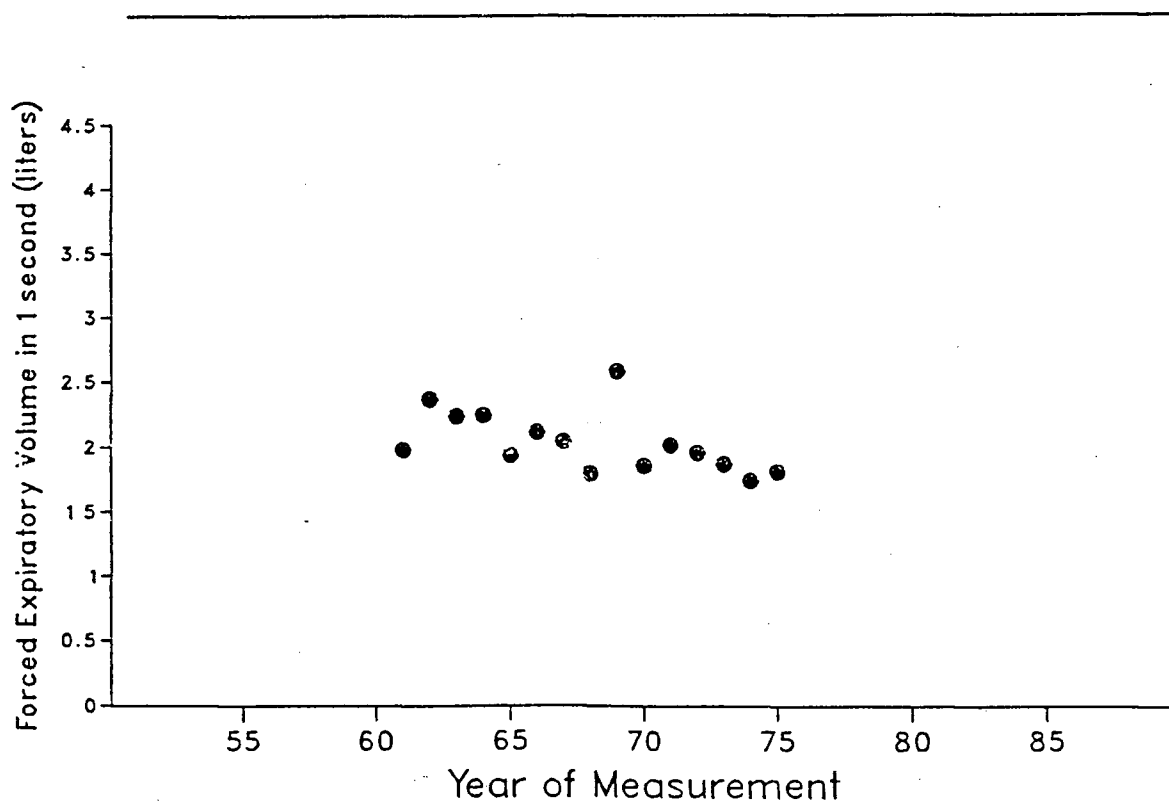


Figure 8.1e: TORONTO - Mean VC by Year

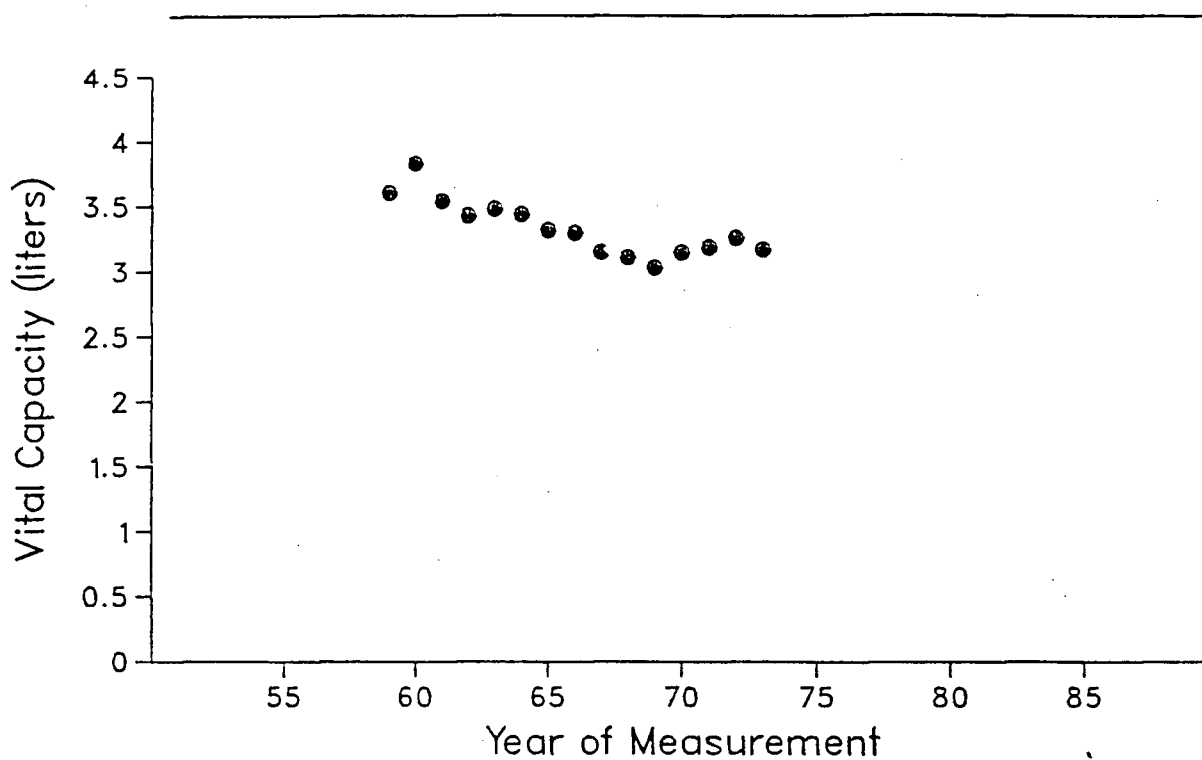


Figure 8.1f: WINNIPEG - Mean VC by Year

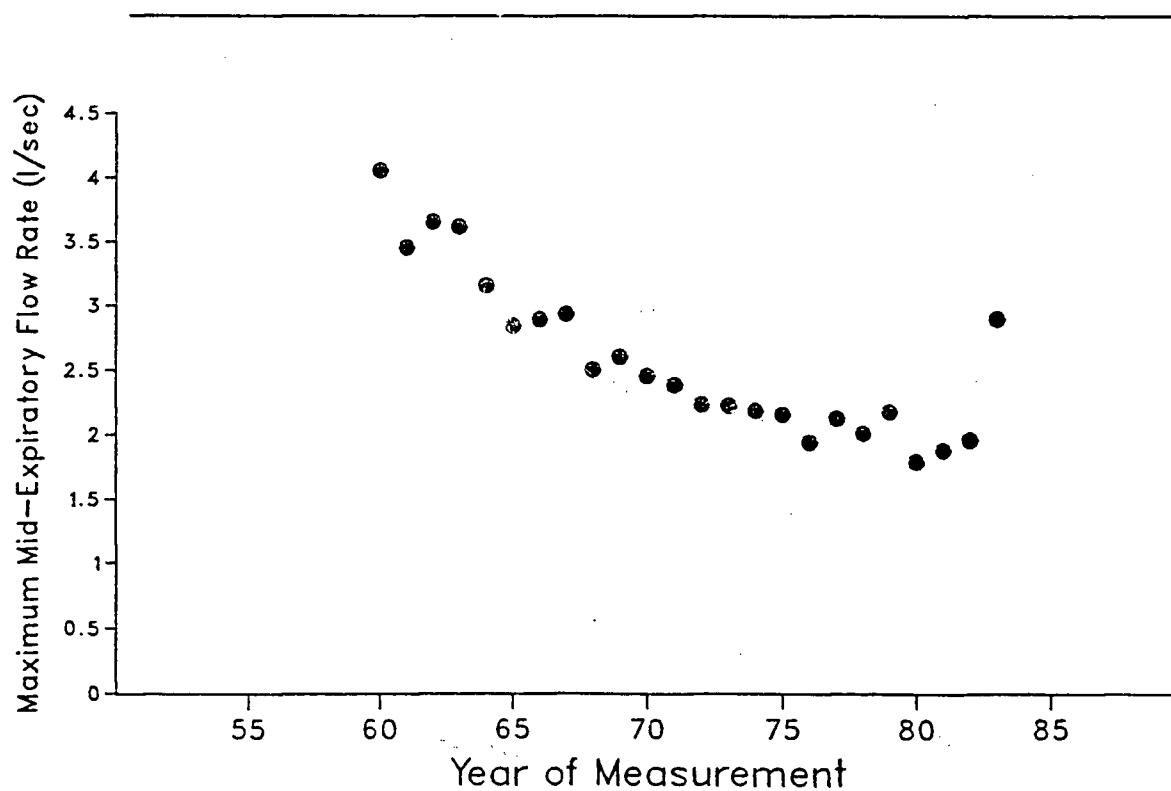


Figure 8.1g: HALIFAX - Mean VC by Year

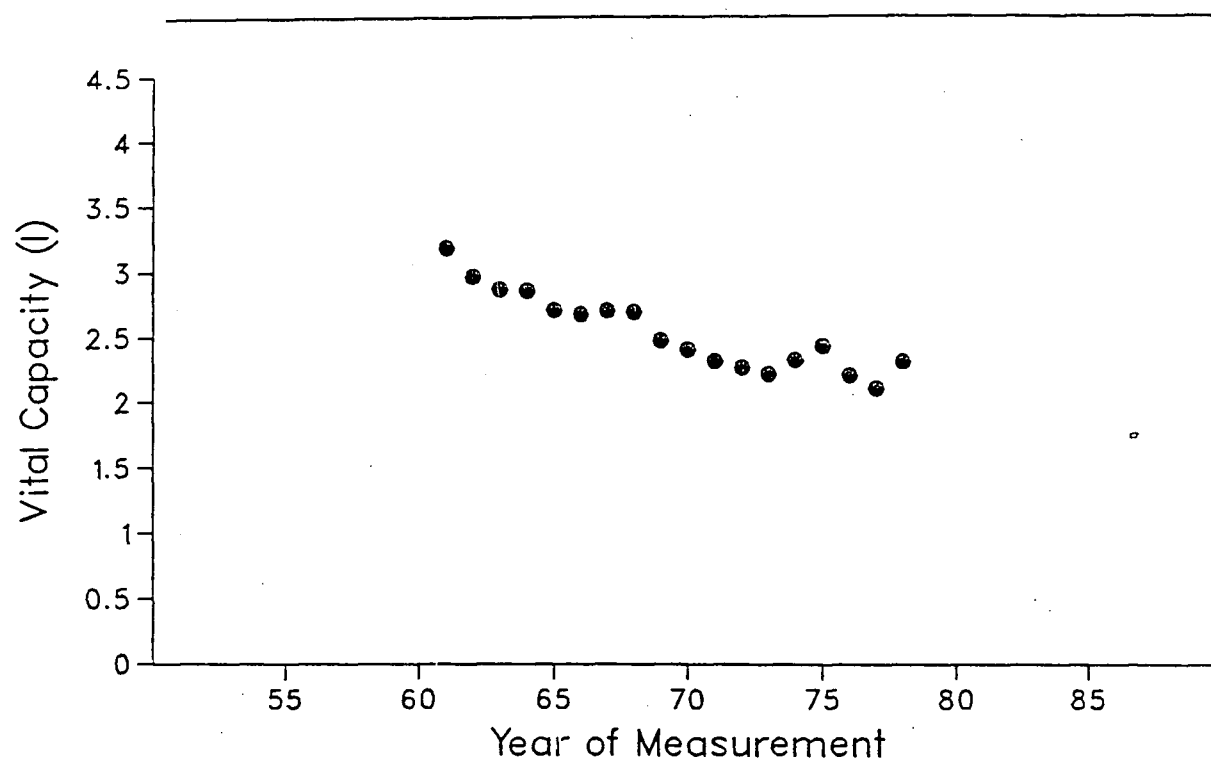
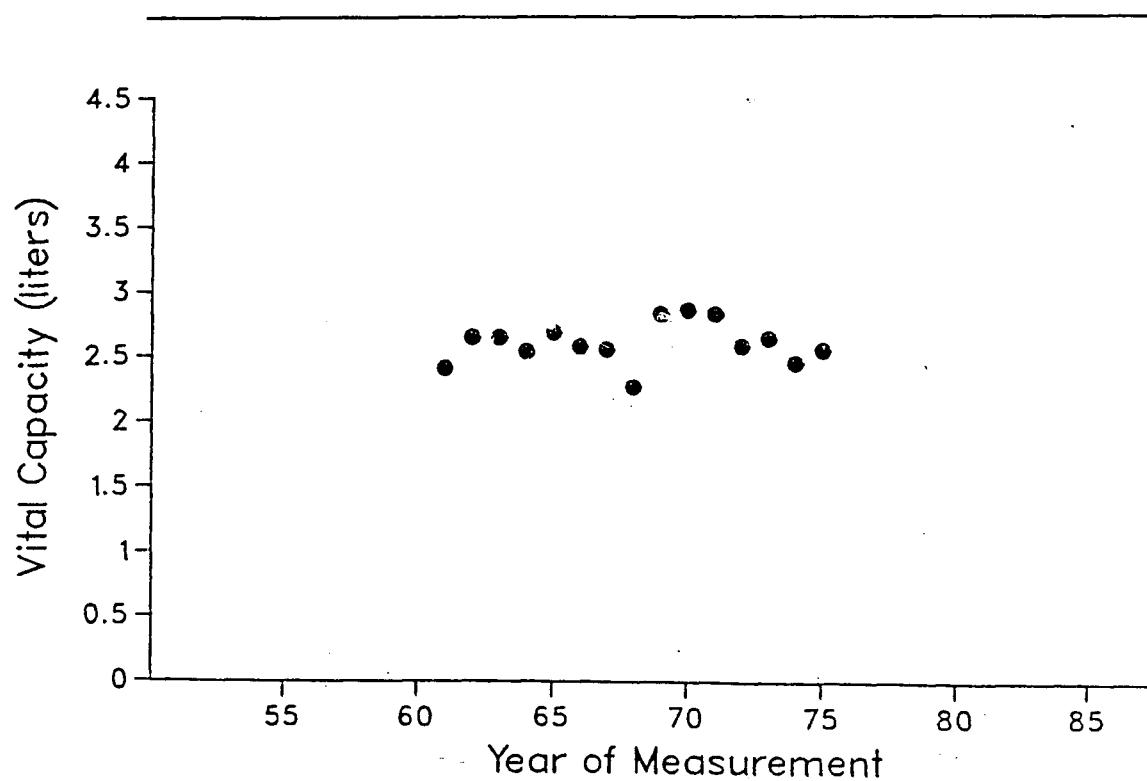


Figure 8.1h: MONTREAL - Mean VC by Year



Grain Handlers and Civic Workers

A subset of the data collected on the grain handlers in the Port of Vancouver and the City Hall civic workers was made available on an SPSSX system file. This consisted of those grain handlers and civic workers who attended at least 3 out of the 4 possible measurement occasions over the period from 1975 to 1984. From the original system file modifications were made to the missing data declarations and corrections made to transcription errors in the FEV₁ measurements taken at the second occasion on the civic workers. Detailed smoking information was made available for the first two measurement occasions on the grain handlers, but information on the quantity of cigarette smoking was limited to classification by ordinal levels, rather than a continuous measure. Characteristics of the working environment such as duration of employment or severity of exposure to grain dust, were not included in the available data set.

The appropriateness of the choice of the civic workers as a control group can be assessed in part, by comparing the initial characteristics between the two groups.

Table 8.4 shows the characteristics of the subset of individuals examined in at least three of the four measurement occasions taken during 1975, 1978, 1981 and finally 1984. The two-tailed t-test was applied to the data to test the significance of the average differences. On average, the grain handlers were younger, and although the differences were not significant they tended to be slightly shorter and had slightly lower initial FVC, than the civic workers.

The *GRAIN* data set was a relatively small subset of those originally studies (327 versus 746) and this was reflected in the relatively large differences in their average age at first measurement; particularly for the subset of civic workers studied who were younger than the original group (40.9 versus 44.3 years of age on average).

Table 8.4: Baseline Measurements of the Grain Workers and Controls Study Group. @

Examination.

	Grain Handlers	Civic Workers
n	269	58
Age (yr) *	36.2 (10.4)	40.9 (9.2)
Age started smoking (yrs)	16.9 (4.0)	17.2 (3.9)
Height (cm)	176.8 (7.1)	178.6 (6.3)
FEV ₁ (l)	3.88 (.73)	3.98 (.65)
Predicted FEV ₁ (l)	4.180 (.62)	4.170 (.54)
FVC (l)	4.893 (.83)	5.137 (.80)
FEF ₂₅₋₇₅	4.006 (1.18)	3.730 (1.05)
FEV ₁ /FVC%	79.2 (6.5)	77.7 (6.2)

@ = average values are given for those with at least 3 measurements
 * = p<0.05 by two-tailed T test
 () = standard deviation
 Prediction equation based on Knudson et al. (1976).

The same technician performed the measurements with the same equipment at each testing. However, as a check to whether differences could be detected in terms of patient effort and technician involvement, regressions were run on the relationship of FEV₁ with both FVC and MMF. Table 8.5 lists the Pearson correlation, the slope coefficient (β_1) and the intercept (β_0) for each lung function comparison using all available measurement points at each measurement occasion.

Table 8.5: Regression Coefficients for the Comparison of Lung Function Measures in the *GRAIN* Data Set

	FEV/FVC			FEV/MMF		
	r	β_1	β_0	r	β_1	β_0
GRAIN1	.91	.80	-89	.81	.49	1913
CONTROL1	.90	.75	+121	.79	.51	2070
GRAIN2	.91	.83	-169	.81	.47	2072
CONTROL2	.91	.79	-76	.82	.50	2148
GRAIN3	.89	.81	-176	.80	.46	2147
CONTROL3	.85	.78	-73	.80	.50	2250
GRAIN4	.90	.82	-193	.77	.44	2221
CONTROL4	.85	.78	-70	.81	.48	2224

No significant differences were observed between the correlation coefficients, as confirmed by the use of Fisher's *Z* transformation. Although the values do appear close, the slope of each comparison was consistently lower for the controls for the FEV₁/FVC comparison, and consistently higher for the controls when FEV₁ was related to the MMF values. This consistency in the differences may indicate a systematic difference in technique and/or effort of the participants involved, or to subtle differences in the maximal expiratory curve shape, although they do not reach statistical significance. A further indication of a possible discrepancy is revealed in a plot of the crude means of FEV₁, FVC and MMF at each measurement occasion. In Figures 8.2a-c, each average lung function value was obtained for all the subjects in the grain handlers versus the civic workers groups who had at least three of four measurements over the entire study period; those who had missed the first or fourth measurement occasion are not included in any further analysis. While the FEV₁ and FVC parameters were found to be on average to be consistently higher at each measurement occasion for grain workers, when the maximal midexpiratory flow rates were plotted the reverse situation was shown; that is, civic workers had consistently lower MMF values on average at each measurement occasion. Again, a possible discrepancy in the effort

displayed by the civic workers could have occurred. If the controls performed somewhat slower FVC manoeuvres, the resultant FEV_1 and FVC values would have been maximized, but the MMF values would be reduced due to a flattening of the maximal expiratory flow curve (Burrows B, Tager I; personal communications). A physiological basis for these results cannot be discounted however as the different flow-volume characteristics may have physiological meaning.

Figure 8.2a: Mean FEV₁ by Occasion for Grain and Civic Workers

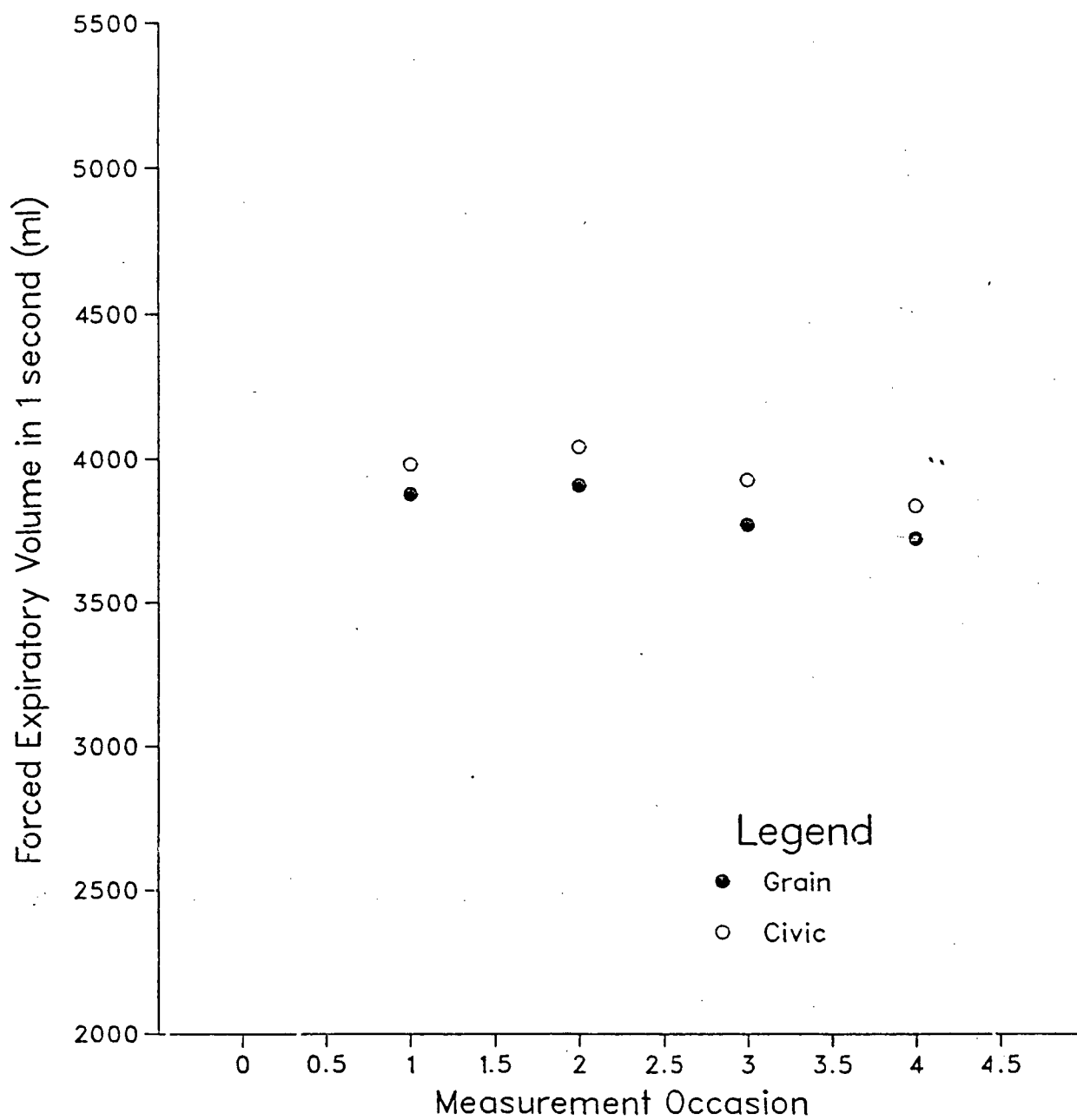


Figure 8.2b: Mean FVC by Occasion for Grain and Civic Workers

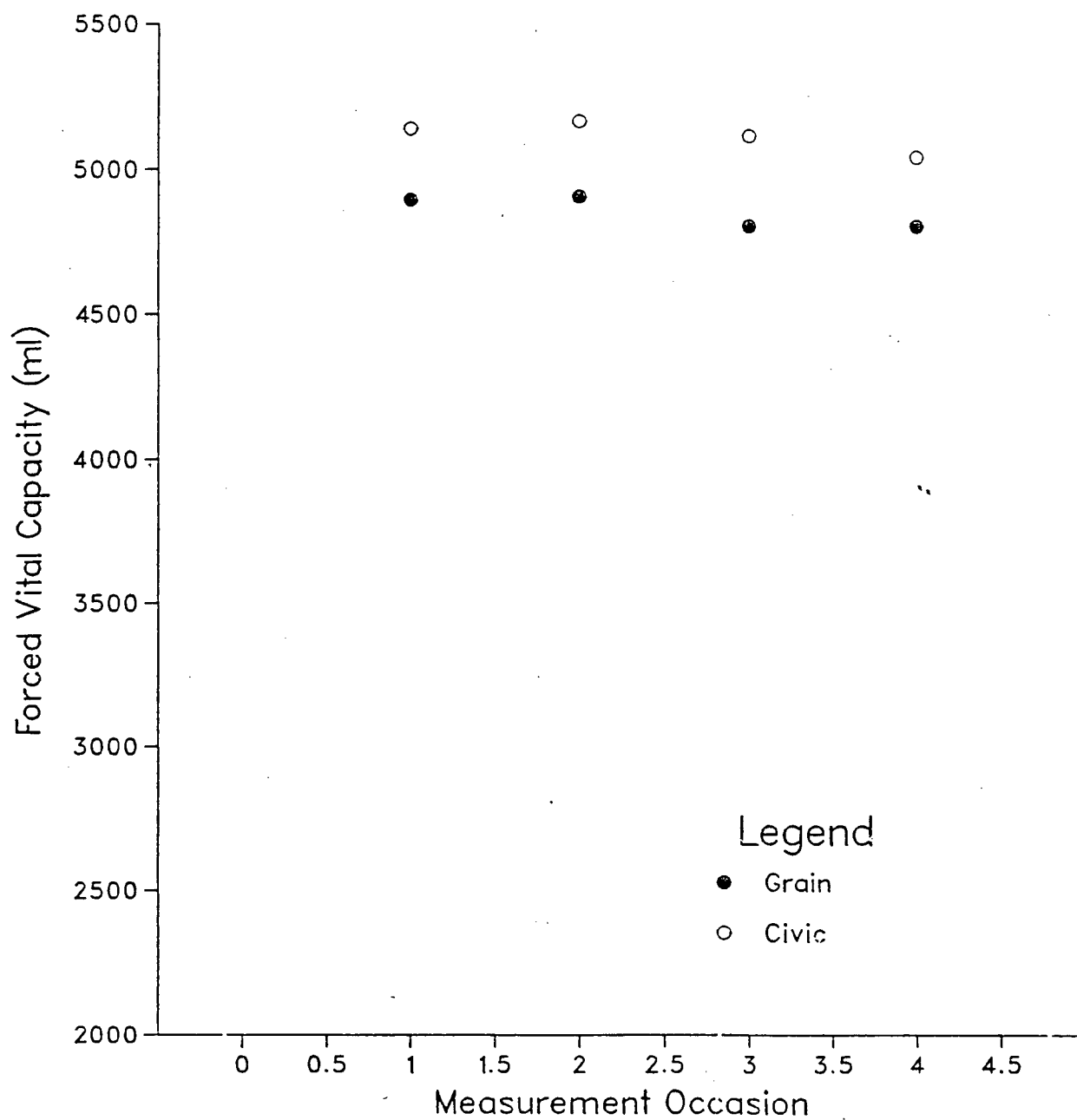
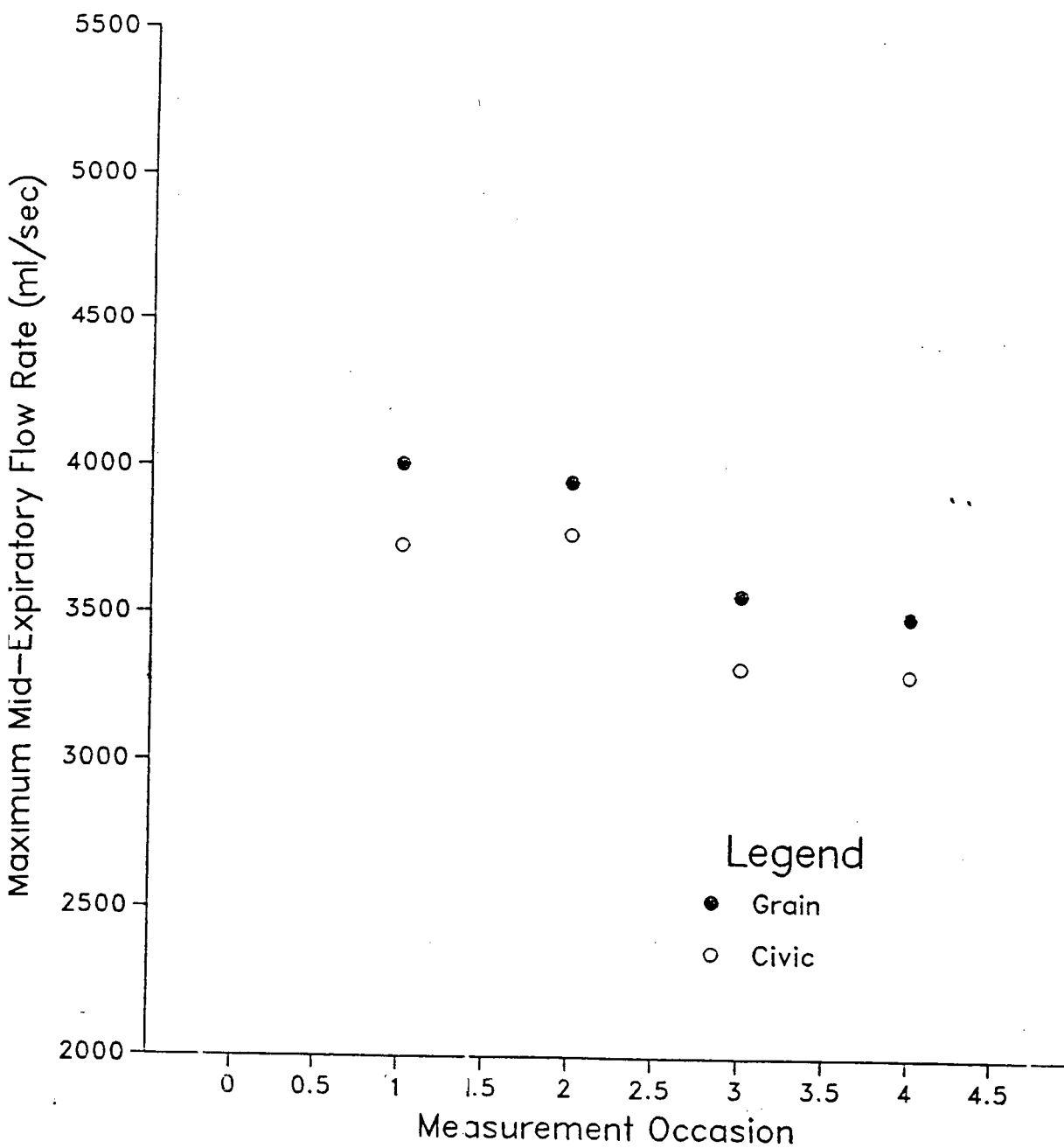


Figure 8.2c: Mean MMF by Occasion for Grain and Civic Workers



The assessment of the quality of each data set has depended on the values of the data points themselves. A residual is defined as the predicted value obtained from a regression analysis, subtracted from the actual value itself (Montgomery and Peck 1982). Plotting the residuals from the fit against other variables, permits a view of what has not been summarised by the straight line (Tukey and Wilk, 1965). Another useful property is its use to identify outliers in the data.

The identification of outliers is an integral part of the initial phases of data clean-up. One approach is reflected in an excerpt from the following article by Tukey and Wilk (1965).

"Using human judgement in selection of parts of the data for analysis, or in cleaning-up the data by partial or complete suppression of apparently aberrant values is natural, sensible, and essential. Data is often dirty. Unless the dirt is either removed or decolorised it can hide much that we would like to learn."

Depending on where an outlier is located in relation to the rest of the data, it can bias any or all of the the slope, intercept or level of any regression line drawn through the data (Daniel and Wood 1980). A more extreme approach to this attitude of deleting outliers is exemplified by the "winsorization" process, as demonstrated by the initial work of Fletcher et al. (1976) and modified by Kanner et al. (1979).

Discarding values which are plus or minus three standard deviations of the mean FEV_1 was a method that was found to improve the correlation of the slope with the level of FEV_1 .

Residuals that are considerably larger in absolute value than others (such as 3 to 4 standard deviations from the mean), can be classified as potential outliers (Montgomery and Peck 1982). Residuals formed as a result of a regression analysis performed on all the measurements are essentially "cross-sectional" in nature as each

data point is analyzed as if it was from a separate individual. Such cross-sectional residuals have their use in evaluating properties of the data for an ordinary least squares analysis, as will be demonstrated below. Longitudinal residuals, on the other hand can be defined as deviations of successive observations from their expected values, given previous responses and the values of the covariates included in the linear model. (Ware 1985). Complications that arise from evaluating longitudinal residuals is that their form, which depends on the autocorrelation characteristics of the data, may vary from subject to subject and from occasion to occasion.

In the few pulmonary function studies which have presented a detailed description of outliers, the option usually chosen was to remove them from the rest of the data set. For example, in a study of young children published by Strope and Helms (1984) seventeen studies on ten subjects were excluded after failing the criterion of being within 3.5 standard deviations of their mean residual value of zero. Each outlier was found to have unacceptable technical quality when subsequently examined. In a data analysis exercise of Mississauga fire fighters data undertaken by Kusiak and Roos (1984), a normal probability plot of the residuals was used to identify outliers subsequently removed from the analysis. Reanalysis of the data was continued in this fashion until the normal probability plots and other indicators were satisfactory. As a result, 45 lung function records were removed from the original 978. No attempt was made to identify the cause of these outliers. A more moderate approach to the decision of whether to discard outliers was adopted by Kanner et al. (1979). Values which fell outside plus or minus 3 standard deviations of a regression line based on the middle observations, were examined individually and were retained if the values were consistent with the subject's other data. In all, 19 observations out of a total of 1625 for each category were discarded. This resulted in a change in slope of the FVC in 9 patients and in the FEV_1 of eight patients.

The residual data analyses used in this chapter, were based upon an overall regression of all FEV_1 values against year of follow-up. Residual plots resulting from individual-based regression analyses produced comparable results. The identification of outliers in the present data sets was based upon a cross-sectional analysis in which each data point is compared to a predicted value based upon an overall regression of FEV_1 on time. Only two coal miners from the alive smokers group had values greater than three standard deviations above the mean. For the *VETERANS*, the most extreme value had a standardized residual of 2.87. Three grain workers had values greater than three standard deviations below the mean, thus their predicted values were much higher than recorded. The data for all of these potential outliers was found to be accurately transcribed from the original source, and thus have been retained in the analysis.

The removal of an outlier that cannot be attributed to measurement or recording error exposes the analysis to bias (Morrison-Ralph 1981). Smoothing of the data may obscure interesting (and possibly important) extremes in behaviour of the process observed. Montgomery and Peck (1982) recommend that there should be strong non-statistical evidence that the outlier is a "bad" value before it is discarded. Deleting outliers to improve the fit of an equation can also give an entirely false sense of precision in estimation or prediction. Outliers may control many key model properties, point to inadequacies in the model, and may yield valuable information concerning the regressor value which produces that outlier response (Montgomery and Peck 1982). The original pulmonary function tracings could not be checked for errors, and therefore a conservative approach was adopted, in which outliers that could not be attributed to errors in data transcription were identified but were not deleted from subsequent analysis.

Examination of residual plots also allows for an evaluation of any violations of the assumptions underlying ordinary least squares (OLS) regression analysis. OLS

regression analysis is the most common method of analysis of longitudinal data in the pulmonary function literature. It has often been used with little regard to the conformity of the data with the assumptions underlying the method. Least squares estimates for the parameter values as well as confidence interval and tests of hypothesis require the assumption that the residuals be independent and have a normal distribution with mean zero and variance σ^2 (Berenson and Levine 1983). For regression models summarised by a typical equation:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

it is assumed that for any fixed level of X, the subpopulation of independent Y values has a normal distribution whose particular mean, μ , changes linearly with X, but whose variance (σ^2) remains constant with changes in X. The assumptions of normality and of equality of variance do not require rigid adherence, as conclusions are not likely to be greatly weakened if the distributions are nearly normal or there is nearly constant variance (Daniel and Wood 1980).

Linearity

Apart from plots of the average values at each measurement occasion (Fig.8.1a-h), inadequacies of the linear model can be investigated through a plot of the residuals versus the corresponding fitted values. Residuals scattered about in a horizontal band about the level of zero are to be expected. A more parabolic form of the residual plot on the other hand indicates non-linearity. The addition of a squared term for the independent variable may be required (Montgomery and Peck 1982). Residual plots (shown in Figures 8.3a-c) resulting from an overall regression of all FEV₁ values with years of follow-up, showed no major deviations from the assumption. However, the plots for the *COAL* and *VETERANS* data sets showed a "truncation" of the positive standardised predicted values, which suggests some inadequacy of the model (a similar result was also found using residual analysis using model E).

Figure 8.3a: COAL - Standardized Plots of Residuals Versus Standardized Predicted Values of FEV₁

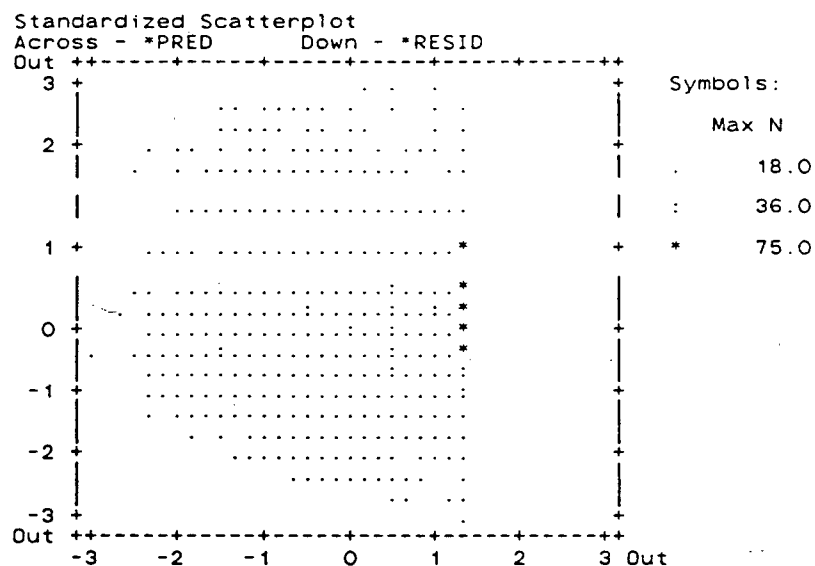


Figure 8.3b: VETERANS - Standardized Plots of Residuals Versus Standardized Predicted Values of FEV₁

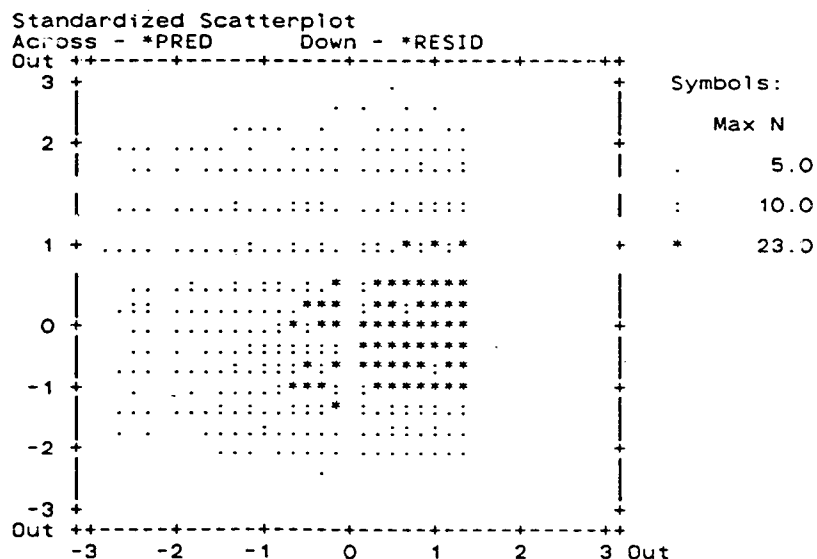
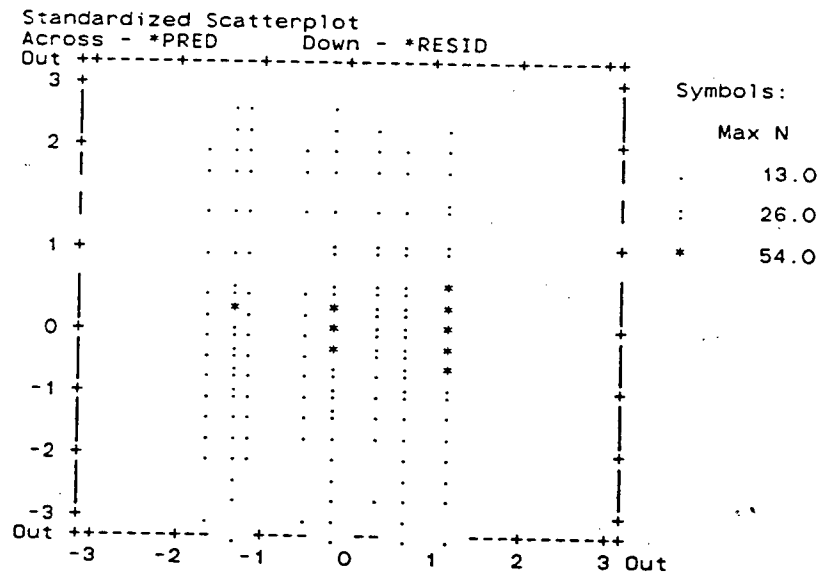


Figure 8.3c: GRAIN - Standardized Plots of Residuals Versus Standardized Predicted Values of FEV₁



Normality

For the purpose of inference, it is assumed that at each fixed X the sub-population of dependent values Y, follows a normal distribution. Normality of the data can be visually assessed through a plot of the empirical cumulative distribution which is a probability plot of the residuals ranked in order, plotted against the approximate percentage points. These are calculated as $100(\text{rank}-0.5)/N$ where N is the number of residuals which are arranged in order of increasing magnitude (Hill 1974). Depicting the residuals in this manner is most informative when the number of residuals exceeds 20. In a publication by Strobe and Helms (1984) normal probability plots were examined and found to be linear, indicating normality of the residuals in this longitudinal analysis of children's pulmonary function data.

Normal probability plots of the standardized residuals based on a linear regression of FEV_1 on time exhibit no substantial departures from a straight line, thus near normality of these residuals is indicated for all three data sets (Figures 8.4a-c).

As an adjunct to the residual analysis, histograms of all the FEV_1 values and the associated descriptive statistics are shown in Figures 8.5a-c, where a normal curve is superimposed. The coefficients of skewness were all low and nonsignificant for the *COAL* and *VETERANS* distributions of FEV_1 values. A significantly negative kurtosis, which indicates a flattening of the normal curve was found, however. For the *GRAIN* workers, there was significantly negative skewness (the higher values were bunched closer to the mean) and significantly positive kurtosis as a result of the peaked form of the distribution (Snedecor and Cochran, 1972).

Figure 8.4a: COAL - Normal Probability Plots of the Standardized Residuals based on a Linear Regression of FEV_1 on Years of Follow-up

Normal Probability (P-P) Plot
Standardized Residual

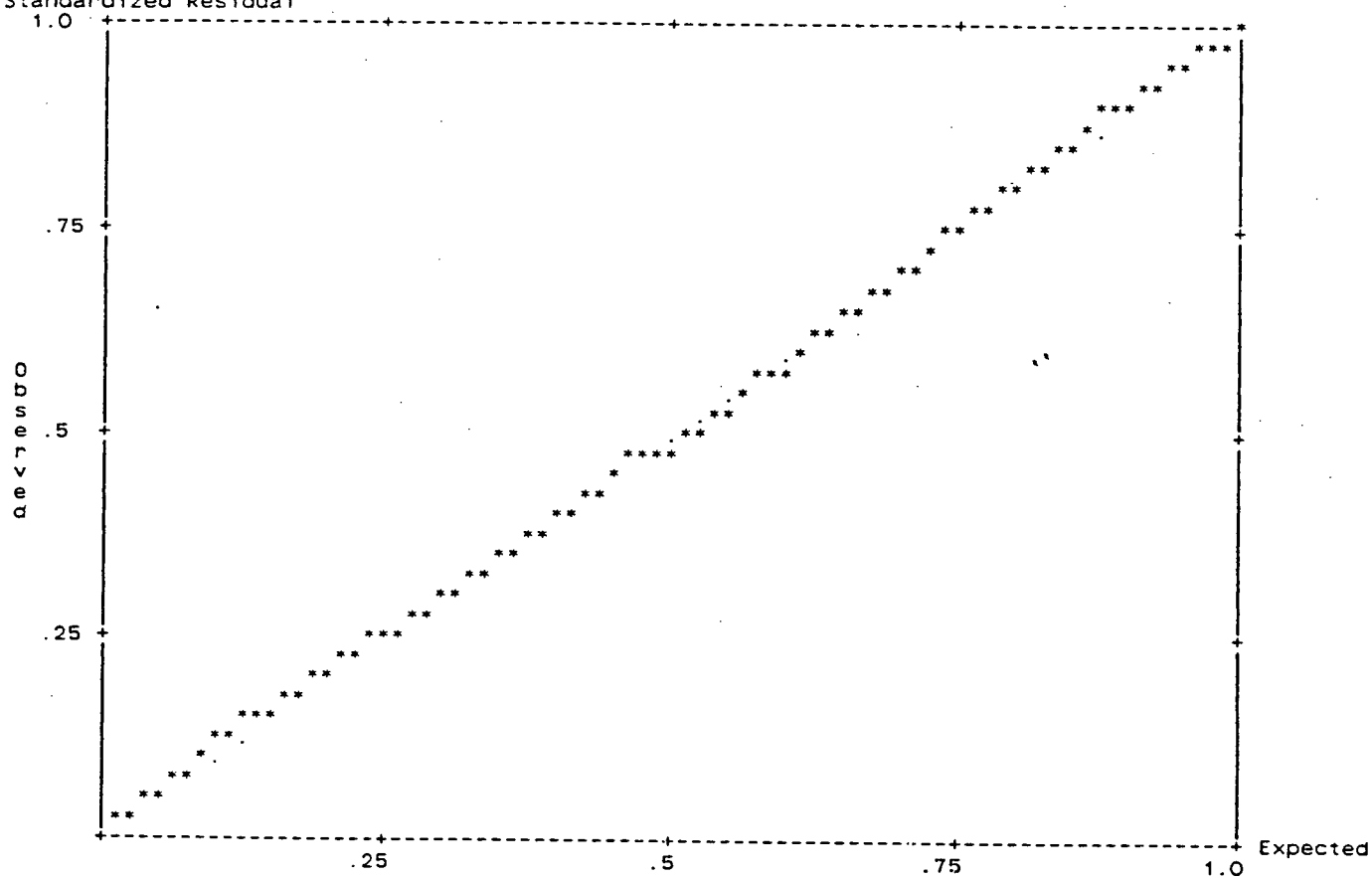


Figure 8.4b: VETERANS - Normal Probability Plots of the Standardized Residuals based on a Linear Regression of FEV_1 on Years of Follow-up

Normal Probability (P-P) Plot
Standardized Residual

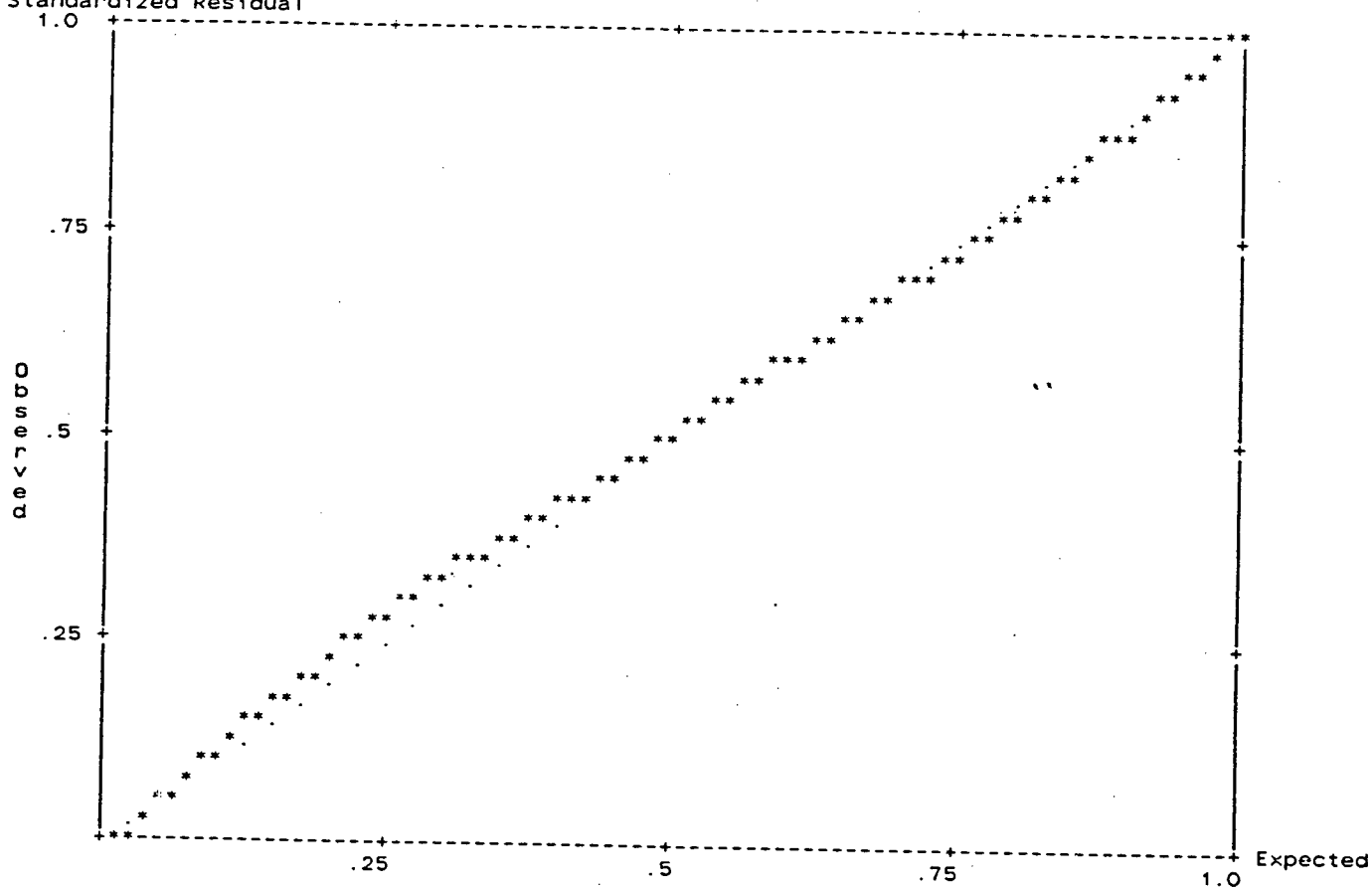


Figure 8.4c: GRAIN - Normal Probability Plots of the Standardized Residuals based on a Linear Regression of FEV_1 on Years of Follow-up

Normal Probability (P-P) Plot
Standardized Residual

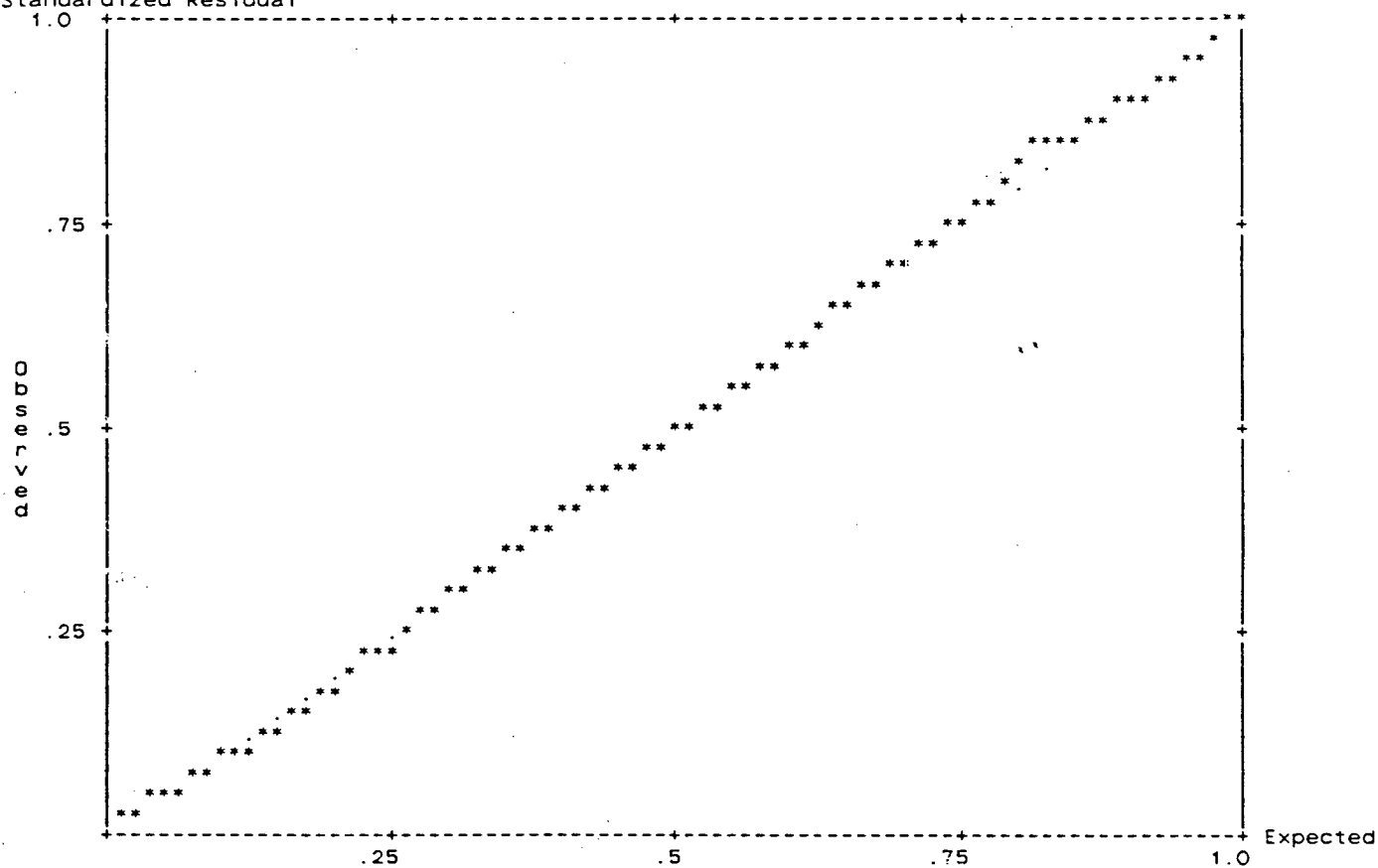


Figure 8.5a: COAL - Histogram of FEV₁

FEV

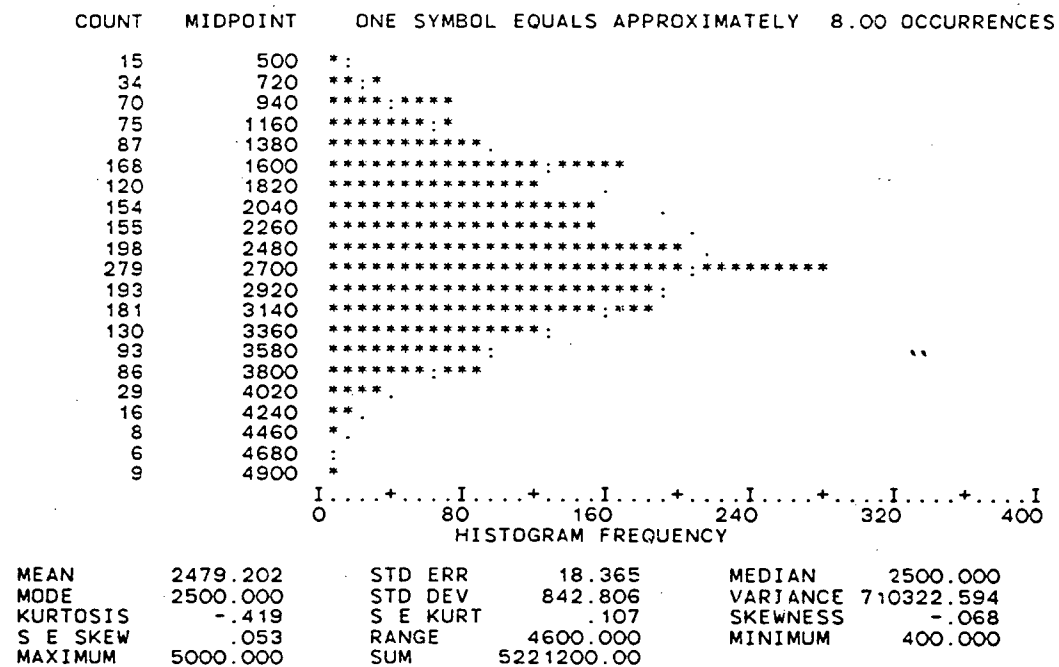


Figure 8.5b: VETERANS - Histogram of FEV₁

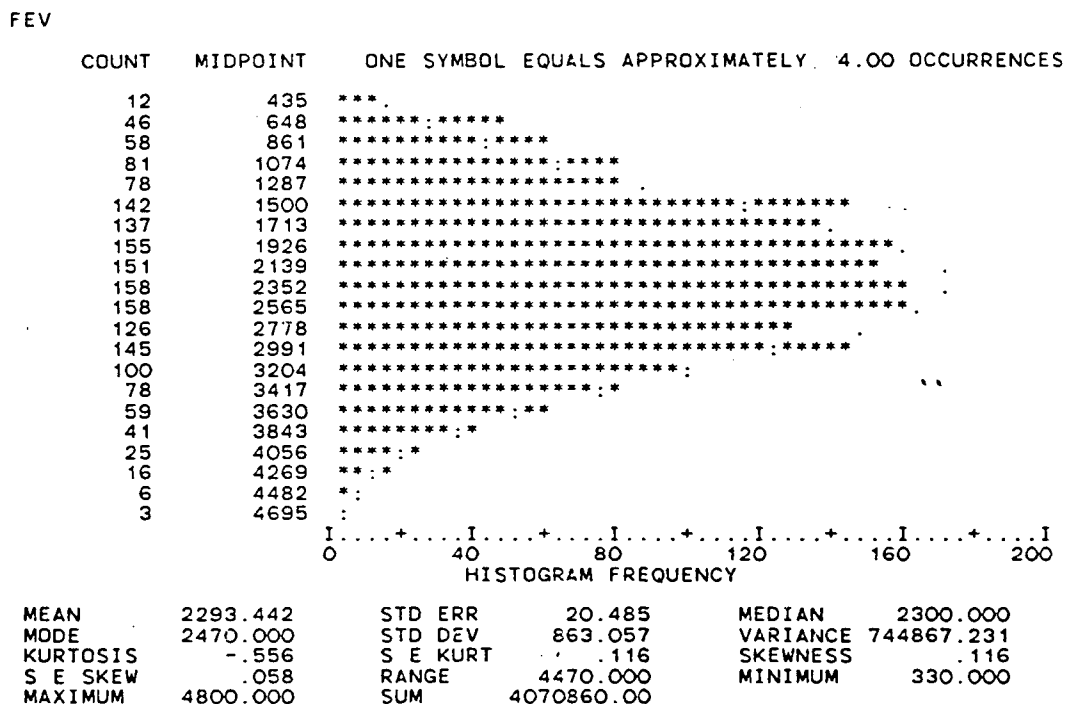
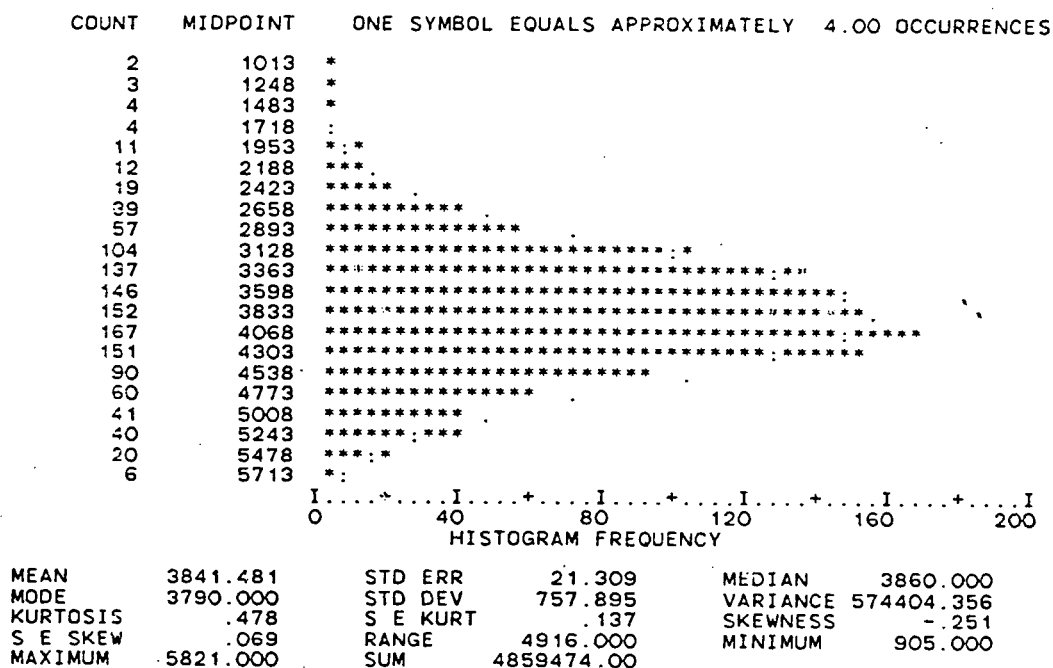


Figure 8.5c: GRAIN - Histogram of FEV₁

FEV



Stebbins (1971) reviewed the pulmonary function literature, and noted a number of studies which evaluated the normality of pulmonary function residuals and found repeated examples where FEV_1 and FVC were normally distributed. For example, in a study by Ashford and Brown (1965) the coefficients of skewness and kurtosis were found not to vary systematically by age and were evenly distributed around zero. Contrarily, in an analysis of pulmonary function data of fire fighters, the regression residuals were found to be skewed and the data was therefore transformed according to a Box-Cox transformation procedure (Kusiak and Roos 1984). In an analysis of the rate of FEV_1 decline, Eisen et al. (1984) found that their data was not normally distributed and was both skewed and had a higher proportion of observations in the tails. An important aspect of such analysis is that regression is generally robust against moderate departures from the normality assumption, and inferences will not be seriously affected (Berensen et al. 1983).

The residual analysis does indicate slight skewness for the *GRAIN* data; and kurtosis in all of the distributions; however the values of the parameters, being close to zero, do not show significant departures from normality.

Homoscedasticity

The application of ordinary least squares (OLS) assumes that the variation or scatter about the line of regression be constant for all values of X, the independent variable. Thus, the dependent values will vary the same amount when X is fixed at a low value as when it is fixed at a high value, so that the spread about the regression line is uniform (Berensen et al. 1983). A scatterplot of the residuals against the explanatory variable forms one basis for the evaluation of heteroscedasticity. If the residual fan out as the X value increases, lack of homogeneity in the variance of Y. Ignoring heterogeneity of variance will not result in bias in the estimated regression

coefficients; however, it can seriously affect the precision of such estimates, as measured through their variances (Berensen et al. 1983). The standardized plots of residuals of FEV_1 by year of follow-up (Figures 8.6a-c), show relative constancy in their scatter over the entire period of follow-up. There is therefore no definite pattern of increasing spread with increasing years of follow-up in any of the 3 data sets. However, there is some suggestion, particularly in the *COAL* data set that there is a tendency for less scatter in the residuals with a lengthy period of follow-up.

A possible method to take heterogeneity of variance into account is to apply weighted least squares. A study by Strobe and Helms (1984) provides an example of this method. Children's FVC residuals were found to increase in value with increasing height (a surrogate of time) To account for this, separate weighted straight line regressions of FVC versus height were fit to each subjects' data using the inverse of height as the weighting factor. The reverse situation may be evident, at least in the *COAL* and *VETERANS* data where a reduction in variability of FEV_1 appears to occur with time. Weighted least squares is evaluated further in Chapter 9.

Figure 8.6a: COAL - Standardized Plots of Residuals of FEV₁ by Standardized Years of Follow-up

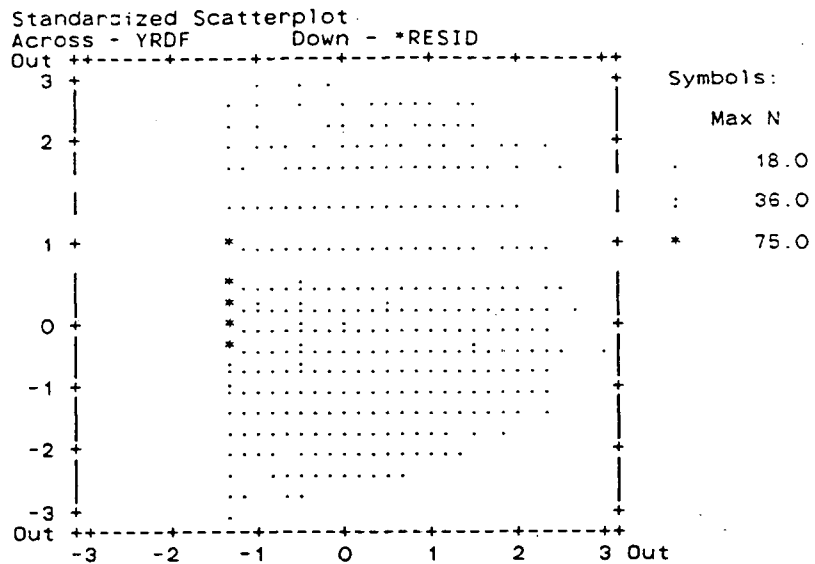


Figure 8.6b: VETERANS - Standardized Plots of Residuals of FEV₁ by Standardized Years of Follow-up

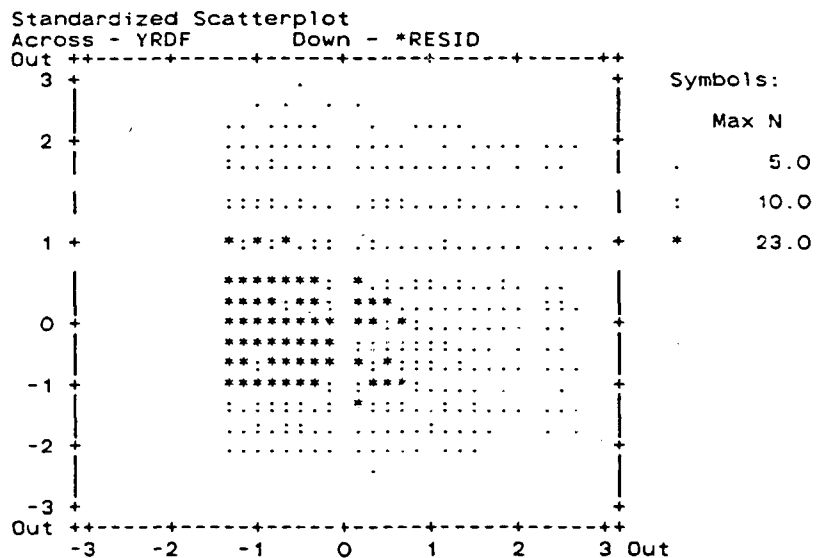
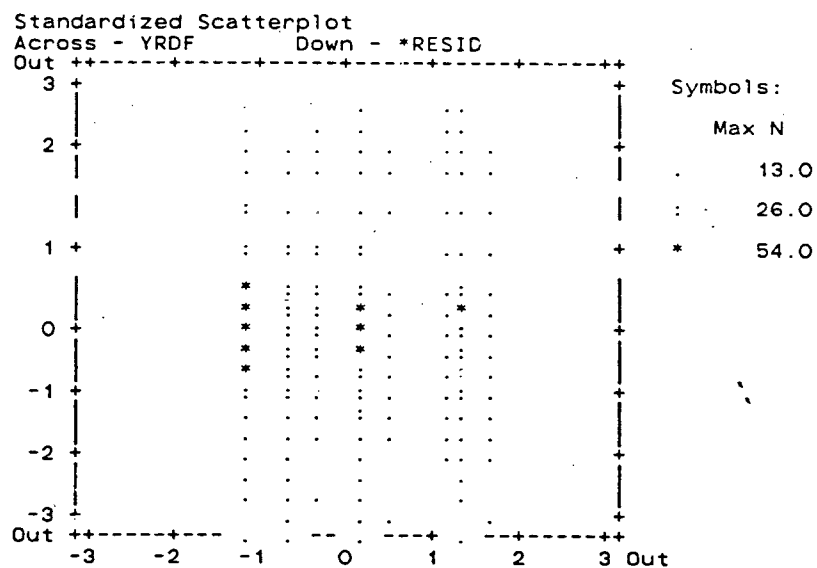


Figure 8.6c: GRAIN - Standardized Plots of Residuals of FEV₁ by Standardized Years of Follow-up



Autocorrelation

A basic assumption of the regression model is the independence of the residual component such that the errors associated with the i^{th} and j^{th} observations are assumed to be uncorrelated. Autocorrelation typically occurs when observations have a natural sequential order, such as found for repeated measurements on an individual. Correlation is a measure of a straight line relationship between two variables as indicated by Pearson's product moment correlation coefficient "R". For the yearly interval *VETERANS* data set the correlations between pairs of adjacent measures were significant ($p < .001$) ranging from 0.92 to 0.98. A slightly larger range in correlations of 0.89 to 0.98 was shown for pairs of FEV_1 values every 3 years apart.

In order to detect autocorrelated errors by statistical methods, the Durbin-Watson procedure has been developed. This makes the assumption that the residuals follow a first order autoregressive model. The process, observed at equally spaced time periods is, $e_t = \rho e_{t-1} + a_t$, where e_t is the error term in the model at time period t , a_t is a normal independently distributed $(0, \sigma_a^2)$ random variable and ρ is the autocorrelation parameter.

The small values of the Durbin-Watson statistic indicated the presence of positive autocorrelation in the data sets according to these overall regression analyses, as each was less than the lower Durbin-Watson statistic of 1.52. The large numbers of pairs of (X,Y) values permitted through this analysis add stability to the Durbin-Watson statistic. With individual data, those having the minimum of 15 measurements needed to use the tabulated test statistic, resulted in 4 of 34 Winnipeg subjects and 5 of 13 Halifax subjects showing significant positive autocorrelation in their FEV_1 measurements. Other

data sets had an insufficient number of measurements to apply this statistic (Montgomery and Peck 1982).

Missing Data

The missing data for each of the three thesis data sets can all be characterised to some extent as having arisen from three attributes of study design:

- 1) The timing of each measurement occasion was such that the intervals of time between measurements were not necessarily equivalent for each person.
- 2) Within the set times of the study, absences from the testing resulted in sporadic missing data.
- 3) Subjects were lost to follow-up before the completion of the designated study time period.

The patterns of missing data, ie. the location and extent of the missing values were analyzed for each data set group using BMDPAM (1981). Percent of missing FEV₁ values for the *VETERANS* and *GRAIN* data sets are shown in Table 8.6. Because the *COAL* data were so imbalanced, empty cells dominated the matrix creating a situation that the programme could not handle. The extent of such an imbalanced structure can be appreciated by examining Table 8.6, where large percentages of missing data dominate the *VETERANS* data particularly at the beginning and end of the follow-up period. This pattern of absence resulted from differing starting and stopping points between the different cities, as well as between individuals. The relatively balanced structure and arbitrariness of the *GRAIN* data is evidenced by the small percentages of missing FEV₁ values.

The identification of personal characteristics that relate to the missing data process, points to an ignorability assumption. (See Chapter 9 for an evaluation of missing data in the *VETERANS* study). The assumption of ignorability does protect against some

kinds of non-ignorable response, and allows most analytical techniques to remain unbiased (Waterneaux 1989).

Table 8.6: Percentages of missing FEV₁ data by Year of Follow-up.

VETERANS					
1959	1960	1961	1962	1963	1964
83.0	76.1	48.9	3.4	2.3	4.5
1965	1966	1967	1968	1969	1970
2.3	3.4	0.0	34.1	4.5	2.3
1971	1972	1973	1974	1975	1976
5.7	4.5	12.5	45.5	56.8	48.9
1977	1978	1979	1980	1981	1982
54.5	52.3	68.2	65.9	69.3	69.3
1983					
98.9					
GRAIN					
1976	1979	1982	1985		
0.0	6.7	6.4	0.0		

Other Data Characteristics

Data which exhibits a regression to the mean (RTM) phenomenon would exhibit a tendency for higher initial values to show steeper declines. For the horse-racing (HR) effect to be present, those whose lung function values were lower, on average, would show greater declines. Whether the raw data displays any tendency for a HR effect as opposed to a RTM phenomenon, can be evaluated by correlating the initial FEV₁ value and the mean FEV₁ values with their slope. According to Burrows et al. (1987) a significant negative correlation between the slopes of FEV₁ with the initial values indicates a RTM effect, as opposed to a positive correlation found between the means of the values with their slope, which is an indication of the horse racing effect. Table 8.7 shows the correlation coefficients for the association between initial FEV₁ and mean FEV₁ with slope for each data set and the groups within them. The slope for each individual was obtained by simple linear regression of FEV₁ with time since first measurement.

Table 8.7: Correlation Coefficients between Initial or Mean FEV₁ and Slope.

	Initial FEV ₁	Mean FEV ₁
<u>COAL</u>		
ALL	-0.14 *	+0.21 *
Deceased Nonsmokers	-0.34 *	+0.07
Living Nonsmokers	-0.18	+0.05
Deceased Smokers	-0.19 *	+0.17 *
Living Smokers	-0.10	+0.26 *
<u>VETERANS</u>		
ALL	-0.18 *	+0.02
Toronto	-0.30 *	-0.11
Winnipeg	-0.27 *	-0.03
Halifax	-0.05	+0.21
<u>GRAIN</u>		
ALL	+0.07	+0.28 *
Grain workers	+0.04	+0.26 *
Civic workers	+0.30 *	+0.44 *
* p < 0.05 (one-tailed)		

A criticism of this method of evaluating the RTM is that the initial value is invariably related to the slope as it helps determine the slope. The *VETERANS* data provides the most useful test of this effect in that the value for the first measurement occasion was actually the average value of approximately 12 monthly measurements taken on each subject. It is most interesting to note that it is in this data set, like the others, that the "regression to the mean" effect is apparent in at least one of the study groups as evidenced by the significant negative correlation coefficients between initial FEV₁ and individual slopes. In this case, this behavior cannot be attributed to this statistical phenomenon. A significantly positive correlation between mean FEV₁ and slope may indicate a horse-racing effect. This effect was found only in the smoking groups of the *COAL* data; whereas this association was found in the grain and civic groups regardless of smoking status.

For the *VETERANS* data, a cross-sectional regression of age and height on the baseline FEV₁ values resulted in an age coefficient of -39.4 ml/yr for Toronto; -43.9 ml/yr for Winnipeg; -41.2 ml/yr for Montreal; and a relatively low value of -20.7 ml/yr for Halifax. For this data set, height was not a significant predictor of initial FEV₁ level. A cross-sectional regression analysis performed on the *GRAIN* data revealed an age related difference of -32.5 ml/yr for grain workers and -29.1 ml/yr for the civic workers. A cross-sectional analysis of baseline data could not be conducted on the *COAL* data, as the date of the first measurement occasion differed for most subjects.

Conclusion

The objectives of exploratory data analysis include:

- 1) Achieving more specific description of what is suspected;
- 2) Employing the data to assess the adequacy of a contemplated model, and
- 3) To find anticipated aspects in the data, and to suggest new models for data summarization and analysis (Tukey and Wilk 1965).

The exploratory data analysis conducted on the thesis data sets achieved all three objectives. Initial transcription errors in one of the data sets were detected, and another set of data was excluded from analysis because initial criteria of validity were not met. The suitability of the ordinary least squares model has been assessed by evaluating its assumptions. Although the normality, homogeneity of variance and linearity of the data has been generally satisfied, based upon an overall regression analysis of all data points, the independence assumption is questionable. The adjacent serial measurements were highly correlated with each other and, based on the "cross-sectional" residual analysis there was evidence of autocorrelation.

An important consideration of the exploratory data analysis is that the findings should be regarded as a guide and not absolute determinants. As Francis Bacon was quoted as saying:

"truth arises more easily from error than from confusion". (Tukey and Wilk 1965, p386).

The systematic application of residual "or error" analysis has proven to offer insight into the statistical properties of the data, which may have relevance when evaluating different methods of analysis.

CHAPTER IX

RESULTS

Comparative Analysis

The initial comparison will be based upon the simpler methods of analysis which use ordinary least squares principles; these methods are commonly used in published reports for longitudinal data and are easily adapted from available statistical packages. Except for Method D, which has as a dependent variable the percent change in FEV₁ per year, each of the methods describes change in FEV₁ over time in terms of ml/yr. Each method however, describes a distinct dependent variable. The methods used are as follows:

Method A - For this "cross-sectional" approach of longitudinal data, the decline estimate is obtained from the regression coefficient describing the relationship of years since first follow-up on all the FEV₁ measurements. Added covariates to this model alter the time coefficient through their relationship with the FEV₁ measurements.

Method B -The regression coefficient of interest describes the relationship of the time interval between an individual's pair of measurements with the associated differences in the pairs of FEV₁ measurements. No intercept is included in this model of unadjusted estimates of FEV₁ decline.

Method C -For each individual, the difference between the first and last measurement occasion divided by the number of years of follow-up yields the dependent variable.

Method D -Percentage change per year is the time variable of interest, calculated as the difference in FEV₁ between the first and last measurements divided by the initial FEV₁ level and the total number of years. The magnitude of this estimate can give an indication of the severity of the FEV₁ decline with time and emphasises the extent of FEV₁ change for those with lower initial FEV₁ measurements.

Method E -As in Method A, this was the regression coefficient of the relationship of difference in years since first follow-up with FEV₁ level. In this case however, regression coefficients are derived for each individual and for group comparisons it is the averages of these regression coefficients that are being compared.

Methods A and B have in common the "cross-sectional" nature of their derivation, where each value, or pairs of values are treated as if they were individual units. Method D, like Method C, is dependent upon an estimate of FEV₁ change based upon the initial and final measurements only. Unlike Method A, in Method E the measurements belonging to an individual are given equal weighting with those of the other individuals in the group. A slope based on 6 measurements is therefore given equal weighting to one based on 12 measurements. In the "cross-sectional" approaches, on the other hand, each measurement or pair of measurements is given equal weighting without accounting for which measurements belong to an individual unit.

Unadjusted Decline Estimates

An initial evaluation of the FEV₁ decline estimates is based on average unadjusted levels within each data set grouping, as shown in Table 9.1. The raw decline estimates calculated from end points (Method C) or by regression of individual values (Method E) gave consistent results to one another, differing by only a maximum of 2.4 ml/yr for the average decline estimate among Toronto subjects. In the *COAL* workers, these two methods produced consistently higher values. Method A had a relatively larger standard error. The importance of the size of the standard error is that it defines the confidence limits of the mean, yielding an estimate of precision. However the regression-based standard errors for methods A and B are not directly comparable with the others. Methods A on the civic workers data yielded standard errors that were so high that the coefficient of FEV₁ decline was not significantly different from zero.

The small standard errors for Method D are derived from the standard deviation of the dependent variable that is expressed as percent rather than ml/yr. Generally, the ratio of the standard error to its mean was relatively small, as was found for all methods (apart from Method A); this emphasizes the greater precision of these methods.

The estimated coefficients of decline in the grain workers are generally much lower than that of the other two data sets (which would be expected from the younger cohort followed-up for a relatively shorter time). The values for the *COAL* and *VETERANS* were quite similar between all the OLS methods. The highest decline estimate overall was a decrement of 71.6 ml/yr in FEV₁, which was found for dead coal miners who smoked using Method C. The same method produced the lowest estimate of a decrement of only 16.7 ml/yr for the civic workers. The percentage decline ranged from -0.46% for grain workers to a high of 2.9% decrement per year, which was once again found for the Halifax workers.

Table 9.1: A Comparison of Mean Unadjusted FEV₁ Change Estimates

	METHOD	A	B	C	D	E
<i>COAL</i>						
1 Dead	X	-59.3	-62.9	-63.1	-2.3	-63.9
Nonsmokers	(SE)	(4.1)	(2.8)	(3.2)	(.12)	(3.1)
n=100						
2 Alive	X	-42.2	-41.6	-47.8	-1.6	-45.7
Nonsmokers	(SE)	(6.4)	(3.9)	(3.5)	(.11)	(4.0)
n=35						
3 Dead	X	-65.0	-64.5	-71.6	-2.5	-71.2
Smokers	(SE)	(4.7)	(3.0)	(3.6)	(.13)	(3.7)
n=104						
4 Alive	X	-52.7	-54.4	-57.6	-1.9	-57.7
Smokers	(SE)	(3.5)	(2.1)	(2.6)	(0.09)	(2.8)
n=145						
<i>Significant</i>		1,3,4	2	3	2,4	3
<i>Differences</i>		vs	vs	vs	vs	vs
		2	1,3	2,4	1,3	2,4
<i>VETERANS</i>						
1 Toronto	X	-65.7	-65.4	-63.4	-2.6	-65.8
n=51	(SE)	(8.5)	(12.7)	(6.1)	(0.25)	(5.9)
2 Winnipeg	X	-37.2	-52.5	-51.2	-1.8	-50.2
n=56	(SE)	(4.5)	(7.6)	(4.9)	(.23)	(5.9)
3 Halifax	X	-67.0	-62.1	-59.3	-2.9	-57.3
n=34	(SE)	(7.2)	(13.3)	(7.5)	(0.43)	(7.6)
<i>Significant</i>		1,3	ns	ns	2	ns
<i>Differences</i>		vs			vs	
		2			3	
<i>GRAIN</i>						
1 Grain	X	-19.2	-18.3	-16.7	-0.46	-19.9
n=269	(SE)	(6.7)	(2.8)	(2.3)	(0.06)	(2.1)
2 Civic	X	{-20.2}	-17.6	-18.0	-0.50	-19.3
n=58	(SE)	(15.8)	(5.6)	(3.8)	(0.10)	(3.8)
<i>Significant</i>		ns	ns	ns	ns	ns
<i>Differences</i>						

{ } = Non-significant FEV₁ decline estimate

() = Standard error of the mean

The most consistent estimates of FEV_1 decline in ml/yr found across the four OLS methods was observed for the Toronto data, which ranged from 63.4 to 65.8 ml/yr. The *VETERANS* data also had the least consistent estimates where for the Winnipeg group, the estimates of FEV_1 decline ranged from 37.2 to 52.5 ml/yr. For the *COAL* workers, the order of the group estimates within each method were consistent with one another. The alive non-smokers had the smallest declines, followed by alive smokers, deceased non-smokers and then deceased smokers. The most statistically significant differences between the *COAL* groups were found for Method D, where the two living groups had smaller percentage declines than those found for the dead smoking groups. Under Method A, all groups differed significantly from the lowest decline of the alive non-smokers. Methods C and E each showed the same pattern where the highest declines found for deceased smokers differed from the alive groups. Therefore, there was agreement between the methods in that vital status was emphasized; the alive smokers showed less of a decline than the two deceased categories.

For the *VETERANS* data the methods uniformly showed the lowest FEV_1 decline occurring among Winnipeg subjects. The magnitude of this estimate however, varied from 37.2 ml/yr for Method A to 52.5 ml/yr for Method B. Without evaluating statistical significance, the order of the remaining two cities was mixed, in that methods A and D ranked Halifax, as opposed to Toronto, as having the greatest decrement in FEV_1 . It was also only for methods that some of the comparisons were statistically significant. Under Method D, the percent decline for Halifax was significantly higher than that found for Winnipeg. Using Method A, both Halifax and Toronto had higher declines in FEV_1 than observed for Winnipeg.

For the *GRAIN* data, All the methods were consistent in showing no significant differences in the decline estimates between the grain and civic workers and the

estimates were close to one another in magnitude. For Method A the standard errors were so high that the estimates for the civic workers were not even significantly greater than zero.

In summary, with the exception of Method A, the estimates of decline derived from each of the methods are generally consistent with one another in magnitude and in the ordering of the groups within each data set. For the *COAL* miners, the group mean estimates are so distinct that statistically significant differences were found among the groups with almost all the different methods applied. On the other hand no statistically significant differences could be seen between the estimates of decline of the *GRAIN* versus the civic workers in any of the methods.

Adjusted Decline Estimates

Initial age and height were common fixed covariates obtained for each of the data sets analyzed. Where the estimates of decline were the dependent variable, as in methods C through E, analysis of covariance was used to adjust for the differing distributions of age and height in each of the groups compared. In the regression analysis of method A, age and height were forced into the model as additional independent variables. These techniques were extended to the fully adjusted models, with the full model covariates used, listed below each data set, as shown in Table 9.2.

Table 9.2: A Comparison of Mean Age-Height and Fully Adjusted FEV₁ Decline

Estimates

	METHOD	A	C	D	E
<i>COAL</i>					
1 Dead	Age-Ht	-66.0	-65.1	-2.31	-65.2
Nonsmokers	Full	-66.9	-67.1	-2.44	-67.1
2 Alive	Age-Ht	-38.2	-48.0	-1.57	-45.9
Nonsmokers	Full	-38.4	-48.0	-1.63	-45.8
3 Dead	Age-Ht	-64.3	-71.9	-2.52	-71.5
Smokers	Full	-66.2	-72.3	-2.52	-71.6
4 Alive	Age-Ht	-55.1	-56.3	-1.86	-56.5
Smokers	Full	-56.2	-57.1	-1.90	-57.0

Full Model Covariates = AGE, HEIGHT, RETYRS, FUP.

VETERANS

1 Toronto	Age-Ht	-65.5	-62.2	-2.58	-64.9
	Full	-70.3	-66.0	-2.78	-68.9
2 Winnipeg	Age-Ht	-41.7	-51.9	-1.85	-50.7
	Full	-51.0	-46.6	-1.53	-45.0
3 Halifax	Age-Ht	-70.5	-60.0	-2.87	-57.9
	Full	-74.2	-61.3	-2.86	-58.2

Full Model Covariates = AGE, HEIGHT, AGESMK, SMKLVL, FEVCV, FUP.

GRAIN

1 Grain	Age-Ht	-20.2	-17.6	-0.48	-20.8
	Full	-22.0	-16.0	-0.45	-19.1
2 Civic	Age-Ht	-20.8	-13.8	-.37	-14.9
	Full	-21.8	-20.2	-0.53	-22.2

Full Model Covariates = AGE, HEIGHT, CIGLVL, FUP.

RETYRS = years from retirement to end of follow-up

FUP = total number of years of follow-up

AGESMK = Age when the subject started smoking

SMKLVL = (ordinal) level of smoking

FEVCV = coefficient of variation of the first 3 measures

CIGLVL = (ordinal) level of cigarette smoking

For *COAL* miners the fully adjusted model contained the additional covariates of years from retirement to the end of follow-up (RETYRS) and also the total length of follow-up (FUP). For *VETERANS* the fully adjusted model included age at which smoking started (AGESMK), the level of smoking (SMKLVL, grouped according to average number of cigarettes smoked per day and whether they smoked pipes or cigars), the total length of follow-up, and the coefficient of variation for the first three measurements of FEV₁ (FEVCV). In addition to age and height, the *GRAIN* full model had an indication of the level of cigarette smoking (CIGLVL) and total length of follow-up. Forced entry of all the covariates in each model was used for these comparisons.

For the *COAL* workers, differences in the decline estimates between the age/height adjusted model and the fully adjusted model were slight; thus the effects of adding RETYRS and FUP to the model had a negligible effect on the decline estimate calculated. Larger differences are observed between the two models for the *VETERANS* data; this could be expected because there are a larger number of covariates added to the model. For models C to E the covariate adjustment effects were consistent; the full models for Toronto all showed a slight increase in the decline estimate, for Winnipeg, the opposite effect was shown, and for Halifax, very little change was evident. All full covariate models applied to the *VETERANS* data resulted in a larger estimate of decline with model A.

It is for the *GRAIN* workers data that the most dramatic differences are seen between the coefficients found for the age-height adjusted models versus the full model. After full adjustment under models C and D, the order of the two groups was reversed. Again, however there were no statistically significant differences between these groups, whether looking at unadjusted decline estimates or the two adjusted models.

A comparison between the unadjusted decline estimates in Table 9.1 to the adjusted models in Table 9.2, shows little difference in the coefficients of decline. Among *COAL* workers, it was the deceased nonsmokers group that showed the greatest extent of change after adjustment. That resulted in a slight increase in FEV₁ decline for all the methods. This same pattern occurred for the Winnipeg group among the *VETERANS* data.

Table 9.3 shows the yearly change estimates for the complete data sets with the age and height coefficients that were obtained along with the coefficient of determination (R^2) for that model. An obvious distinction between Method A and the others is the relative magnitude of the age and height coefficients. This is because the dependent variable used in the regression equations is FEV₁ level. Age and height have been repeatedly shown in various epidemiological studies to be strongly related to FEV₁ level, and are predictors that are included in equations used to determine normal pulmonary function level. Very few of the age-height coefficients for the other methods were significantly greater than zero. The relatively high R^2 values for model A indicate that a linear cross-sectional relationship of years elapsed, and age and height on all FEV₁ measurements, account for a good proportion of the variance of the measurements.

Table 9.3: Comparison of Age-Height Adjusted Prediction Equations

METHOD	DEPENDENT VARIABLE	DATA SET	YEARLY CHANGE	AGE	HEIGHT	R ²
A	Level	COAL	-57.8	-34.5	+32.6	0.40
		VETERANS	-35.0	-27.0	+36.2	0.22
		GRAIN	-20.9	-35.4	+43.3	0.48
C	Endpoint Slope	COAL	-62.1	(.4)	-0.6	0.08
		VETERANS	-57.6	(.5)	(1.5)	0.03
		GRAIN	-17.0	-1.1	(0.01)	0.10
D	% Endpoint Slope	COAL	-2.13	(-0.01)	(0.01)	0.00
		VETERANS	-2.36	(0.00)	0.06	0.09
		GRAIN	-0.46	-0.03	(0.01)	0.14
E	Individual Slope	COAL	-61.9	(0.3)	-0.7	0.08
		VETERANS	-57.6	(0.4)	(0.5)	0.03
		GRAIN	-19.8	-1.1	(-0.07)	0.11
() = Non-significant coefficients						

The decline estimates presented in Table 9.2 for the full models resulted from forced entry of all the predictors listed in a regression or analysis of covariance relationship. Table 9.4 lists the covariates that were found to be significant predictors of the dependent variable in each data set. The *COAL* data was unique in that all independent variables used in the model were found to be significant for Method A only. This method had the distinction in all the data sets of demonstrating the most number of significant predictors and FEV₁ level, not change was the dependent variable. It is the group and/or age variables that were usually found to be significant predictors in all the data sets for all the methods. For the *GRAIN* data, the group variable was not significant in any of the methods.

Table 9.4: Comparison of Significant* Predictors

METHOD		SIGNIFICANT COVARIATES				
<i>COAL</i>						
A	GROUP	AGE	HEIGHT	RETYRS	FUP	..
C	GROUP	AGE
D	GROUP	FUP	..
E	GROUP	AGE	FUP	..
No nonsignificant variables						
<i>VETERANS</i>						
A	GROUP	AGE	HEIGHT	SMKLVL	..	FEVCV
C	GROUP	SMKLVL	AGESMK	..
D	GROUP	SMKLVL	AGESMK	..
E	GROUP	SMKLVL	..	FEVCV
FUP was a nonsignificant variable						
<i>GRAIN</i>						
A	AGE	HEIGHT	CIGLVL			
C	AGE			
D	AGE			
E	AGE			
GROUP, FUP were nonsignificant variables						

* $p < 0.05$ according to 2-tailed t distribution

AGE = age at first measurement

RETYRS = years from retirement to end of follow-up

FUP = total number of years of follow-up

AGESMK = age when the subject started smoking

SMKLVL = (ordinal) level of smoking

FEVCV = coefficient of variation of the first 3 measures

CIGLVL = (ordinal) level of cigarette smoking

The coefficients shown in Table 9.3 were unstandardized. For the regression equation used in Method A, which had the most number of significant predictors, the contribution of each independent variable in the regression relationship can be better assessed by looking at standardized beta coefficients. Standardization can be achieved by multiplying the coefficient by the standard deviation of the independent variable over the standard deviation of the dependent variable. Each regression equation was applied separately to each group in the data sets. The FEV_1 measurements, not their decline, was the dependent variable. The results are presented in Table 9.5. Note that the values of the coefficients for CIGLVL and SMKLVL are difficult to interpret as the variable was categorical, but treated as continuous for this analysis. Among the *COAL* workers (except for the case of the live nonsmokers where age has a greater effect) the strongest predictor was years of follow-up (YRDIF). The largest standardized betas were generally found among the alive nonsmokers group. Among the *VETERANS*, initial age as well as FEVCV were among the strongest predictors of FEV_1 level. With a greater FEVCV value, (that is higher variability of the initial values of FEV_1), there tended to be lower FEV_1 levels recorded among the veterans. For the *GRAIN* data set, age and height were dominant predictors of FEV_1 level, with year of follow-up having a comparatively small effect. In general, the signs of the coefficients conformed to prior expectations of the relationships. For example, for the data, the number of years of follow-up, initial age, and the level of cigarette smoking, were all negatively associated with the FEV_1 measurements, while height showed a positive relationship.

How well these models actually fit the data can be judged in part by evaluating the coefficient of variation (R^2) and the standard error of the estimate (SEE). The R^2 and SEE values for the unadjusted model, the age-height adjusted model and the fully adjusted model are given for each data set in Table 9.6.

Table 9.5: Standardized Regression Coefficients by Group Using Method A Full model

GROUP	INDEPENDENT VARIABLES						
<i>COAL</i>	YRDIF	Age	Height	RETYRS	FUP		
Dead	-0.58	-0.22	+0.26	-0.19	(+.10)		
Nonsmokers							
Alive	-0.42	-0.65	+0.37	+0.34	-0.25		
Nonsmokers							
Dead Smokers	-0.54	-0.13	+0.16	-0.11	+0.12		
Alive Smokers	-0.49	-0.24	+0.31	-0.09	(+.08)		
<i>VETERANS</i>	YRDIF	Age	Height	SMKLVL	AGESMK	FEVCV	FUP
Toronto	-0.34	-0.40	+0.14	-0.25	-0.13	-0.37	(+.03)
Winnipeg	-0.38	-0.31	+0.15	(-.04)	(-.06)	-0.36	(-.06)
Halifax	-0.42	-0.20	+0.15	(-.08)	+0.15	-0.24	(+.02)
<i>GRAIN</i>	YRDIF	Age	Height	CIGLVL			
Grain	-0.10	-0.52	+0.37	-0.08			
Civic	-0.09	-0.40	+0.46	-0.16			
() = Nonsignificant coefficient							

It must be restated that the dependent variables evaluated in each method differ from one another. As a result, the standard errors of the estimate cannot be directly compared to one another. When comparing these terms between the unadjusted model and the fully adjusted model, it can be seen that they both give similar information as to whether an improvement in the goodness of fit occurred; an increase in the R^2 is reflected in a decrease in the SEE. The more dramatic change occurred in the *GRAIN* data under Method A where the R^2 from an unadjusted to an adjusted model increased from 0.01 to 0.49, with a decrease in the standard error of the estimate of 211 ml.

Given rounding errors, which inevitably occurred, the *GRAIN* and *COAL* data show very little difference in the sum of squares with each method, while for the overall cross-sectional method (Method A) on the *VETERANS* data, the sum of squares was smaller, indicating a better fit of the model. It is noticeable that the coefficients in each of the models were very similar to each other, apart from the *VETERANS* Method A procedure.

An evaluation was made of what factors affected the SEE derived from regression of each individual's data. For the *GRAIN* data, a significant negative association was found between length of follow-up and number of measurements with the SEE. The opposite finding of a significant positive relationship was observed for the *COAL* data, which was also supported by a similar trend found in the *VETERANS* data.

Table 9.6: Comparison of Goodness of Fit Criteria for each Model

METHOD	DATA SET	d.f.	R^2_1	R^2_2	R^2_3	SEE ₁	SEE ₂	SEE ₃
A	<i>COAL</i>	2105	0.30	0.46	0.47	705	620	618
	<i>VETERANS</i>	1774	0.21	0.35	0.45	767	696	645
	<i>GRAIN</i>	1264	0.01	0.49	0.50	754	543	539
C	<i>COAL</i>	384	0.05	0.08	0.09	32.5	32.1	32.1
	<i>VETERANS</i>	140	0.02	0.03	0.11	41.0	41.1	40.0
	<i>GRAIN</i>	326	0.00	0.10	0.11	35.7	33.9	33.9
D	<i>COAL</i>	384	0.08	0.08	0.10	1.14	1.14	1.14
	<i>VETERANS</i>	140	0.06	0.08	0.19	1.95	1.94	1.86
	<i>GRAIN</i>	326	0.00	0.14	0.14	0.98	0.91	0.91
E	<i>COAL</i>	384	0.05	0.08	0.09	33.5	33.0	32.9
	<i>VETERANS</i>	140	0.02	0.03	0.12	43.5	43.7	42.4
	<i>GRAIN</i>	326	0.00	0.11	0.12	34.1	32.3	32.3

1 = Unadjusted model

2 = Age-Height Adjusted Model

3 = Fully Adjusted Model

d.f. = degrees of freedom

R^2 = coefficient of determination

SEE = standard error of estimate

Additional Statistical Methods

Three types of transformations were applied to the simple regression of all FEV_1 values on time elapsed. The dependent value, FEV_1 , was transformed either as $\log_{10}FEV_1$, the inverse of FEV_1 , or the square root of FEV_1 . In each data set the results were uniform: the untransformed relationship of FEV_1 with year of follow-up produces the most normal residuals. This was evidenced by the histogram of the residuals, a normal probability plot, and slightly smaller range of standardized residuals. Therefore, no further analysis of transformed variables was conducted.

Apart from normality considerations, transformations are often used to linearize the regression model and to stabilize the variance of the dependent variable if the homoscedasticity assumption is violated. Apart from the residual plots used in Chapter 8 which did not indicate heteroscedasticity, a crude way of assessing the time related variability was to observe the standard deviation of the FEV_1 values at each level of time. For the *COAL* data, the most extreme differences were observed, in that the standard deviations of the group of values at one time of measurement differed by a factor of approximately two times.

To emphasize the differences in precision or length of follow-up between slopes, a weighted average of the unadjusted slopes, which were derived using Method E, was conducted using the various weighting criteria suggested by the residual analysis described in Chapter 8.

The weighted averages of unadjusted individual slopes is shown in Table 9.7. There is little difference between the unweighted estimates and those weighted by the standard error of the estimate, the follow-up time and the square of the follow-up time. Weighting by the precision estimate of the reciprocal of SEE, produced greater differences from the other weighting schemes. In the *COAL* data set, these decline

estimates for the smokers were each significantly higher than that of the nonsmokers, regardless of vital status. With the *VETERANS* there is little difference between any of the decline estimates using this weighting, although Winnipeg had a median decline in comparison to the others. Also, with this type of weighting, declines of the grain workers at 10.6 ml/yr were almost twice that of the civic workers at 5.4 ml/yr.

The results for the weighted least squares analysis using year since first measurement (YRDIF) and its reciprocals, $1/\text{YRDIF}$ and $1/\text{YRDIF}^2$ as weighting factors in a regression analysis of all data points are shown in Table 9.8. In the *COAL* data weighting by $1/\text{YRDIF}$ and $1/\text{YRDIF}^2$ resulted in coefficients which were similar in magnitude and order. They were also similar to the unweighted coefficients. Using year of follow-up as the weighting factor for this group achieved a similar result to when the unadjusted slope estimate was weighted by reciprocal of SEE; that is the declines of the smokers were emphasized more than that of the nonsmokers.

Among the *VETERANS* the reciprocal of time weighting emphasized the Toronto group more, so that the decline estimates were somewhat higher than those of Halifax. For the grain workers a weighted least squares analysis using $1/\text{YRDIF}^2$ produced estimates of -11.8 in comparison to -5.8 ml/yr for the civic workers. These estimates are comparable to that given when the unadjusted slope estimates were weighted by the reciprocal of SEE.

Table 9.7: Weighted Averages of Unadjusted Slopes

GROUP	PARAMETER	NONE	1/SEE	SEE	FUP	FUP ²
<i>COAL</i>						
1 Dead Nonsmokers	B	-63.9	-49.4	-68.8	-64.5	-65.0
	Int	2792	2765	2861	2840	2880
2 Alive Nonsmokers	B	-45.7	-44.5	-44.8	-49.3	-43.2
	Int	3126	3210	3087	3124	3114
3 Dead Smokers	B	-71.2	-70.0	-79.0	-70.1	-68.8
	Int	2943	2845	3071	2952	2957
4 Alive Smokers	B	-57.7	-57.7	-61.8	-56.8	-56.3
	Int	3234	3373	3239	3247	3248
<i>VETERANS</i>						
1 Toronto	B	-65.8	-46.9	-76.4	-66.7	-65.8
	Int	2571	2441	2758	2558	2545
2 Winnipeg	B	-50.2	-53.3	-40.8	-51.2	-52.0
	Int	2973	3030	2861	3060	3120
3 Halifax	B	-57.3	-55.5	-60.0	-58.6	-60.8
	Int	2229	2239	2271	2247	2255
<i>GRAIN</i>						
1 Grain	B	-19.9	-10.6	-18.3	-20.0	-20.1
	Int	3908	3814	3917	3910	3912
2 Civic	B	-19.3	-5.4	-20.3	-19.3	-19.3
	Int	4024	3732	4040	4024	4024

Table 9.8: Weighted Least Squares Estimates of Decline

	1/yr dif	1/yr dif ²	yr dif
<i>COAL</i>			
1 Dead Nonsmokers	-63.2	-65.7	-52.1
2 Alive Nonsmokers	-40.7	-42.9	-35.6
3 Dead Smokers	-68.0	-71.3	-57.0
4 Alive Smokers	-50.7	-51.5	-61.8
<i>VETERANS</i>			
1 Toronto	-63.0	-60.1	-68.1
2 Winnipeg	-36.9	-39.4	-40.4
3 Halifax	-57.9	-50.0	-75.3
<i>GRAIN</i>			
1 Grain Handlers	-16.1	-11.8	-20.8
2 Civic Workers	-11.9	-5.8	-38.3

Random Effects Model

The random effects method of analysis models FEV_1 as the dependent variable and produces an overall coefficient of change for the entire data set and for each group, through the use of dummy variables.

The random effects method, using BMDP5V, requires a Z matrix to be declared which has as its first column a series of ones and as a second column the time interval between successive measurements that is in common with all the subjects in the data set. In the *COAL* data set, measurements were made irregularly and there were a different number of measurements for each individual. In consequence, the specified Z matrix was highly imbalanced, having a contrast of 33 time elements separated by yearly intervals. The time element (X) had its first value at zero for the very first year of measurement that any individual had during the follow-up series, which was in 1950. With both BMDP software versions, however, this data set could not be analysed because there was insufficient space possible for the array designation. The matrix inversion was impossible under the constraints of the systems available.

In the *VETERANS* data set the time contrast array was also very large, having 25 separate values according to the calendar time of measurement. However, because of the prospective design of yearly follow-up, most individuals had the first set of measurements complete and missing values were only found at the end of the measurement series. However, even with this more complete model, the iterations exceeded the default maximum of 15, and the estimates failed to converge even after attempting 100 iterations and substituting starting values, based on previous results. Despite this, these decline estimates were made available for a few models that were attempted using the 1990 version of the BMDP program. It must be cautioned that these estimates are subject to error.

Table 9.9: Results of Random Effects (RE) and Unstructured (UNSTRUC) Covariance Models in the *GRAIN* data set.

METHOD	SLOPE Grain	ESTIMATES Civic	SIGNIFICANT VARIABLES
RE(1)	-11.2	-11.5	intercepts, time
UNSTRUC(1)	-12.5	-11.5	"
RE(2)	-11.2	-11.5	intercepts, time, age, height
UNSTRUC(2)	-18.9	-21.3	"
RE(3)	+5.4	+2.8	intercepts, time, ciglvl, height
UNSTRUC(3)	-20.0	-19.4	intercepts, time, age, height

(1) = unadjusted model
(2) = age, height adjusted model
(3) = full covariate model

Table 9.10: Variability Estimates using Random Effects (RE) and Unstructured (UNSTRUC) Covariance Models in the *GRAIN* data set.

METHOD	VARIANCE OF SLOPE	WITHIN-SUBJECT VARIANCE	AIC
RE(1)	591	37679	-9225
RE(2)	590	37685	-9101
RE(3)	112	20361	-8582
UNSTRUC(1)	—	—	-9214
UNSTRUC(2)	—	—	-8980
UNSTRUC(3)	—	—	-8914

AIC = Akaike's Information Criterion (Units = ml²)

No significant differences in slope were found between any of the *VETERANS* groups using unadjusted, age–height adjusted, and full covariate random effects models. The FEV_1 decline estimates were generally lower than those found for the other ordinary least squares models. The unadjusted decline estimates were 36.6 ml/yr for Winnipeg, 39.8 ml/yr for Toronto, and 52.3 ml/yr for Halifax. The age–height adjusted model produced a slightly higher decline of 42.2 ml/yr for Toronto. With the full model, the group factor, age, smoke level and the coefficient of variation of the first 3 measurements of FEV_1 (FEV₁CV) were significant, producing lower decline estimates that ranged from 23.0 to 40.3 ml/yr but retained the same order. In the full model both the log–likelihood and the within–subject variance of slope estimates were slightly higher, while the error mean square was smaller, as could be expected from a model with more parameters. A completely unstructured full model was used and there, a significant difference was found for the higher decline estimates of Toronto (–63 ml/yr) compared to the –48 ml/yr decline estimates of the other two cities. It must be stressed that these results are unsubstantiated, as there was no convergence of the estimates.

It is with the *GRAIN* data set that convergence was achieved to yield stable estimates; the results of the maximum likelihood analyses are presented in Tables 9.9 and 9.10. No significant group differences in slope were demonstrated for any of the models. The age–height adjustment made very little difference to the unadjusted slope estimates using the Random Effects model, even with starting age and height being significant predictors of FEV_1 . The coefficient for the time–dependent covariate, which was an ordinal variable of the amount of cigarettes smoked, was found to be positive and significantly greater than zero, in the full Random Effects model. With the inclusion of this variable, starting age was no longer a significant determinant of FEV_1 . In comparison to the unadjusted model, within–subject variance actually increased slightly

for the age–height adjusted model but showed a more definite drop in the full covariate model. The variance of the slope estimates from the population slope decreased to 112 ml² in the full model from 590 ml² for the age–height adjusted RE model. However, the ratio of the between–subject to within–subject variability was lowest for the full model form, implying a comparatively less favorable goodness–of–fit.

An unstructured full model applied to this data set, also resulted in nonsignificant group differences in FEV₁ decline. The FEV₁ decline estimates for the unadjusted model were very similar between the two approaches, but for the age/height and unadjusted models, the unstructured covariance model yielded higher estimates, ranging from –18.9 to –22.0 ml/yr. In this case, CIGLVL, with a coefficient of –13.0, was not a significant predictor, while AGE at the start of the study was. A limitation of this analysis is the use of categorical rather than continuous estimates of smoking behavior. Both the log–likelihood ratio test, a measure of the deviance and Akaike’s Information Criteria (AIC) showed that the four covariance parameter based random effects model was more appropriate when all available variables were used. However, for the age and height adjusted model, the unstructured covariance model had a better goodness–of–fit.

Nonlinearity of Decline

All of the models presented so far describe a linear relationship between FEV₁ or FEV₁ decline and year of follow–up. Constraining the analysis of longitudinal decline in lung function to a linear based estimate of change with time, fails to take into account any pattern of accelerated decline or of levelling off. In order to overcome this limitation in the analysis two methodologies were applied to detect the type and extent of curvature of the individual FEV₁ values over time. A quadratic function of time was used in order to describe a type of curve that showed a progressive increase in the decline. This "accelerated" decline is indicated by a significant negative coefficient

for the regression of the squared term of years of follow-up (YRDIF). Initially, a "cross-sectional" regression was done with the inclusion of both the linear and the quadratic terms in the analysis. This allows for the maxima, or highest level of FEV_1 reached in the subjects lifetime to be centered at any point along the x-axis of time. On a "cross-sectional" basis, neither the linear nor quadratic terms were significant on any of the data sets with the exception of the Winnipeg *VETERANS* group and the alive smokers group of the *COAL* data set. A problem associated with this analysis was collinearity between the linear and quadratic terms of time.

On an individual basis, for the *GRAIN* data set in particular there were many individuals with less than the required minimum of 4, and very few individuals (13 grain and 4 civic workers) showed statistical significance in their coefficients. Of the 385 coal miners, 62 had only three points and therefore could not be analyzed in this way. A greater proportion of individuals who have significant quadratic terms was found in the dead nonsmokers and the alive smokers groups. Among the *VETERANS* only 15 individuals in total showed a significant quadratic decline according to this model of linear and quadratic terms.

Table 9.11: Mean Quadratic and Allometric Decline Estimates (Method E)

GROUP	Bsq	SE Bsq	Blog-log	SE Blog-log
COAL				
1 Dead Nonsmokers	-3.8	0.7	-0.18	0.01
2 Alive Nonsmokers	-2.4	0.4	-0.10	0.01
3 Dead Smokers	-4.7	1.0	-0.17	0.01
4 Alive Smokers	-3.4	0.7	-0.13	0.01
Sig. Differences	(2,4 vs 3)		(2,4 vs 1,3)	
Overall:	$FEV_1 = 2878 - 3.7\text{yrdif}^2$			Mean $R^2=0.79$
	$\log FEV_1 = \log 2818 - 0.15\log\text{yrdif}$			Mean $R^2=0.65$
VETERANS				
1 Toronto	-8.0	1.5	-0.13	0.01
2 Winnipeg	-3.9	0.9	-0.13	0.02
3 Halifax	-6.0	1.8	-0.17	0.02
	No Sig. Differences			
Overall:	$FEV_1 = 2527 - 5.9\text{yrdif}^2$			Mean $R^2=0.54$
	$\log FEV_1 = 2455 - 0.14\log\text{yrdif}$			Mean $R^2=0.44$
$\log\text{yrdif} = \text{logarithm to the base 10 of years since first measurement}$				

Longitudinal analysis of quadratic or allometric declines were based upon a regression analysis of individual data (Method E). As the estimates for the *GRAIN* data were highly unstable, with only a maximum of four data points of follow-up per individual, this method of analysis was not applied to the *GRAIN* set. However, another indication of nonlinearity in the FEV_1 decline was the significance of initial age as a predictor of decline. For *GRAIN* workers a pronounced gradient of an increase in FEV_1 decline was evident with older age groups (-36.1 ml/yr in the 50+ age group versus -5.5 ml/yr in the under 30 group), especially for the *GRAIN* handlers group.

The average estimates of the quadratic coefficient (BSQ) and the coefficients obtained by the log-log relationship of FEV_1 with year of follow-up (Blog-log) and their standard errors are shown in Table 9.11 for the *COAL* and *VETERANS* data. The *VETERANS* had larger coefficients of quadratic decline than that found for the coal miners, but very little difference was found for the logged coefficients.

Although the focus is on the nonlinearity aspects of the declines, both models are linear-based, and thus the coefficient of determination (R^2) provides a useful guide of the proportion of variance explained by the linear model. The highest R^2 was found for the quadratic model of the *COAL* data. In both the quadratic and allometric coefficients for the *COAL* groups there was a parallel pattern, in that the deceased groups showed a statistically significant relationship of both having the largest quadratic and allometric decline estimates than that of the alive groups. For the *VETERANS*, the standard errors were generally larger than those found for the coal miners. No statistical significance in the differences of the average quadratic coefficient were found between the cities; although there was a trend with Winnipeg having a lower average value than the other two cities. The *VETERANS* allometric analysis also yielded no significant differences between the groups, although the Halifax level was higher than

Toronto or Winnipeg.

In order to best describe how a quadratic or allometric analysis can be used to estimate a "nonlinear" decline, individual plots were chosen from each group which visually at least appeared to show either a quadratic or an allometric decline of FEV_1 with time. Figures 9.1 to 9.7, at the end of the chapter, each are suggestive of the appearance of a quadratic type of plot first, followed by an allometric one next. While evaluating the individual plots for this exercise, a fair representation of each of the different types of plots could be found in the Toronto and Winnipeg groups. For Halifax, the majority of plots which appeared to be nonlinear, showed the accelerating decline characteristics of the quadratic model. For the *COAL* data, the alive groups tended to favour a greater representation of the quadratic curves, while the individual plots for the dead nonsmokers group appeared to be predominantly allometric in its decline.

For each plotted curve, Table 9.12 lists the FEV_1 decline terms based on either a linear, quadratic or allometric model. The first of each pair of individuals has a plot that, on at least a visual basis, conforms to a quadratic decline. For Toronto, the steep decline after a relatively shallow initial slope is indicated for case 131 by the largest quadratic term of -17.4 ml/yR^2 . The linear coefficient is also highest among all the *VETERANS* data examples. The R^2 value for the linear and quadratic models exceed that of the allometric term which in addition, has a nonsignificant coefficient of decline. The second Toronto case shows a plot that gradually tapers off in its decline over time. Such a curve is least suited to a quadratic model as indicated by the lowest R^2 of the three models. The coefficient of decline for the allometric model was 0.36% of FEV_1 per 1% of the yearly decline.

In all of the examples the linear coefficient of time is significant; suggesting that a linear relationship may be most suitable in the majority of instances, although the magnitude of the linear decline may be underestimated when the type of decline is more aptly described by a logged relationship. The only nonsignificant quadratic coefficient was observed for the plot of ID 2028 of the alive nonsmokers *COAL* group; its R^2 value of 0.50 contrasted markedly to the 0.91 value found for the allometric model for that same curve, which has a clear pattern of demonstrating a slowly decreasing change until it flattens out towards the end of follow-up period. Of the allometric models shown in Table 9.12, a Toronto data plot and three of the *COAL* data plots were found to be nonsignificant. These nonsignificant allometric curves were all examples of a progressively increasing decline with the passage of time. Each of these three examples were best fit by a quadratic model, with each having an R^2 of 0.95 or greater. The largest FEV_1 decline in each data set was indicated by both the linear and quadratic estimates. The highest coefficient of the allometric relationship was -0.50 for a Halifax veteran; it was not among the higher linear or quadratic coefficients, demonstrating a different type of relationship of FEV_1 with time.

A useful parameter that emerges from using the allometric model is that one can predict the number of years it would take to reach 50% of the initial value, regardless of the level of the initial value. The higher Blog-log values predict a precipitous drop in FEV_1 in a short period of time; the more typical coefficients which range between 0.08 and 0.17 would take anywhere between 16 to 256 years in order for a value to fall to half its original level, using the log-linear model.

Table 9.12: A Comparison of Individual FEV₁ Decline Coefficients Using a Linear, Quadratic or Allometric Model

GROUP	Individual Case #	Linear* Term	R ²	Quadratic Term	R ²	Allometric Term	R ²
<i>VETERANS</i>							
1 Toronto	131	-157.1	0.56	-17.4	0.64	(-0.23)	0.31
	156	-76.3	0.80	-5.9	0.63	-0.36	0.83
2 Winnipeg	258	-32.2	0.59	-1.6	0.70	-0.08	0.37
	250	-103.5	0.65	-8.2	0.54	-0.34	0.71
3 Halifax	421	-117.9	0.67	-7.9	0.82	-0.24	0.72
	403	-112.0	0.93	-5.9	0.81	-0.50	0.80
<i>COAL</i>							
1 Dead	1074	-84.9	0.80	-4.9	0.95	(-0.19)	0.41
Nonsmokers	1029	-68.1	0.93	-3.1	0.79	-0.17	0.88
2 Alive	2036	-57.7	0.89	-3.4	0.97	(-0.08)	0.55
Nonsmokers	2028	-43.7	0.70	(-2.1)	0.50	-0.17	0.91
3 Dead	3101	-105.3	0.91	-7.1	0.99	-0.17	0.61
Smokers	3107	-77.8	0.80	-4.3	0.56	-0.29	0.88
4 Alive	4069	-84.4	0.94	-3.2	0.99	(-.22)	0.53
Smokers	4102	-87.9	0.93	-4.2	0.85	-0.20	0.69

() = nonsignificant, $p > 0.05$

* = Method E

Among the *COAL* workers, 264 out of 382 had significant coefficients of quadratic decline; for 98 out of 141 veterans, this was also the case. A descriptive analysis was performed in order to determine the characteristics that distinguished those whose nonlinear decline was significantly greater than zero. In the *VETERANS* study, whether the coefficient was significant was positively related to the length of follow-up, as well as the number of measurement occasions. A higher average linear slope, a higher average age of smoking onset and a lower final FEV₁ measurement were also found for those with a significant quadratic decline. For the *COAL* data on the other hand, no significant differences were found, with the exception of a higher average linear intercept and a positive relationship with the number of measurements found for those who showed significant quadratic decline.

Those with significant allometric coefficients, did not have any results which differed from the previous analysis except that of the average initial vital capacity, which was shown to be slightly lower for those with a significant allometric decline.

Practical Applications

Smoking Behaviour

Using the generalized least squares methods on both data sets resulted in significant coefficients, but for the Random Effects model it was positive, implying an increase in FEV₁ level with an increase in the level of smoking, after the effects of other covariates on the dependent variable were accounted for. On the other hand, all of the OLS methods applied to the *VETERANS* and *GRAIN* data showed significant coefficients relating the level of smoking to greater FEV₁ decline or lower FEV₁ levels. The limitation of using this measure of smoking level is that it is an ordinal variable and thus does not imply equivalent distances between values. Therefore, an alternative

analysis was used in which the median number of cigarettes smoked in each category was substituted for each level. The partial correlation observed for this smoking variable was even smaller when trying this approach.

Smoking behavior, as shown in Table 9.13, is separated into four categories:

- 1) Lifelong never smokers;
- 2) No smoking during follow-up;
- 3) Either stopped smoking or smoked intermittently during the follow-up;
- 4) Continuous smoking during the follow-up period.

Table 9.13: Smoking Characteristics of the *VETERANS* and *GRAIN* subjects during follow-up

GROUP	Never	No Smoking	Stopped/ Intermittent	Continuous
<i>VETERANS</i>				
1 Toronto	0 (0%)	5 (9.8%)	18 (35.3%)	28 (54.9%)
2 Winnipeg	0 (0%)	5 (8.9%)	13 (23.2%)	38 (67.9%)
3 Halifax	1 (2.9%)	4 (11.8%)	6 (17.6%)	23 (67.6%)
<i>GRAIN</i>				
1 Grain	59 (22.4)	70 (26.6%)	79 (30.0%)	55 (20.9%)
2 Civic	22 (39.3)	19 (33.9)	9 (16.1)	6 (10.7)

A chi-squared analysis of the crosstabulation of the *VETERANS* groups showed no statistical significance in the proportions of the different categories, unlike that shown for the *GRAIN* data. Only one of the total of 141 veterans was a lifelong nonsmoker; the predominant category was that of continuous smokers, with over half of the subjects in each city belonging to that category. The grain handlers were fairly evenly distributed among the smoking categories; for the civic workers, the largest percentage were never smokers or nonsmokers during the follow-up.

As a further analysis, these four smoking categories were collapsed into two groups. For the *GRAIN* data, the nonsmokers group consisted of the never and nonsmokers during follow-up, with the smoker group forming the other two categories. An analysis of variance was performed, using methods C, D, and E to assess any main effects of smoking and group variables as well as interactions between them. None of the terms were statistically significant in the analysis. In the regression analysis of Method A, there was also no difference in the levels of FEV_1 or the decline estimates according to the two smoking categories in the unadjusted and full models. The same result was found for the *VETERANS* data. Another characteristic looked at in the *VETERANS* data, apart from the covariates in the full model, were age at which the subject started smoking and age when they first developed a cough. Both variables showed statistical significant differences between the smoking groups according to an unpaired t-test. Those in the nonsmoker category that had quit smoking before the study began, started smoking at a later age (19.7 years) than the smokers (17.8 years). As well, the nonsmoker group complained of coughing which first started at a later age of 32 years compared to an average of 28 years in the smoker group.

The *COAL* data set was presented as four separate groups, with each being a combination of vital status and smoking category. A two-way analysis of variance was performed so that smoking and vital status were each entered separately in the models.

No interactions between the two terms were found to be statistically significant, however, the main effects were each statistically significant in methods C and E. With covariate adjustment for differences between the groups in height, age at first measurement, the years since retirement to last measurement and the total number of years of follow-up, the smoking category for methods C and E were no longer a statistically significant main effect. The unadjusted FEV₁ decline estimates had average levels in ml/yr of 56.1 for nonsmokers versus -65.0 for smokers; those that were alive at follow-up had a decline estimate of 53.9 compared to the dead subjects value of 68.9.

Characteristics of the Dead individuals in the VETERANS Study

The only information available on vital status that was available for the *COAL* data set was whether the group of subjects studied were alive or dead at the time of ascertainment. With the *VETERANS* data, however, information is available to separate whether individuals were alive when they completed the follow-up series; whether they were lost to follow-up; or whether their deaths were recorded within two years of their last follow-up measurement occasion. Of the 141 total individuals in the *VETERANS* study, 75 were classified as alive, 31 were lost to follow-up and 35 were dead by the end of the study. Among those 31 individuals in the lost to follow-up category, 12 left for unknown reasons, 9 moved, 8 refused and 2 stopped participating because they were unwell. The descriptive analysis was based on a one way analysis of variance and Scheffe test for multiple comparisons. The means of the individually-based estimates of the regression slope and intercept were averaged for each of the groups. The intercept was found to be significantly higher for the alive group and lowest for the dead group. For the average of the unadjusted linear slope estimates, a trend was observed where the greatest decrement was shown in the alive group and the lowest decrement among the dead group. The alive group, as could be expected, were found

to have significantly more measurements and were followed for a greater number of years, while the dead group was significantly older than the other two groups. The age when coughing started was highest for the dead group and lowest for the lost to follow-up subjects. Of the pulmonary measures, it was for initial vital capacity and initial fraction of carbon monoxide uptake, that the mean values for the dead group were significantly lower than the others. The initial residual volume showed the expected opposite pattern in which the average value was higher for the dead group than for the other two groups. Although the initial FEV₁ were somewhat lower for the dead group this comparison just failed to reach statistical significance. For the FEVCV variable, a trend was shown in which the mean was higher for the dead group in comparison to the lowest value for the live group.

The cross-tabulation of vital status by smoking behaviour revealed a statistically significant difference in the categories with 74.3% of the dead group and 83.9% of the lost to follow-up group being continuous smokers in comparison to only 49.3% of the alive group. Halifax did show a trend of having the greater percentage of deaths (38.2%) in comparison to Winnipeg (26.8%) and Toronto (13.7%). The trend is consistent with Toronto having the shortest follow-up period.

For the preceding descriptive analyses, individuals that died more than two years beyond the last measurement occasion were treated as lost to follow-up. In a subsequent analysis these individuals who died within the total follow-up period were classified with the dead group, such that the two resulting categories contained those who were alive up to the end of follow-up and those who died within the follow-up period. The 39 veterans who died during the course of the follow-up period, were significantly older than the 102 alive veterans. Although there were more continuous smokers in the dead group (76.9%) than found in the alive group (57.8%) the differences were now not significant.

On an individual basis, of those 35 veterans who died within 2 years of the last measurement, 2 died from respiratory causes, 14 died from cardiovascular disease, 5 died from respiratory cancer while the other deaths were attributed to unknown or other causes. The plot of FEV_1 over time for two individuals who died of respiratory causes showed, for one individual, an inverse "S" curvilinear form such that both the quadratic and the allometric coefficients were significantly negative. For the other individual, the data points appeared as a cloud of values, with very little shape or form possible to be discerned among them; the coefficient of variation value for FEV_1 was very high for this individual. This variability of the FEV_1 measures was extreme. Among the individuals with respiratory cancer, 3 showed significant allometric declines of 0.10 to 0.11 percent FEV_1 change per percent yearly change.

The Effects of Retirement on FEV_1 Decline

The policy in effect over the working life of most of the *COAL* miners was for mandatory retirement at age 50; but with working time extended for pension adjustments, the average age of retirement of the *COAL* miners was 54.3 years and ranged from 35 to 73 years. A regression of all the data points showed a significantly higher level of FEV_1 before retirement as opposed to after retirement, which could be expected on the basis of time related changes alone. Although the slope was 8.1 ml less after retirement, the difference was not significantly different. The significance of these relationships stays the same even when age of retirement is adjusted for in the model. When it was the change in FEV_1 that was the dependent variable (Method B), not only was there no difference in slopes before and after retirement, there was also no difference in the change in FEV_1 levels in the two different periods. The age of retirement was not a significant predictor variable in this case.

It was when the *COAL* groups were divided according to smoking status that differences emerged in this method of analysis. Before retirement, nonsmokers showed a high average decline of 69.5 (S.E.=6.1) ml/yr; after retirement the average FEV₁ decline coefficient was 49.6 (S.E.=3.5) ml/yr and the difference was significantly different. With smokers, the opposite relationship was evident. This group experienced a rate of decline of 56.4 (S.E.=4.5) ml/yr before the onset of retirement. After retirement, the FEV₁ decline estimate increased significantly to 67.4 (S.E.=3.3) ml/yr. This interaction of smoking status with pre- versus post-retirement follow-up time had resulted in an overall nonstatistically significant change in pre- versus post-retirement slopes using this methodology. When using an overall regression of FEV₁ level with time (Method A), no such pattern emerged in this profile analysis.

Characteristics which may distinguish smokers from nonsmokers in these groups were assessed through an unpaired t-test analysis. The nonsmokers had a slightly longer average period of follow-up, their average age of retirement was 55 years (as compared to 53 years in smokers) and their age at the first test was 48 years on average, in comparison to 45.7 years for the smokers. Even though the age of retirement was higher, the number of years they were followed-up since retirement was significantly longer at 13, as opposed to 11.1 years of follow-up. The only significantly different initial respiratory function parameters were vital capacity and FEV₁ which were both significantly less for the nonsmokers group. It appears that the 66 nonsmokers were, on average, an older group with lower initial spirometric measures of lung function than those found among the 107 smokers of this *COAL* miners group. The fact that they were followed on average for a longer period after retirement, yet still had lower decline in FEV₁ change between measurements, supports the observation that reduction in the rate of FEV₁ decline after retirement was not biased by a shorter period of follow-up

Initial Lung Function

There is debate in the literature as to whether FEV_1 values should be used as an adjustment for FEV_1 decline. The first measurement occasion FEV_1 was found to be highly predictive of average FEV_1 and of FEV_1 slope in most cases. This was true even in the case of the *VETERANS* data where initial FEV_1 was actually the mean of up to 12 monthly measurements. As the initial value of a slope is an usually an influential observation, a strong relationship between initial value and slope is not unexpected. Are any other of the measured lung function variables highly predictive of later decline in FEV_1 ?

Table 9.14 lists the results of an analysis of covariance on the unadjusted slopes (Method E) and the percentage of FEV_1 change (Method D). Similar results were obtained when a more complete model was used. In the *COAL* data set, where the initial values of each lung function measurement are separately used to adjust the relationship of FEV_1 decline by group, initial vital capacity was significant for Method E, but not Method D. What is surprising about this analysis is that both diffusing capacity and fractional carbon monoxide uptake were significant predictors of FEV_1 change even though these measurements were taken sporadically on the individuals and never during the first measurement occasion. The DLCO coefficient was not however, a significant predictor of FEV_1 decline. Using Method E, the correlation with each of the initial values was not as high as was found with initial FEV_1 level. The initial vital capacity was a stronger predictor than that of carbon monoxide uptake or diffusing capacity.

Table 9.14: Initial Lung Function as Predictors of FEV₁ Decline.

DATA SET	PREDICTOR	Method E (Δ FEV ₁)	Method D (Δ %FEV ₁)
COAL	FEV ₁	-7.3	0.46
	FVC	-13.7	(0.03)
	FCO	80.1	5.3
	DLCO	(.62)	0.07
VETERANS	FEV ₁	-9.6	+0.61
	VC	-13.6	(+0.24)
	MMF	(-1.9)	+0.45
	FCO	(90.1)	11.3
	DLCO	(0.93)	0.16
	RV	(0.28)	-0.50
GRAIN	FEV ₁	(3.5)	0.23
	FVC	(1.8)	0.15
	MMF	3.3	0.16
() not significantly different from zero using ANCOVA F test			

With the *VETERANS* data there was a much greater choice of lung function values and not only were all measurements taken at the initial testing session, but the initial values are actual averages of monthly measurements taken. It is to be noted, in this data, that the VC was not a forced vital capacity measured in the same spirometer as the FEV₁, but was a slow vital capacity measured on a helium closed circuit. As shown in Table 9.14, when each were entered independently in the full analysis of covariance model (Method E), only initial vital capacity and FEV₁ were significant covariates. Almost the opposite results were seen if percentage change in FEV₁ was used as the dependent variable (Method D). There, initial vital capacity was no longer a significant predictor; in fact it was the initial FEV₁, MMF, residual volume, fractional carbon monoxide and diffusing capacity that were each predictive.

For the *GRAIN* data, MMF was a significant predictor for both methods. All spirometric measures were positively related to the percentage-based slopes.

Longitudinal Decline of Various Lung Function Measures

FEV₁, and its changes with time, has been used as the dependent variable in all the analysis so far. Each data set has at least one other lung function which can be used as the dependent variable in a longitudinal analysis. Results of such analyses are found in Table 9.15. In the *COAL* data, FVC was the only other lung function measure taken at each testing occasion. The coefficient of variation for all of the measurements of forced vital capacity was 20.2% in comparison to that of 34% found for FEV₁. Regression analysis was performed on each individual's data using FVC as the dependent variable. The decline coefficient was significant for all the groups, with the alive nonsmokers having the lowest value of -38.8 ml/yr while the deceased smokers had the highest FVC change of -64.8 ml/yr. The group order was the same as that found for FEV₁ decline; but the average FVC decline was now significantly less for the alive nonsmokers compared to the dead nonsmokers. No significant correlation was found between the two decline coefficients.

Among all the lung function measurements found in the *VETERANS* group, MMF had the highest coefficient of variation of all the points with fractional carbon monoxide uptake having the lowest. A linear regression of all the *VETERANS* data points showed significant declines in all the lung function variables, but for residual volume the change was positive as would be expected. The MMF decline coefficients were found to have the highest significant correlation with those of FEV₁ ($r=.59$) followed by that of DLCO and VC. Group estimates were much more informative in distinguishing the value of each of the lung function measures. For the variables of VC, MMF and FCO the average of the individual regression coefficients relating the

dependent variable with years of follow-up, produced coefficients which were significantly different from zero for all *VETERANS* groups. The average annual change in vital capacity was similar in magnitude to that of FEV_1 for Toronto and Halifax. The VC decline for Winnipeg was so much lower that it differed significantly from the other two cities. For MMF, a different pattern of decline was observed in that Winnipeg showed the greatest annual decline, being significantly different from Toronto's average MMF change, which was 25 ml/yr less. It was also for Winnipeg that the greatest average annual increase in residual volume was seen; for the Halifax group a negative coefficient was estimated, implying an unexpected decline with time, however, the coefficient was not significantly different from zero. Winnipeg was the only city in which the *VETERANS* showed a significantly negative average DLCO coefficient. No significant differences between the cities was apparent for FCO decline.

For the *GRAIN* data set the coefficient of variation of the data points was much higher for MMF. Neither the average annual change in forced vital capacity (FVC) or MMF were significantly different between the two groups, which was in agreement with the FEV_1 results. For this data set the values were consistent in magnitude. Both FVC and MMF decline were highly correlated with FEV_1 decline in this data set, yielding significant correlation coefficients of .74 and .67 respectively.

When a random effects model was applied to the *GRAIN* data, there was very little difference in the unadjusted FVC decline (11.3 ml/yr for the grain workers, versus 11.8); the larger difference found for MMF decline (63 ml/sec/yr versus 55.5 in the civic workers) was also not statistically significant. Significant predictors of the full random effects MMF model were group, time, FEV_1 , height, and initial age. The standard deviation of the individual MMF slopes was 45.8 ml/sec, which was almost equal to the yearly decline estimate.

Table 9.15: A Comparison of Unadjusted Lung Function Decline Estimates Based Upon Regressions of Individual Data.

LUNG FUNCTION					
COAL	1. Dead Nonsmokers	2. Alive Nonsmokers	3. Dead Smokers	4. Alive Smokers	Significant Differences
FEV ₁	-63.9 (3.1)	-45.7 (4.0)	-71.2 (3.7)	-57.7 (2.8)	3 vs 2,4
FVC	-61.7 (3.7)	-38.8 (4.1)	-64.8 (3.6)	-50.2 (2.8)	3 vs 2,4
VETERANS					
FEV ₁	1. Toronto -65.8 (5.9)	2. Winnipeg -50.2 (5.9)	3. Halifax -57.3 (7.6)		N.S.
VC	-71.3 (10.6)	-21.6 (5.3)	-52.4 (7.7)		2 vs 1,3
MMF	-46.5 (7.9)	-77.9 (7.3)	-70.1 (10.0)		1 vs 2
RV	+29.7 (7.7)	+44.8 (6.6)	<-15.9> (11.7)		1,2 vs 3
FCO	-0.32 (0.15)	-0.39 (0.08)	-0.51 (0.11)		N.S.
DLCO	<+0.12> (0.11)	-0.19 (0.08)	<-0.11> (0.07)		1 vs 2
GRAIN					
FEV ₁	1. Grain -19.9 (2.1)	2. Civic -19.3 (3.8)			N.S.
FVC	-12.1 (2.4)	<-13.0> (3.7)			N.S.
MMF	-62.0 (4.6)	-61.6 (8.8)			N.S.

() = Standard Error

< > = Not significantly different from zero, using a t statistic (p<0.05)

Figure 9.1a: TORONTO - FEV₁ versus Year for Case #131

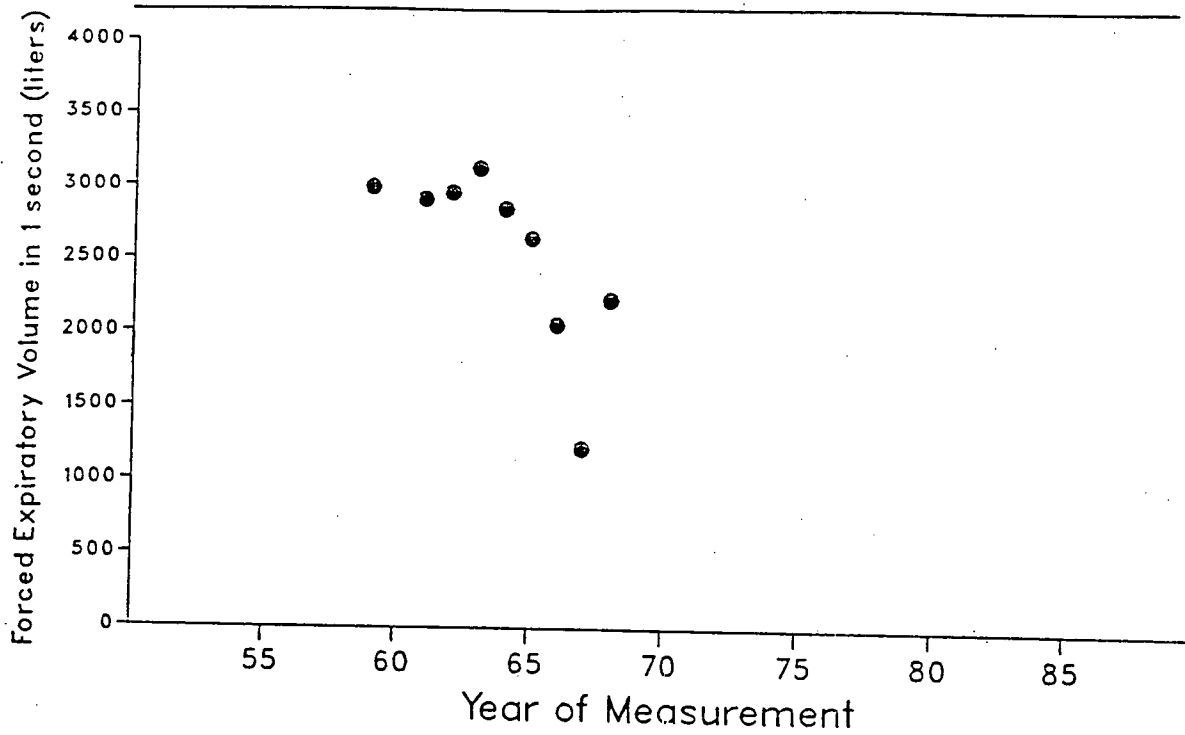


Figure 9.1b: TORONTO - FEV₁ versus Year for Case #156

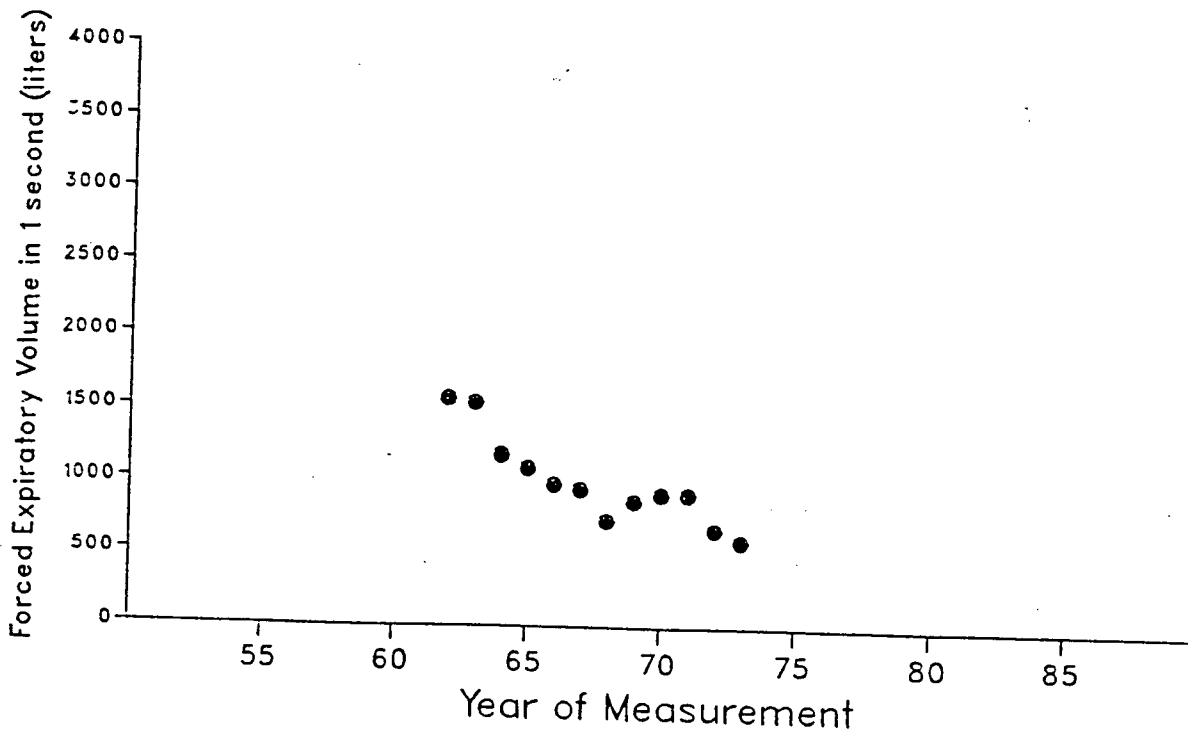


Figure 9.2a: WINNIPEG - FEV₁ versus Year for Case #258

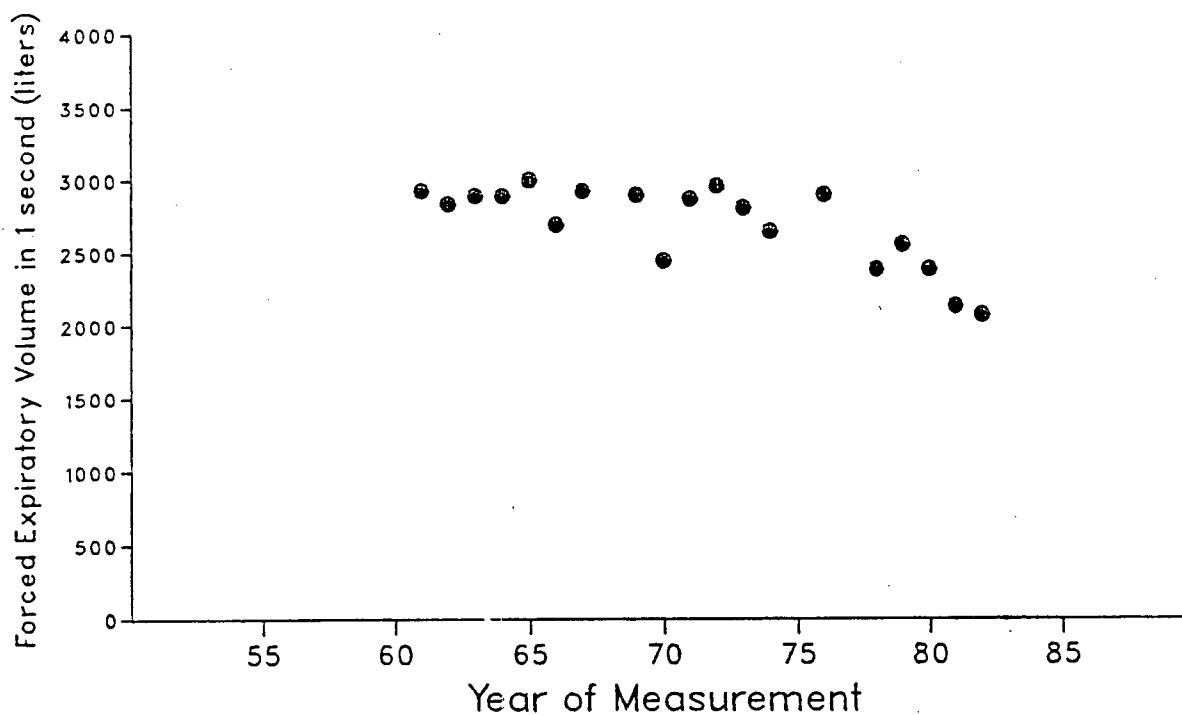


Figure 9.2b: WINNIPEG - FEV₁ versus Year for Case #250

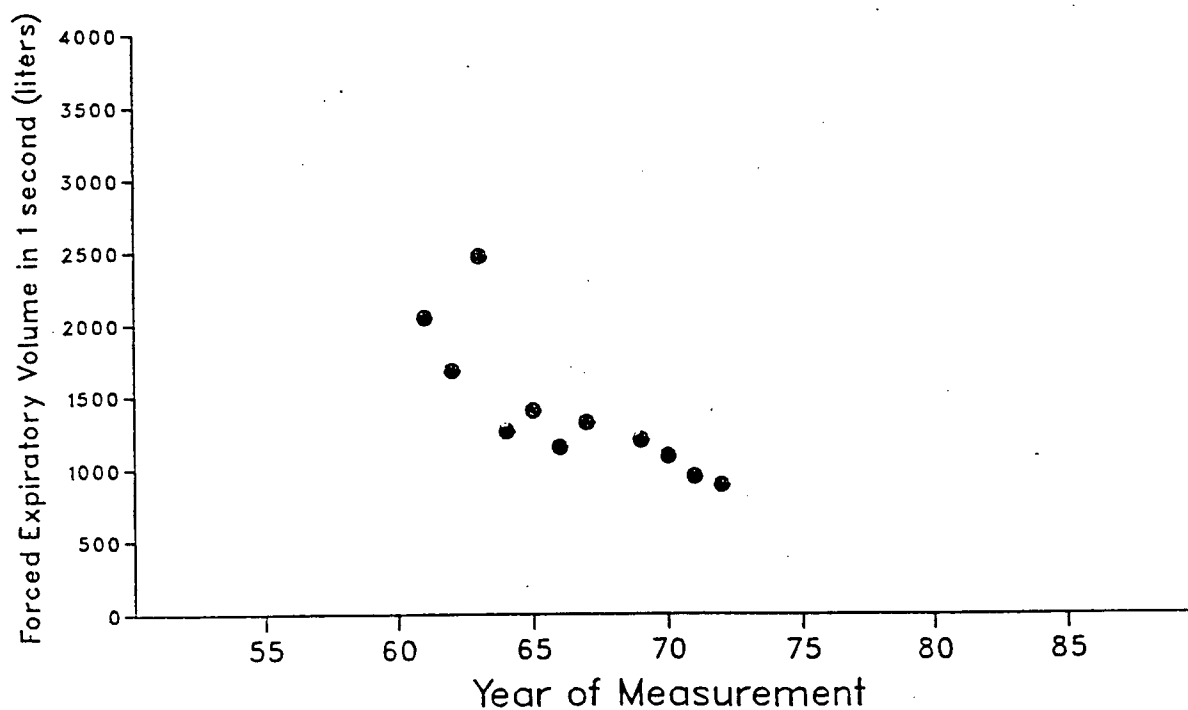


Figure 9.3a: HALIFAX - FEV₁ versus Year for Case #426

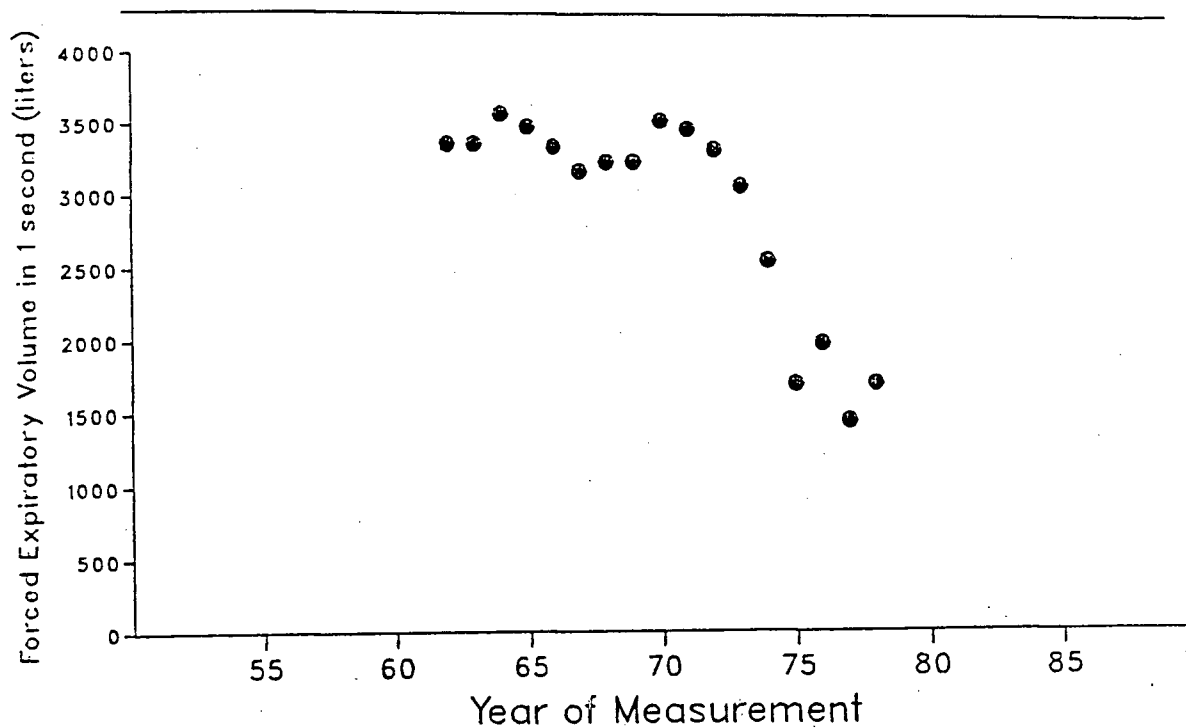


Figure 9.3b: HALIFAX - FEV₁ versus Year for Case #403

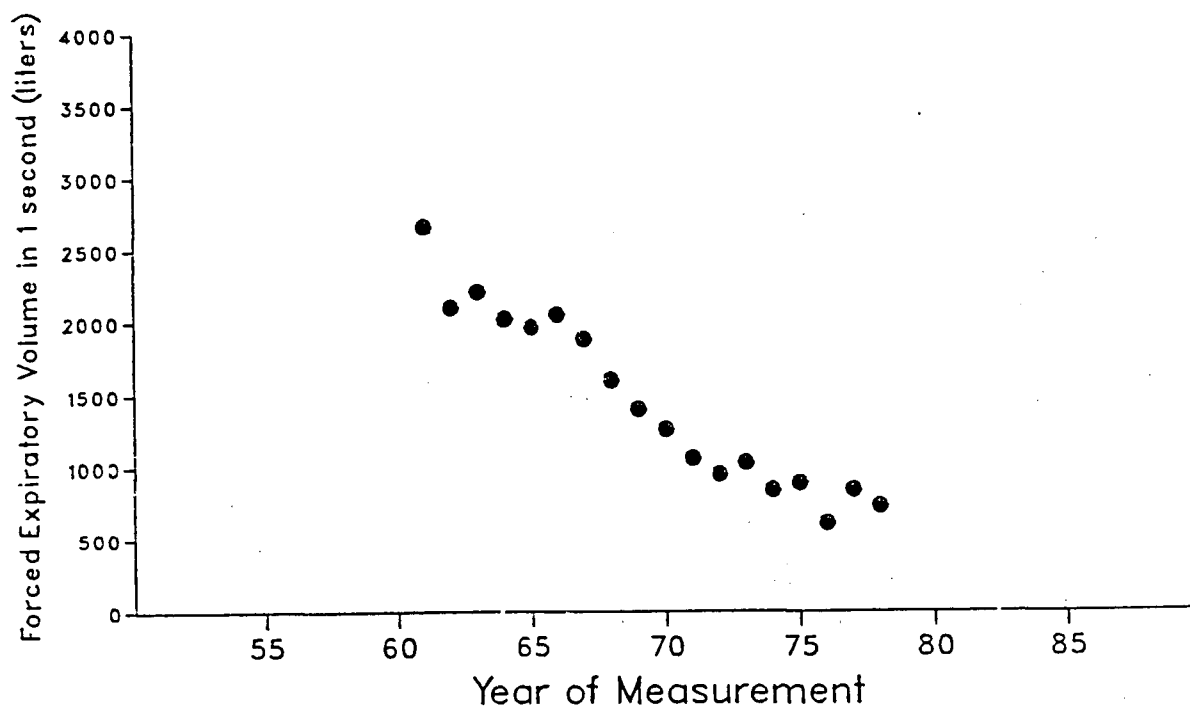


Figure 9.4a: DEAD NONSMOKERS - FEV₁ versus Year for Case #1074

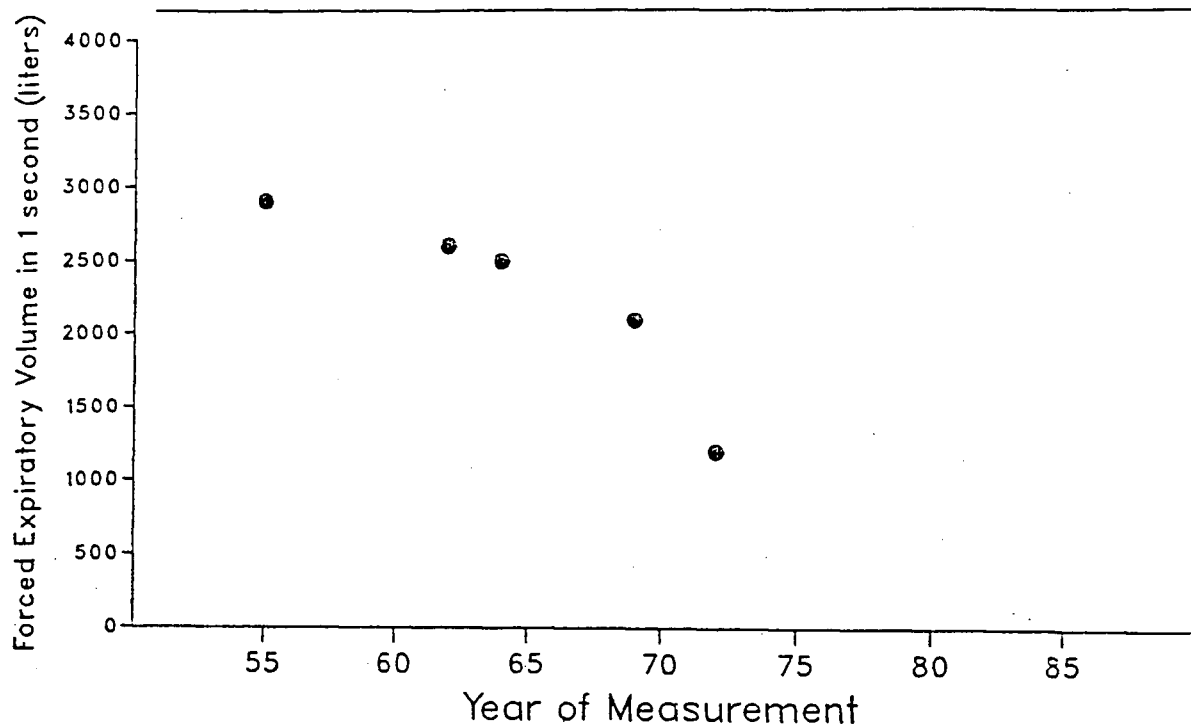


Figure 9.4b: DEAD NONSMOKERS - FEV₁ versus Year for Case #1029

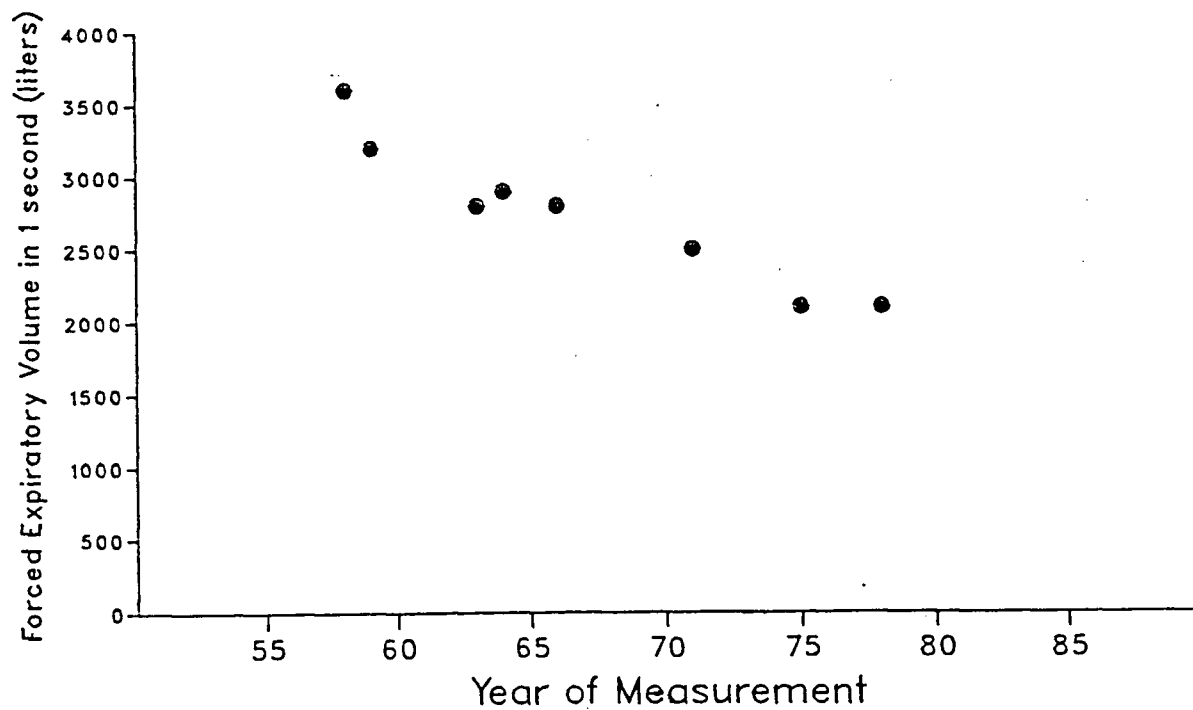


Figure 9.5a: ALIVE NONSMOKERS - FEV₁ versus Year for Case # 2036

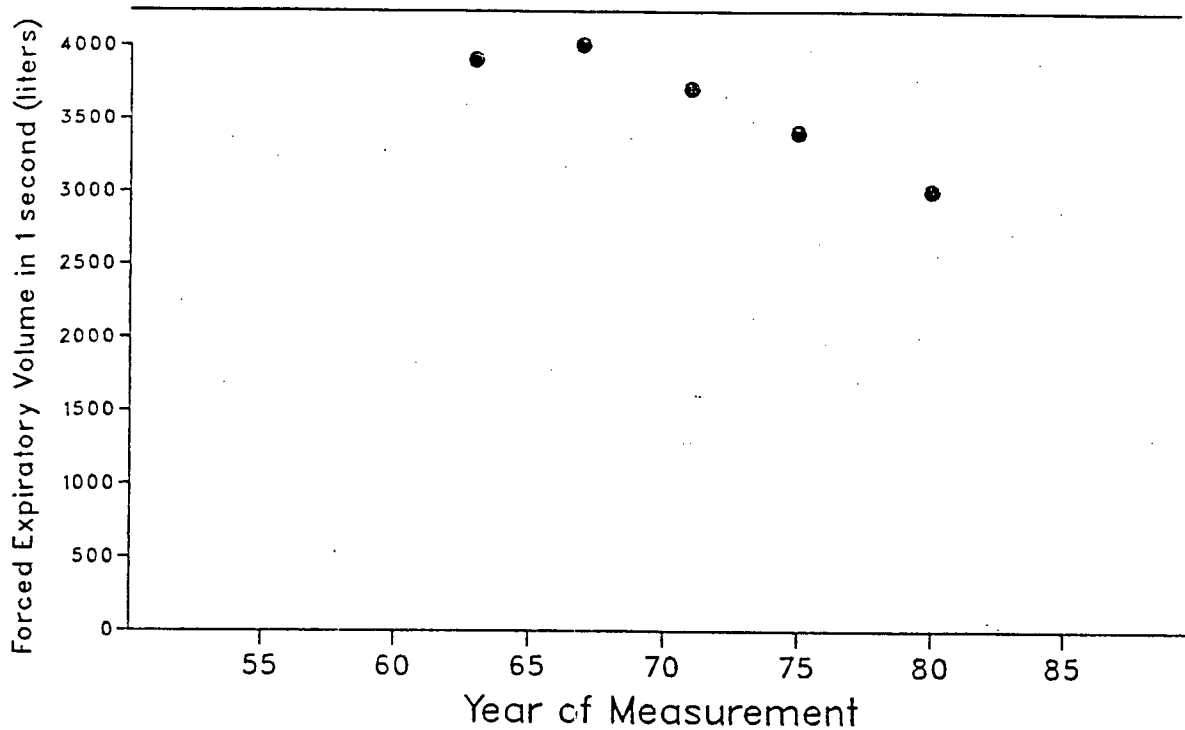


Figure 9.5b: ALIVE NONSMOKERS - FEV₁ versus Year for Case # 2028

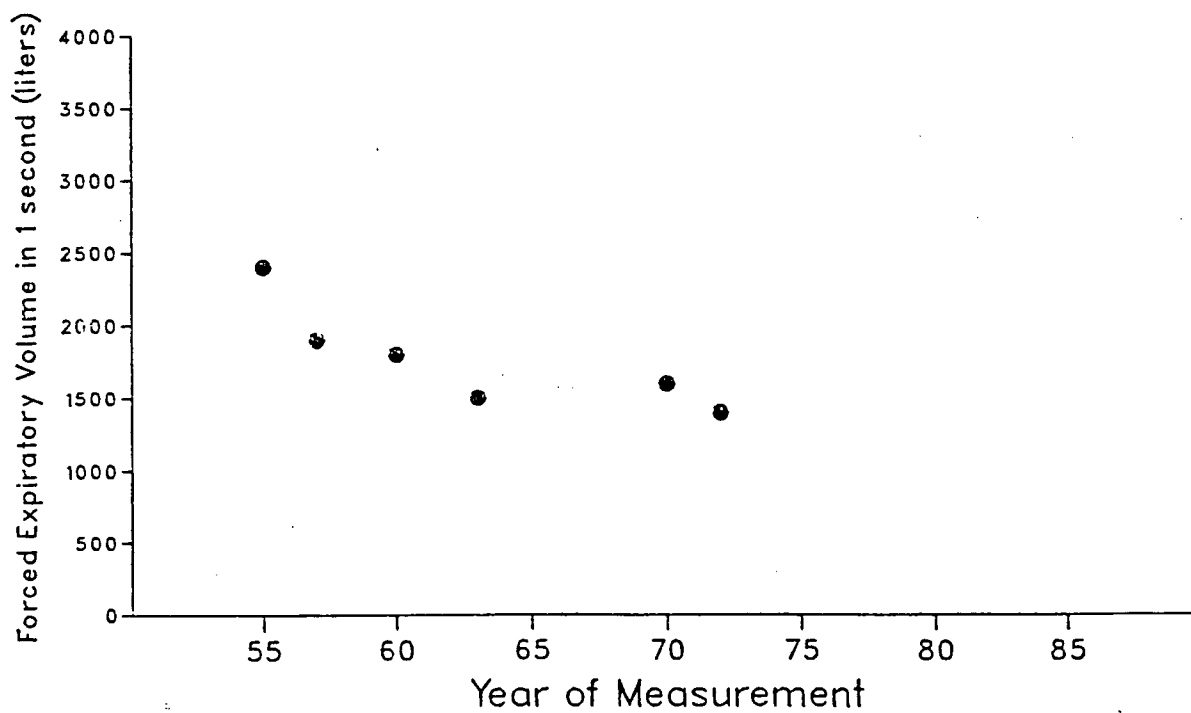


Figure 9.6a: DEAD SMOKERS - FEV₁ versus Year for Case #3101

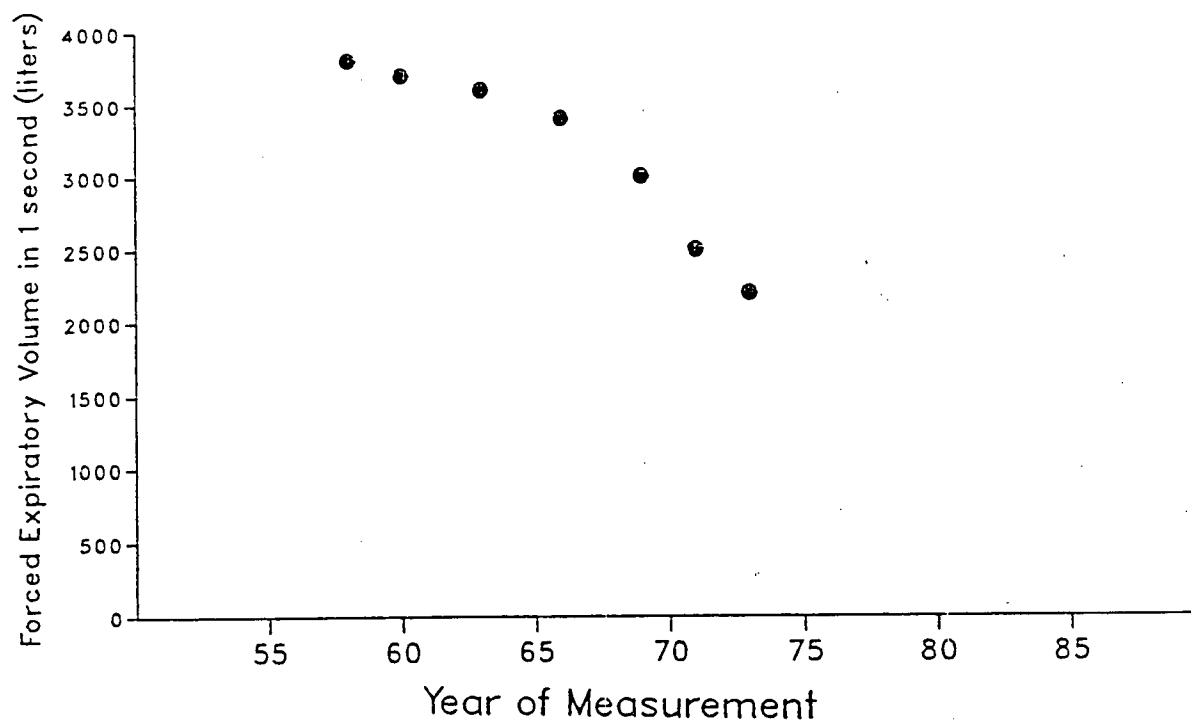


Figure 9.6b: DEAD SMOKERS - FEV₁ versus Year for Case #3107

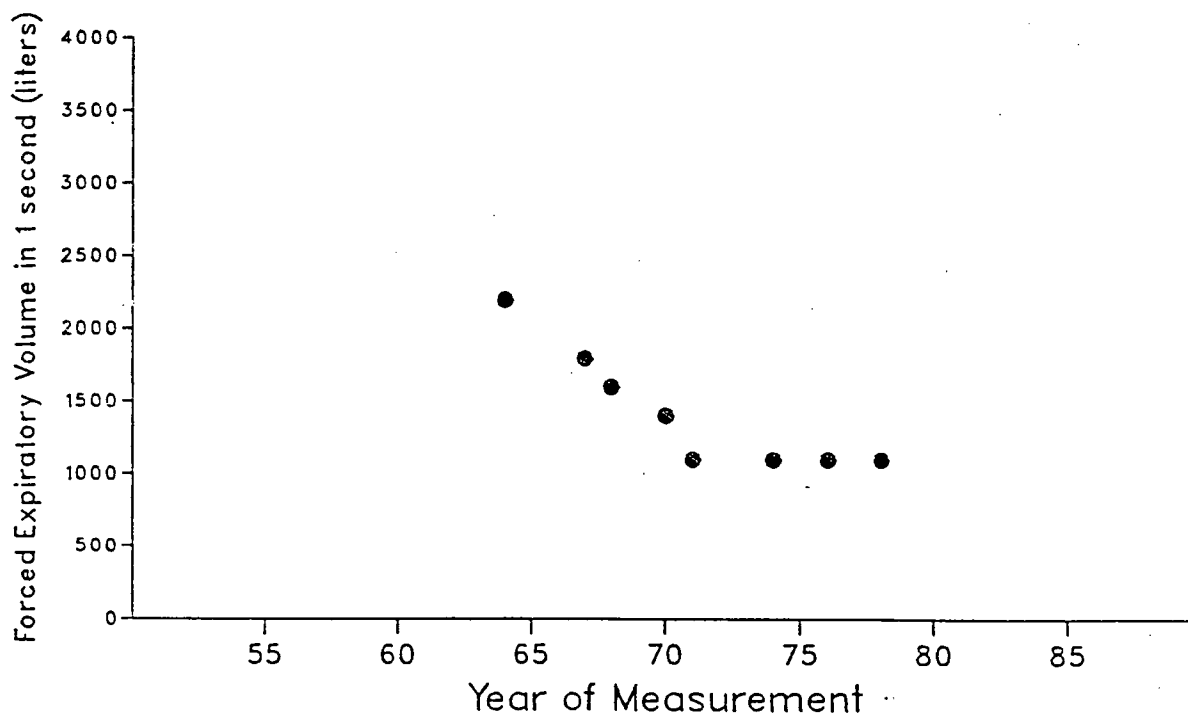


Figure 9.7a: ALIVE SMOKERS - FEV₁ versus Year for Case #4069

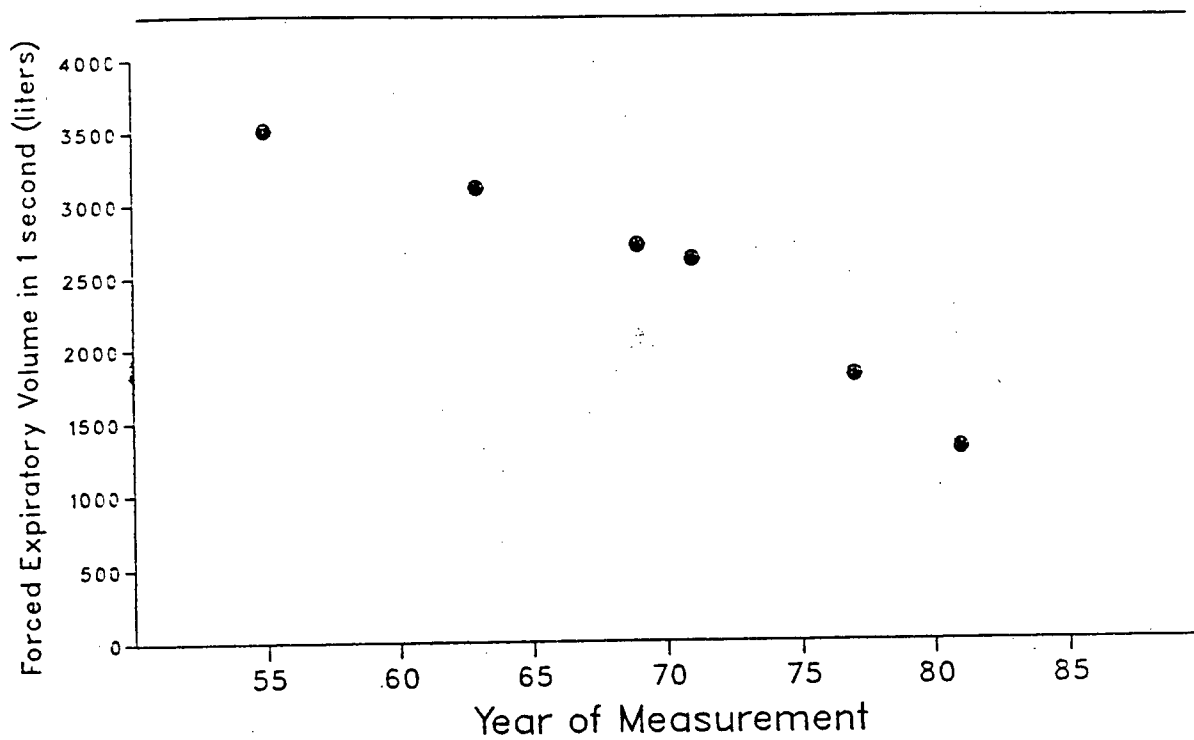
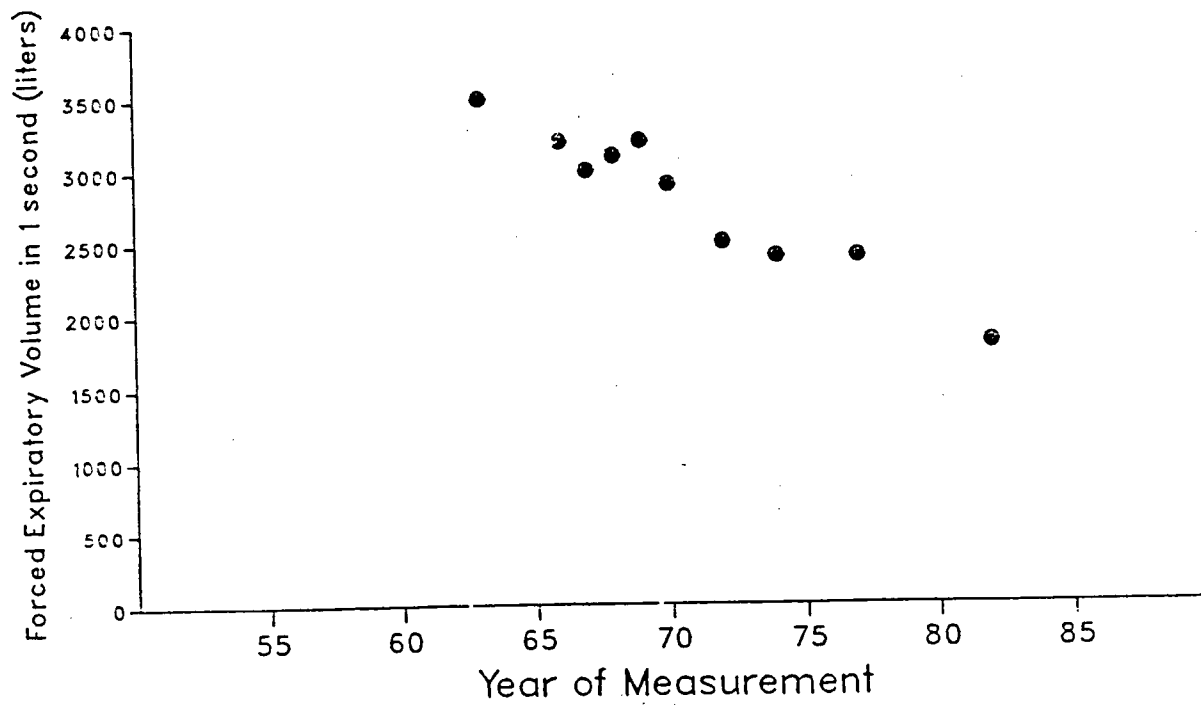


Figure 9.7b: ALIVE SMOKERS - FEV₁ versus Year for Case #4102



CHAPTER X

DISCUSSION

In the course of the present investigation, numerous issues in the application of longitudinal data analysis were addressed. This chapter therefore, will be organized into a series of discussions of the results as they relate to pertinent topics.

The first question this thesis addresses is: what difference do different methods of analysis of longitudinal data applied to the same data set, make to the conclusions that might be drawn? It is clear from the present analysis that an assessment of statistical significance of different group decline estimates will be affected by the choice of method. The variance estimates are affected by violations of the assumptions of the model chosen. For example, most of the significant differences found between group estimates of decline were the results of an overall regression method which ignores between-individual variability. Apart from differences in the magnitude of the estimates of decline, the order of decline among the groups compared can differ, as can the number of significant predictors; these have effects on the interpretation of the data. As an example of this, either the veterans from Toronto or from Halifax showed greater declines in lung function depending on the statistical method used. Similarly, the grain workers compared to the civic employees might or might not have shown a trend toward a faster rate of lung function decline. In the *VETERANS* data, a conclusion of whether age when smoking started, or the variability of the initial FEV_1 values, were significant as predictors of later decline depended upon whether regression-based decline estimates or endpoint-based slopes were used as the dependent variables.

The observation that, for the *COAL* data set, the alive nonsmokers group invariably had the lowest coefficient of decline, as did the Winnipeg group among the *VETERANS* cities, would support a conclusion that the lower declines observed in

these two groups were real, and not artifacts of the chosen method of analysis. Each finding is supportive of the literature based hypotheses that (1) nonsmokers, even in a dirty (coal mine) environment (2) coal miners remaining alive at a certain time of follow-up and (3) mild chronic bronchitics living in a nonpolluted city, all consistently showed relatively lower linear FEV_1 declines with aging.

The second question to be addressed was; Do different statistical methods produce the same comparative results when applied to different data sets? The answer depends on what results are being considered. For the *COAL* workers, the order of the groups, using the percentage change in FEV_1 estimate (Method D) was exactly the same as that found for the individual-based change estimates (Method E). Yet, this was not the case for the other two data sets. Discrepant results were produced using these same methods for assessing the relationship of different lung function variables to FEV_1 decline. The standard error of the decline coefficients were consistently lower for the individual-based regression approach in comparison to the other OLS methods for the *GRAIN* and *COAL* data sets; but not for the *VETERANS* group estimates. These results illustrate the general conclusion that the properties of any given method are dependent on which data set is being analyzed. Relevant considerations include the degree of balance in the data structure, the population characteristics, and nonlinearity of decline.

Appropriate Models of Analysis

The question arises as to what is to be gained by using a more complex method of analysis. To answer this question, one must first define what a "good" model is. It is not merely a model which contains a large variety of parameters and interactions thereof; rather, the model of choice is one which can explain a greater portion of the variance with the fewest parameters (Greenland, 1989; McCullagh and Nelder, 1983). It

should provide a closed form of solution for the parameter estimates or at least, an iterative solution that should converge easily. What is most important is that it is understandable, conveys meaning in the interpretation of its coefficients, is broad in scope, and is relevant to biological behavior. The more complicated models are those that involve modelling of the covariance structure of the repeated observations in order to more accurately estimate the standard error of the coefficients. Generalized least squares methods can be used to take into account the expected covariance of the error terms, particularly where there is evidence of autocorrelation which may be encountered in longitudinal data (Hanushek and Jackson, 1977).

What is gained by the use of an appropriate model for longitudinal data is unbiased estimates for the coefficients and their variance. This has direct relevance when assessing the significance of the differences of two population slopes. An ordinary least squares approach, which does not take between-subject variability into account will, when applied to autocorrelated longitudinal data, most often result in a nonconservative bias; that is, population decline estimates will be incorrectly found to be significantly different (Neter et al, 1985; Ware and Cook, 1982). The overall cross-sectional regression method (Method A) was an obvious example of an inappropriate model and resulted in more instances of significant differences in group estimates.

Use of the random effects model in a generalized least squares approach merits application in that the slope estimates are weighted according to their precision in estimation and the variance about the population slopes is partitioned between the variance of the individual slopes about the population slope and the variance of observations about the individual slope. For the *GRAIN* data, the most complete linear random effects model appeared to be most appropriate according to the value of the Akaike's Information Criterion (AIC). The opposite conclusion was made in favor of the maximally parameterized unstructured covariance model when assessing the performance

of the age and height- adjusted and unadjusted models. Taking into account the relatively greater within- to between-subject variability; and the biologically implausible finding of a positive coefficient for level of smoking and a positive FEV_1 slope in the full random effects model; the age/height adjusted unstructured covariance model which does not include the time-dependent level of smoking variable, appears to be the more appropriate method of analysis for this data set. The unstructured model favors application to data sets with relatively few observations taken on many individuals.

The performance of the random effects model in the present analysis might have been improved by substituting more appropriate initial parameters based on OLS results for the default values, in order to decrease the number of iterations needed to reach convergence (Laird et al, 1987). The unbalanced structure of the *VETERANS* data and relatively large number of measurements (up to 25) per individual, imposed memory limitations for the software that proved impossible to overcome. The use of time contrasts in these covariance structure models required that all measurements be taken during the same occasions for the entire group studied. This was not precisely the case for the *GRAIN* data set, where the civic workers were measured subsequent to the grain handlers and the intervals between measures were slightly smaller. Apart from the need for more flexible software, a separate analysis of each group, with more precise (monthly) intervals may improve the accuracy of the estimates.

The more balanced a data set is, the easier it is to analyse. With a fully balanced data set with no missing values, any ordinary least squares procedure which takes explicit account of between-subject variability, would be expected to give unbiased estimates of both the coefficient and its variance (Laird, 1988). The FEV_1 decline estimates for the *GRAIN* data, which was closest to a balanced structure, showed the smallest range between the different methods applied; and none of the methods showed a statistically significant difference in lung function decline between the two groups.

The choice of the best simple linear model to use on all the data sets does not rely merely on the statistical techniques of goodness of fit, as reflected by the R^2 and standard error of the estimate values. Comparing the OLS models is somewhat like comparing apples and oranges. The models differ with respect to the choice of dependent variable (FEV_1 level versus change versus slope versus %change) and they differ in what covariates can be used in the full model. For example time-dependent covariates could not be assessed in an endpoint analysis; as well, the level of smoking variable is limited to the first occasion value in an analysis where the dependent variable is either slope or percent slope.

There was no single optimal statistical method for all the data sets analyzed. The simple statistical method that offers flexibility and provided reasonable estimates of the magnitude of the slopes for all the data sets was the method of averaging the unadjusted individual regression decline estimates (Method E). The main criterion for choosing this method, apart from the relative precision of the model, was that the estimates obtained were very similar to those found for the random effects or unstructured covariance models, and the order of the group declines were similar to that found by the majority of methods used. Using the criterion of biological plausability, the method of choice was the weighting of decline estimates by a measure of the within-subject variability ($1/SEE$). The declines in FEV_1 for the coal workers who smoked were emphasized, as were those of the grain workers in comparison to their controls.

There were specific methods that proved to be better suited to one data set only. For example, the long and sporadic intervals of time between measurements in the *COAL* data were appropriate for the pairwise overall regression analysis. With the differing duration of time between measurements the FEV_1 difference estimates were more stable; as well, the method accounts for the first order autoregression

characteristics of the data, using time since first measurement differences as the dependent variable. In the profile analysis of changes in FEV_1 decline due to retirement, statistically significant differences in the slopes were found using this pairwise approach as opposed to the overall regression method of all data.

As previously noted, the estimates of decline by each method were similar in magnitude for the *GRAIN* data set, but the order of group decline estimates differed between methods. Such inconsistencies can be partly attributed to the variability in the decline estimates, as many of the younger individuals in this data set showed an increase in FEV_1 over the nine year follow-up period. An improvement in the analysis of this data set would be to separate the younger individuals (those less than 30 or 40 years of age at first measurement) in order to eliminate most of the individuals that showed an increase in FEV_1 with time.

The twenty-five yearly follow-up of the *VETERANS* data is an important documentation of the natural history of chronic bronchitis; yet it also was most difficult to model the covariance structure, and the standard error of the model estimates were all relatively high for this data set. Generally, the models based on linear declines appeared to be the least appropriate for the *VETERANS* data set, which consisted of an older cohort with lung disease who were observed for a lengthy period. Modelling linear relationships should be supplemented by models which show nonlinear behaviour, whether it be quadratic, exponential, allometric or perhaps logistic, in form. For instance, regression models have the flexibility to incorporate quadratic terms to test for acceleration of decline.

Uses of Longitudinal Pulmonary Data

The major purpose of collecting longitudinal data in pulmonary studies is to measure the rate of change of lung function with aging and to distinguish aging from exposure effects. As long ago as 1741, Sussmilch observed that, "one needs a series of good and average years if one is interested in obtaining something reliable in terms of age relationships" (Baltes and Nesselroade, 1975).

Single measurements of lung function tend to be poor predictors of eventual disability or disease, unless the measurements are already at low levels (Postma et al., 1979). The rate of change of FEV_1 has not only been related to chronic obstructive lung disease and other respiratory disorders but it has also been shown to be predictive of mortality due to non-circulatory disorders and various cancers (Ashley et al., 1975; Beatty et al., 1982). Four possible mechanisms for this phenomenon have been postulated by Menkes et al. (1985).

- Risk factors that produce lung disease could also be risk factors for other diseases that lead to death.
- Decreases in forced expiratory volumes may reflect a non-specific decrease in the strength or vigour of the subjects who subsequently die.
- Normal lung function may be necessary to protect other systems from the toxic effects of exogenous agents or endogenous metabolic processes.
- Lung dysfunction may be a secondary manifestation of diseases in other systems of the body.

Cross-sectional versus Longitudinal data

It is commonly stated in the pulmonary literature that predicted declines of age obtained from cross-sectional data are invariably greater than those obtained from longitudinal studies due mainly to the cohort effect of past noxious influences on cross-sectional observations (Ng et al., 1977). The work of Glindmeyer et al., (1982) is usually cited as being supportive of this conclusion in that the annual decline in lung function determined cross-sectionally was 3 to 5 times greater than that estimated longitudinally.

As shown in Chapter 8, it was only for the *GRAIN* data set that the cross-sectional estimates (32.5 ml/yr for FEV₁ decline among grain workers and 29.1 ml/yr for that of the civic workers) were greater than longitudinal indices, and hence the predicted discrepancy between cross-sectional and longitudinal estimates of decline was confirmed. The longitudinal FEV₁ decline estimates for the *GRAIN* data had a small range (13.8 to 21.3 ml/yr).

For the *VETERANS* data, the cross-sectional prediction of age-related FEV₁ decline of 43.9 ml/yr for Winnipeg was similar in magnitude to the estimates derived from all of the longitudinal methods used. For Toronto and Halifax, the age-height adjusted estimates of longitudinal FEV₁ decline were even higher for the simple OLS models, and most of the weighted models, than the cross-sectional estimate of 39.4 ml/yr. Siracusa et al. (1984) expressed surprise when their analysis of FEV₁ change in subjects exposed to asbestos showed comparable cross-sectional and longitudinally derived estimates; instead of the the cross-sectional values being 3 to 5 times higher than the longitudinal. They explained this observation by noting that the asbestos exposure could cause a non-linear decline such that an acceleration of FEV₁ decline occurred 10 to 15 years after the first exposure to asbestos. Among the data sets studied, the highest

averaged negative quadratic coefficient was found for Toronto veterans. Halifax showed a relationship completely contrary to that expected; the cross-sectional age coefficient of -20.7 ml/yr was only about one third of the value of the longitudinal coefficients. Non-linearity of decline may be an explanation for this unexpected finding.

For the *COAL* data set the retrospective follow-up resulted in a wide range of dates when subjects were first enrolled in the study. Because the measurements were not taken at one point in time, no cross-sectional estimates were possible.

The use of cross-sectional analyses can be a poor substitute for longitudinally derived estimates of change in lung function with aging. The value of longitudinal data are that the cause and effect relationships are preserved. Therefore it can be determined with greater confidence as to which risk factors are predictive of eventual pulmonary function decline. Determinants of the level of pulmonary function are not necessarily those that affect its decline. For example, height is a significant predictor for all groups when all the data points are regressed upon time (Method A) using FEV_1 as the dependent variable. Yet in none of the longitudinal analyses predicting slope, was height found to be a significant factor. Van der Lende et al. (1981) emphasized the value of longitudinal investigations; they found a significant effect of air pollution based on a longitudinal analysis of their data, where no association was found in their cross-sectional study.

Statistical Aspects

The usefulness of longitudinal data is subject to a number of factors associated with the original study design, its conduct and the analysis. Despite careful planning, the execution of a longitudinal design often results in an unbalanced structure for analysis due to drop-outs or deaths of some of the study subjects.

The different aspects of the longitudinal data collection and design process as represented in the three data sets, have potential to affect the applicability and outcome of the analytical methods employed. Unlike the *COAL* data set, the two prospective studies were designed to have relatively even follow-up intervals. Even with this design however, a balanced data set could not be achieved as unplanned losses to follow-up and deaths occurred sporadically in the data. The subjects for the *VETERANS* and *GRAIN* data sets all were measured initially within two years of the initiation of the study. For the *VETERANS* the maximum length of follow-up varied for each city, (15 years for Toronto, 25 years for Winnipeg and 30 years for Halifax). The average follow-up for all the groups in both data sets therefore substantially exceeded the minimum of six years required for a slope to exceed its within-subject variability (Diem, 1982).

In the present analyses, the inclusion criteria was a minimum of three data points of follow-up. This criterion has been commonly adopted in the literature. For example, Berry et al. (1973) required at least 3 data points and a duration of at least 18 months of follow-up. This allowed for a possible non-linear estimate of decline to be calculated for each subject.

It was decided not to delete any outliers observed, as recommended by Montgomery and Peck (1982). The extreme values may have been influential with individual and subsequent population estimates. Comparison of the results with previous publications describing the complete cohort population (Chapter 6) showed comparable baseline characteristics.

As all pulmonary function measurements show within-subject variability, and are correlated, serial measurements necessarily display regression to the mean (Buist, 1982; Dales et al. 1987). Therefore extreme values on one occasion are likely to be followed

by measurements closer to the mean at the next measurement occasion (Dockery et al., 1985). A negative relationship of initial value on its decline has been used to define regression to the mean (Burrows et al, 1981). Yet even with the *VETERANS* groups, whose initial value was actually the average of the up to twelve monthly measurements, most groups showed a significant negative coefficient relating the initial FEV₁ value and slope, indicating that higher initial FEV₁ values would show a steeper decline with time. Caution must be exercised in the interpretation of this relationship as it may well be a real phenomenon, and not merely a statistical artifact. Supportive evidence for this conclusion was the finding of significant coefficients of yearly percentage change in FEV₁ in each of the *VETERANS* groups. For a fixed percentage yearly change in FEV₁, a greater absolute level of decline would be found in those with higher initial values over the same period of time. Anthonisen et al. (1986) ignored the baseline FEV₁ in the calculation of FEV₁ decline in order to avoid the inevitable relationship of initial value with its slope (Oldham, 1975).

It is primarily the group population decline estimates that are being compared with one another in this thesis. Unlike many other investigators, no adjustment for initial FEV₁ value was done for the full covariate models. Burrows (personal communication, 1988) pointed out that any adjustment for initial value due to this regression to the mean effect, is only valid on homogeneous subgroups as otherwise, the estimated declines may be distorted. A study by Beck et al. (1982) is one of many in which an adjustment for initial lung function was performed on the entire data set, followed by a comparison of decline estimates for subgroups by type of work and smoking status. In this case, making an adjustment for initial FEV₁ value would exaggerate the differences in the slopes between the smokers and nonsmokers as the distribution of the initial values typically is found to differ. Another reason for not adjusting for the initial lung function value, is that it may be an additional effect of

the risk factor (Buist and Vollmer, 1988). As well, initial lung function is measured with random errors which distorts its own relationship with change; adjusting for its confounding effect could bias the group regression coefficient (Irwig et al, 1989).

Serial measurements of lung function on an individual typically display significant correlation; for example, Lebowitz et al. (1987) found a simple correlation for yearly FEV_1 values in adjacent measurements of children to be 0.77, decreasing to 0.53 for values three surveys apart. Higher correlations ranging from 0.89 to 0.98 were found for FEV_1 values from 1 to 3 years apart in the yearly interval *VETERANS* data. First order autocorrelation, as indicated by the Durbin-Watson test was noted in all three longitudinal data sets, based on residuals formed by a regression of all data points. While the estimates of decline using such a "cross-sectional" approach may not be biased, the significance of their comparison tends to be nonconservative (Neter et al., 1985). Based on a limited analysis of a subset of veterans with a minimum of fifteen measurements, the Durbin-Watson test for first order autocorrelation was significant in only a minority of individuals.

The GLS methods take account of autocorrelation. No statistically significant differences in FEV_1 decline could be found between the two *GRAIN* groups, whether using the unstructured or random effects GLS models; but this result also occurred with the OLS methods. For the *VETERANS*, the percentage decline and the overall cross-sectional method showed significant differences between the groups. This was not the case using in the random effects model, or any other OLS models. According to tentative findings of the completely unstructured model, a significant difference was found between Toronto and the other two cities.

The technique of regressing yearly differences between pairwise measurements on FEV_1 differences, is one method of accounting for autocorrelation (Neter et al, 1985).

With the *COAL* workers, this technique proved effective as significant differences were shown for pre- and post-retirement according to smoking group. This was not demonstrable using the overall "cross-sectional" regression method, although for the majority of analyses this produced the most significant differences.

A further complication, due to the autocorrelation of longitudinal data, is assessing goodness of fit of the model. The usual data exploratory analysis techniques rely on the cross-sectional nature of the data, in which each value is treated independently. Longitudinal residuals, on the other hand, are dependent on each individual, each measurement occasion, as well as risk factors (Cook and Ware, 1982). Further statistical developments in this area could prove most useful for exploratory data analysis.

Group Decline Characteristics

The groupings within each data set were all classified according to different conditions. For the *GRAIN* data, it was the type of work that separated the two major groups. In the *COAL* data, it was the groupings according to smoking and vital status that were compared. The *VETERANS* data group was distinguished by all subjects having symptoms of chronic bronchitis; the groupings were based on residence in one of three cities in Canada.

The average unadjusted estimates of FEV_1 decline for each group ranged from 5.8 ml/yr for the civic workers (based on a weighted least squares analysis using the inverse of the squared time term); to a high of 79.0 ml/yr found for the deceased coal miners who were nonsmokers, using a weighted average of unadjusted slopes. This range of FEV_1 decline encompasses roughly the lower two-thirds of the range of estimates found in the literature where published values of FEV_1 decline revealed a low estimate of 1.4 ml/yr for cotton operators (Fox et al., 1973) to a high value of

147 ml/yr found for toluene diisocyanate workers (Peters et al., 1970). What distinguishes these two extreme examples is the small period of follow-up, which was two years for the cotton operators and 1.5 years for the chemical workers; both were based on an endpoint analysis only. For these reasons, the reported magnitude of FEV₁ decline should not be accepted without qualification. For those studies in which the follow-up period averaged at least six years, the range of estimates in the literature generally reflected that found in the data sets analyzed. The lowest decline estimate of 4.6 ml/yr was found by Burrows et al. (1987) for asthmatic bronchitics, while the highest value of 78 ml/yr was found in a large group of smokers by Bosse et al. (1981). Within each group analyzed, the statistical methods produced similar estimates of FEV₁ decline, with the maximal discrepancy among the OLS methods due to the overall regression approach, which represents the crudest, more "cross-sectional" approach to longitudinal data analysis.

It is well documented in the literature that coal mine dust exposure is related to poor lung function (see Hurley and Soutar, 1986). Coal mine dust exposure has been implicated in chronic obstructive lung disease (Morgan et al., 1973) and it increases the risk of death from bronchitis and emphysema (Miller and Jacobsen, 1985). In the present analysis, the estimates of lung function decline found for the *COAL* data set were among the highest found. Cigarette smoking, rather than coal mine dust, might be considered to be the cause of respiratory impairment (Ames and Hall, 1985). Among the OLS methods FEV₁ decline estimates for deceased smokers were consistently significantly greater than that found for the alive nonsmokers. When unadjusted estimates were weighted by their precision, the declines of both the dead and alive smokers were greater than that of nonsmokers. Except for the overall "cross-sectional" regression method, which overly favours significance in the differences, no statistically significant differences were found between alive nonsmokers and alive smokers, although,

the trend was for a higher estimate of decline for the smokers group. Lyons and Campbell (1976) also performed a retrospective analysis of their longitudinal data on coal miners, collected over 15 years. For various categories of pneumoconiosis the estimates of deficits in FEV_1 above that expected due to aging, were greater in smokers than nonsmokers, but none of the differences were statistically significant. They concluded that the impairment of ventilatory capacity was not due to the effects of smoking and the smokers and nonsmokers were affected in equal measure by the pneumoconiosis. Love and Miller (1982) concluded that smoking had a comparable effect on FEV_1 decline to that of coal dust exposure when dust exposure was high; otherwise the smoking effect gained prominence. Although it cannot be measured, selection bias may be operative in many of the studies, as was indicated by Hurley and Soutar (1986).

In a two point survey of U.S. coalminers by Attfield (1984) the FEV_1 decline estimates for the coal miners who were over age 50 by the middle of the 9 year follow-up, was 58 ml/yr for the current smokers, and 45 ml/yr for the nonsmokers. These estimates are very similar to those found by the OLS methods in the present data set; the decline estimates for the alive smokers ranging from 52.7 to 57.7 ml/yr, while that of the alive nonsmokers ranged from 32.2 to 47.8 ml/yr. The published estimates of FEV_1 decline of this same group of miners (Bates et al., 1985) used a two point slope averaging technique on the *COAL* data set which contained at least two points of follow-up. With this technique, the slopes calculated for every two points of follow-up were weighted equally regardless of the number of years between the two points. The estimates were of the same rank order, and similar but slightly larger than the estimates found from the endpoint technique of method C.

The stated purposes of the original study of the Canadian *VETERANS* were two fold:

- 1) to observe the progression of chronic bronchitis over time; and
- 2) to compare lung function decline in cities with differing levels of atmospheric pollution. A relatively slow lung function decline would be expected to occur in the majority of chronic bronchitis patients at this stage, that is, at about the age of 40 when symptoms are present (Bates 1979). The range of unadjusted decline estimates found over all the methods used (36.6 ml/yr for Winnipeg to 95.7 ml/yr for Halifax) was slightly higher than the published estimates of FEV₁ decline among normal smokers (23 to 78 ml/yr), as reviewed in Chapter 3.

"Reducing type" of air pollution, as opposed to the ozone-based "oxidizing type" of air pollution, contains sulphur oxides, sulphuric acid and sulphate salts, as well as suspended particulates, primarily from fossil fuels. It has been shown that with increased levels of the reducing type of pollution an increased prevalence of respiratory disease is observed (Ferris, 1979). Both these types of pollution are found in Toronto; intermediate levels are observed in Halifax; but Winnipeg is much less polluted. Ishikawa et al. (1969) compared autopsy data from lungs of persons from Winnipeg with a similar sample from St Louis. They found that after controlling for cigarette smoking habits, the lungs from those that had lived in Winnipeg showed much less emphysema; this was attributed to the lower air pollution in Winnipeg.

Of the simple OLS methods used, the all points "cross-sectional" regression analysis and the percentage change approach showed a statistically significant difference in the unadjusted decline estimates, with Winnipeg showing the smallest decline. For the unstructured GLS model, the FEV₁ decline estimate for Toronto at 63 ml/yr was significantly higher than that found for the other two cities, although convergence was not achieved. It is important to note that the percentage of continuous smokers in Winnipeg (67.9%) was similar to that of Halifax (67.6%) and higher than that of Toronto (54.9%).

Comparison between three Australian cities by Campbell et al. (1985) showed a significantly smaller decline of FEV_1 for chronic bronchitics residing in an unpolluted region compared to a more polluted one after only 4 to 6 years of follow-up. Van der Lende et al. (1981) conducted a 9 year follow-up on two Dutch towns, and found among the middle-aged heavy smokers an average decline of 34.5 ml/yr in an unpolluted region in comparison to 40.2 ml/yr for the polluted one. These groups studied were not specifically chosen for having characteristics of chronic bronchitis, but instead were "normal" residents of the town. It was concluded that exposure to moderate air pollution levels caused a significantly greater decline in VC and FEV_1 with increasing age than was found among nonexposed individuals; however, no results of the significance testing was presented in any of the tables.

It has been repeatedly demonstrated that an increase in the prevalence of the indicators of chronic lung disease is a health effect associated with grain dust exposure (see Becklake, 1980). Many components of grain dust have the potential as respiratory irritants; grain dust consists of plant matter of various grains, antigens, insect parts, rat hair, mold spores and pesticides (Dosman 1980). In two recent studies matched controls were used and it was concluded that grain handling had an adverse effect on lung function that was of the same or smaller magnitude than that of smoking; however, these conclusions were both based on cross-sectional analyses (Cotton et al., 1983; Dopico et al., 1984).

Longitudinal change in lung function among grain workers was studied in a series of reports based on a larger sample of the *GRAIN* data set, at earlier stages in the study. Chan-Yeung et al. (1981) studied longitudinal changes in lung function of 396 grain workers and 111 civic workers who took part in both surveys over 2.5 years of follow-up. The annual decline in FEV_1 and MMF was found to be greater for grain workers than civic workers especially among the older age group (>50 years of age)

where the differences were statistically significant for both grain and civic nonsmokers and current smokers. The range of values found for FEV₁ change among nonsmokers was from +21.2 to -87.4 ml/yr; among current smokers the range was -16.0 to -78.1 ml/yr among the three different age groups. The interrelationships between dust loading of the lung, particularly if there is small airways involvement, and ageing, which results in progressive loss of recoil, is complex. This factor might be responsible for accelerated FEV₁ decline with ageing in those exposed to dust during their working lives. With three measurement occasions conducted over six years, two more studies were published on the *GRAIN* data set. Grain workers who attended all three measurement sessions and did not change their smoking habits showed the smallest declines (among exsmokers) of 17.2 ml/yr. Declines among the smokers (37.2 ml/yr) and nonsmokers (33.9 ml/yr) were similar (Tabona et al, 1985). Age and height adjusted FEV₁ decline estimates showed an increase in FEV₁ per year in the civic workers (Schulzer et al., 1985).

The three cited publications on the grain data set all have important distinguishing characteristics which make their estimates of decline not directly comparable to those in the present analysis. The numbers of workers in the two groups progressively decreased as follow-up was extended. Reducing the group to those with consistent smoking habits, reduces the representativeness of the total group. Decline estimates found at two measurement occasions over 3 years were generally much higher than those found at three points over 6 years; which are again higher than those estimated after four points of follow-up over a 9 year period. Without considering smoking status, the age-height adjusted decline estimates using the OLS methods were 17.6 to 20.8 ml/yr for the grain workers as opposed to 13.8 to 20.8 ml/yr for the civic workers. These declines are much lower than those found in previous publications. The published estimates all reveal increases in FEV₁ over time, particularly among the

civic nonsmokers and the younger age groups in all the smoking groups. With only 58 subjects forming the civic workers group, no age by smoking stratification was conducted. The unadjusted estimates of slope did show a strong gradient especially among the grain workers, where the older age group (> 50 years of age) decline estimates were much higher than that found for those less than 30 years of age. Individual regression coefficients showed a higher frequency of positive slopes for the younger age groups as opposed to the older ones, although there were many instances of cross-overs. Those showing increases in lung function were primarily never smokers or did not smoke during the follow-up. Apart from individual systematic responses attributable to a learning effect, this may indicate that FEV₁, rather than reaching a maximum at the mid-twenties may in fact peak at a later age, up to the mid-40s for healthy individuals who are nonsmokers or work in a clean environment.

The *GRAIN* data set consisted of only those individuals who attended the first and fourth measurement occasion, and one or both of the middle testing occasions. This "survivor" group of studied grain and civic workers may differ from those individuals who either left their place of employment or failed to return for repeated measurements. It has been demonstrated that workers who are most affected by grain dust exposure may subsequently leave the industry, and thus would not be studied in a population of workers with a longer employment history (Dosman et al., 1980). The original cohort were found to be older, on average, at baseline. Another selection process may be operating if asthmatics are discouraged from working in grain elevators (Cotton et al., 1983). For the thesis data set, information was not available on whether excluded individuals quit work or refused to attend the measurement session; as well no information on duration of employment or intensity of grain exposure was provided.

Another selection bias which is probably manifested in all data sets was the "healthy worker effect". The Veterans enrolled in the study had to be working or at

least, capable of employment. Working populations are generally considered to be more fit than those in the general population. Vogel described this process in 1885 (Miller and Jacobsen, 1985); "...miners are a body of picked men. No very weakly man is likely to take to the occupation; and ... many who become weakly have to abandon this form of labour for lighter work." The influence of this selection bias can not be readily measured, but its importance tends to diminish as follow-up proceeds in a longitudinal study (Fox and Collier, 1976). The *GRAIN* data set is thus more likely to be prone to this healthy worker effect; the generally low unadjusted estimates of 10.9 to 19.9 ml/yr found by the OLS methods are generally much lower than the decline estimates found among normal subjects. However, Burrows et al. (1987) found their smokers to have an average FEV₁ decline of 23.4 ml/yr while their nonsmokers showed a significantly smaller decline of 3.6 ml/yr. Chan-Yeung et al. (1989) attributed their finding of no differences in lung function decline between a group of aluminum smelter workers and controls, in part, to the healthy worker effect.

Despite the selection bias operating, results of the analysis of the three data sets are each generalizable; this is due to the relative normality of the initial lung function measures; the relatively large number of subjects studied; long duration of follow-up, which promotes stability in the estimates; and the comparability of these estimates with those found in other studies of similar groups. In addition, the *VETERANS* data was carefully planned so that the selection criteria would provide more comparable subpopulations to be selected from the participating cities. The average magnitude of FEV₁ decline for all *VETERANS* and *COAL* groups, as indicated by the majority of models, would be classified as "abnormal" based on the criteria of normality of decline being up to 40 ml/yr as suggested by Rosenstock and Cullen (1986).

The Effects of Baseline Characteristics on FEV₁ Decline

Apart from the grouping factors used in the methods comparison exercise, initial age was the most common significant baseline predictor, especially in the case of the *GRAIN* data set, where initial age was a significant covariate for each OLS method used. As was noted by Chan-Yeung et al. (1981) dependency of annual change in lung function on age, implies a quadratic relation of lung function level with age. The "cross-sectional" analysis of the *GRAIN* data produced a significant quadratic coefficient of decline; but with only four points of follow-up, individual based quadratic estimates were not calculated.

It has been stated by Cotes (1979) that decline in lung function is not properly established until after the age of 35. The inclusion of young men particularly in the grain workers group, resulted in many showing a rise in FEV₁ over the follow-up period. This behaviour was even evident for those subjects who were in the mid-30's at the beginning of follow-up, particularly among the civic workers. Flood et al. (1985) noted that the inclusion of young men in the analysis reduced the calculated FEV₁ loss.

Where the slope of FEV₁ was the dependent variable, age was a significant predictor for the *COAL* data set. Attfield (1984), in a longitudinal study of U.S. coal miners, found the 11 year decline estimates to be significantly related to age as well as to height. The rationale suggested for a significant association of height with FEV₁ decline is that tall persons with larger lungs could have larger changes in ml/yr for the same percentage loss of function, compared with short persons with smaller lungs (Beatty et al., 1984). Height was only found to be a significant predictor of the FEV₁ measurements as found through a "cross-sectional" analysis (Method A) in all data sets, where the level of FEV₁, rather than its slope, was the dependent variable.

In the *VETERANS* data, the initial values were an average of up to 12 monthly measurements in the first year. Thus the direct relationship of initial value of FEV_1 with FEV_1 change was not so obviously dependent upon regression to the mean. Despite this, there was a significant association overall of initial FEV_1 with FEV_1 decline for each data set.

Due to the close intercorrelation of the ventilatory flow measurements of FEV_1 , MMF and FVC (Burrows et al., 1965), the lung function parameters were assessed individually for their impact on FEV_1 decline. A different outcome was observed, depending upon the data set evaluated and whether absolute change (Method E) or Percentage change (Method D) of FEV_1 was used as the dependent variable. For the *COAL* data, FVC, unlike FEV_1 , was not a significant predictor of percentage change of FEV_1 yet it was a strong predictor of absolute FEV_1 decline. FCO, measured at some time during follow up, was predictive of both. Other than initial FEV_1 , VC a static volume, was the only significant predictor of absolute FEV_1 changes in the *VETERANS* data set. Conversely, VC was the only initial lung function variable not predictive of FEV_1 percentage change, while the initial measures of FEV_1 , MMF, RV, DLCO and FCO were. For the *GRAIN* data, MMF was the only significant spirometric variable associated with absolute FEV_1 decline, yet each spirometric variable was positively associated with percentage change in FEV_1 .

Barter et al. (1974) did not find a significant correlation between initial value of FEV_1 and the rate of deterioration of the FEV_1 , in their prospective study of mild chronic bronchitic patients. They suggested that this may be attributed to the group as a whole having well preserved lung function at the commencement of the study. In contrast, Beck et al. (1982) found initial lung function level to be highly significant, in all regressions of FEV_1 with time in their cotton textile workers. Pham et al. (1977) found initial FEV_1 , whether expressed in absolute terms, as percent predicted or as a

ratio of FEV_1/VC to be predictive of FEV_1 change; however residual volume and the fraction of carbon monoxide uptake coefficients were not significant. It can be concluded, based on the present results that the prediction of FEV_1 decline by a lung function variable taken on a single measurement occasion did not demonstrate "predictable" behavior, and was dependent on the which statistical approach was used.

The Influence of Follow-up Events on FEV_1 Decline

The variable defining length of follow-up (FUP) was not found to be significant for any of the *VETERANS* or *GRAIN* groups. With the grain data set the subjects all had completed the first and fourth measurements of follow-up. Thus, there was very little or no variation within or between the groups. For the *VETERANS* data, within each group there were differences in length of follow-up because of drop-out or deaths; however this variable was not a significant predictor. Although greater differences in follow-up were found between cities, length of follow-up added very little explanatory power to the model compared to the grouping factor.

Within each *COAL* group the largest variation of follow-up was observed; and length of follow-up was a significant predictor of FEV_1 level and decline, although the direction of its effect was not consistent. It has been shown (Berry, 1974; Lebowitz et al., 1987) that with increasing duration of follow-up, the standard error of the estimate of the individual regression slope becomes smaller, indicating a more precise estimate of decline. This impact on standard error of the estimate was much greater than that found for the actual number of measurements per individual for a given length of follow-up. An opposite relationship was observed in the *COAL* data, where a significant positive association was found between the total length of follow-up, as well as number of measurements, with the standard error of the estimate of the regression analysis. The same pairs of variables were not significantly correlated for the *VETERANS* data but

the same trend was found. The observation that greater precision in the decline estimate necessarily results from increasing the duration and number of measurements in the follow up, may not be applicable to data sets of long duration or which exhibit nonlinear decline tendencies.

For the *COAL* miners groups, it was the variable which measured the time between retirement and the last follow-up measurement that was found to be a significant factor in each of the groups. For the majority of coal workers, a significantly negative relationship was found for years since retirement and the FEV₁ measurements. Paradoxically, for alive nonsmokers a significant positive relationship was found, implying that a longer length time from retirement was related to higher FEV₁ values on average. Regression analysis of all values using dummy variables for retirement status showed significant differences in the levels of FEV₁ before and after retirement; the slopes however, in the entire group were not significant. When the change in FEV₁ over every two points was regressed on the change in years for that interval, a distinct difference was found in the slopes before and after retirement, depending on the smoking status of the coal miners. Nonsmokers showed an expected relationship where the pre-retirement rate of decline in FEV₁ decreased in the post-retirement period when the coal workers were no longer exposed to coal dust. Yet, in smokers the opposite pattern emerged as post-retirement declines in lung function showed an increase. Information on whether there were changes in the amount of smoking upon retirement was unavailable.

Studies concerning retirement have been conducted on cotton textile workers, as summarized by Beck and Schachter (1983). Workers who had retired from the mills were found to have similar or greater declines than those who remained actively working. No comparisons were made for individuals before and after their retirement. Beck et al. (1982) noted that smoking, in addition to cotton textile work, resulted in

significantly larger losses in lung function. They concluded that not only was the lung function loss and respiratory disability irreversible, but that such loss and disability may progress after exposure ceases. This observation on cotton textile workers appears to apply to this group of coal workers studied, with the qualification that this effect appears to be applicable only to the workers who smoke.

Smoking behaviour has been identified in previous publications on the *GRAIN* data, as being an important predictor of FEV_1 decline. In both the *GRAIN* and *VETERANS* data sets smoking behaviour was analyzed by the more pertinent classification of smoking behaviour during the follow-up period; as opposed to their baseline classification. Groups were thus divided into those that never smoked, those that did not smoke during the follow-up, those that showed intermittent smoking behaviour or stopped smoking at some time during follow-up; and those that continued smoking throughout the follow-up period. Whether analyzed by this four level classification system or as two groups (nonsmoker versus smoker groups), no significant differences in slopes were found between any of the smoker groups in the two data sets. This result is contrary to that of the expected relationship of nonsmokers showing a smaller decline than smokers (Wilhelmson et al., 1969; Huhti and Ikkala, 1980; Camilli et al., 1987 etc.). In concurrence with our results, Barter et al. (1974) found no relationship between the quantity of tobacco smoked and the rate of decline of FEV_1 , nor was there significant differences between the smoker and nonsmoker groups.

Howard (1970) noted that although the smokers and nonsmokers among their industrial workers showed symptomatic differences, there were no significant differences in their rate of decline of FEV_1 . They explained this observation by suggesting a multifactorial cause of FEV_1 decline, such that, if one factor alone were sufficient to cause a maximal response then it would not be possible to detect contributions from the other possible causes. In this case it was the heavy atmospheric pollution to which

all these men were exposed to that was suggested as the cause of the maximal response. Certainly, in the *GRAIN* study, responses to grain dust may have contributed to the non-association. As well, with the wide age range of individuals studied, there were a number of individual smokers and nonsmokers who had positive coefficients of FEV_1 change, which increases the standard deviation. Among the *VETERANS* the pollution levels in the cities of Toronto and Halifax were relatively high, at least in comparison to Winnipeg. As well, the smoker groups were similarly chosen as showing early symptoms of chronic bronchitis.

Further refinement of the analysis was attempted by assessing the effects of changing smoking behaviour over the follow-up period in terms of the time-dependent covariate. With this analysis, using the random effects covariate structure, the smoking variable was significantly positive, which implied that increasing categories of smoking were related to increasing FEV_1 levels, a counterintuitive result. However, for the Unstructured model, the expected relationship of a negative coefficient for smoking level was found. This finding was supported by the overall regression full model results for both data sets. The limited use of restricting the time-varying smoking behavior variable to an ordinal rather than continuous scale, compounded the problems of inaccuracy in exposure ascertainment and subsequent misclassification. As well, goodness-of-fit criteria of the two covariance structure models suggest that the Unstructured model was more suitable for the *GRAIN* data set.

For the *COAL* workers, smoking status did have a significant effect on average unadjusted FEV_1 decline estimates. Attfield's analysis of U.S. coal miners (1984) found the loss of FEV_1 over time to be significantly related to smoking, in that current smokers had 0.1 liter excess decline in their FEV_1 compared to those who had never smoked over the 11 years. This results in just under a 10 ml/yr difference in decline. For the *COAL* data, only the unadjusted FEV_1 decline estimates were significantly

different; they differed by 9 ml/yr between the two smoking categories.

Dontas et al. (1984) attributed a lack of a statistical smoking effect on FEV_1 decline on a possible bias due to the dead subjects having smoked more and having had lower lung volumes. This observation may have particular relevance to the *VETERANS* data set in that a greater percentage of the deceased in all the cities were current smokers and the initial FEV_1 level was lower for the deceased group. Of the *COAL* miners, those that died during follow-up showed a significantly higher decline of 68.9 ml/yr in comparison to the 53.9 ml/yr decline of the alive subjects. This observation has been supported in the literature; for example, Howard (1970) found that those who died during follow-up had a faster rate of FEV_1 decline. In the *VETERANS* data the various OLS methods showed contradictory results in that a lower estimate of FEV_1 decline for the deceased group was observed. For the endpoint estimate of slope (Method C) the comparison was significant, with the alive group having a greater decline of 61.9 ml/yr in comparison to 46.2 ml/yr estimate for the dead group. For those individuals who were lost to follow-up (any deaths recorded occurred at least two years after their last follow-up measurement), the decline estimates were intermediate between those found for the alive and deceased groups. This finding differed from that of Eisen et al, (1983;1984) where granite workers lost to follow-up showed greater declines of FEV_1 than those who remained unemployed or retired during the study period.

Differences between the study groups may explain these discrepancies. The *VETERANS* were an older mildly chronic bronchitic group, whose respiratory health was monitored over a long period of up to 30 years. One could speculate that those who had died may have had a decline of function that was more exponential in form so that a levelling off process was occurring. The resulting straight line slope estimates would be less for those who showed this form of decline.

Variability of Lung Function

In the OLS models where the unadjusted slope of FEV_1 change with time was used as a dependent variable, the proportion of the variance explained (R^2) for the full models reached a maximum of 19% when the percent change method was used for the *VETERANS* data; but most frequently the R^2 values varied from 9 to 12% of the variance explained. These R^2 estimates compare favourably to those found in other longitudinal studies. For example, Attfield (1984) found in their linear models of coal workers' FEV_1 decline, based on two measurements over 9 years, that even with the inclusion of such covariates as age, height, smoking status, mine effects and various aspects of the mine environment, the resulting maximum variance explained was only 12%. The mine effects were based on differences between 24 mines used in the analysis. With its exclusion, the variance explained dropped to 6%, in accordance with that found by Love and Miller (1982). Explanations put forward for the large unexplained variation in the decline of FEV_1 , include measurement error, and other aspects of within-subject variability as discussed in Chapters 2 and 4.

The full models used for the thesis data sets were kept simple for comparative purposes. The use of interactions or polynomial terms as well as the ratios of various terms, could perhaps have contributed to a larger proportion of the variance being explained. Adding more predictor variables would have had the same effect. With only the linear terms of age and height being included in the model, from 10 to 14% of the variance explained was accounted for in the *GRAIN* data set. For the *VETERANS* the most dramatic increases in R^2 were shown for the full model; but then the full model contained the most number of terms in comparison to the other data sets. The standard error of the estimate (SEE) remained consistently higher for the *VETERANS* data set, indicating that the average deviation of the individual values from the predicted straight line slope was greater than that in the other data sets.

One attempt at defining between-measurement variability in an individual's FEV_1 decline, was to use the coefficient of variation of the first three measurements (over 2 years) in the *VETERANS* data set. This variable was a significant negative predictor for both the FEV_1 measurements and the unadjusted decline estimates as dependent variables. Thus, for higher values of the predictor variable, the levels of FEV_1 were lower but the linear decline estimates were greater.

Within-subject variability has been evaluated by a number of investigators using different measures. For example, in order to control for spirometric variability, attempts have been made to set exclusion criteria, whereupon the two best forced expiratory volumes must be within 200ml of one another. A test of this criterion was conducted by Eisen et al. (1984), who found that those who failed the test criterion had a greater coefficient of decline of 81.1 ml/yr, compared to the 45.9 ml/yr decline of those who did not show such variability. Apart from the different sources of variability (within-occasion versus between-occasion), these results concurred with the findings that a measure of variability may be predictive of greater lung function decline with time.

Airway responsiveness, another source of variability in lung function, has been studied as a test of the Dutch hypothesis that smokers with chronic and obstructive lung disease shared with asthmatic patients an increased nonspecific bronchial activity. From the earlier work of Barter et al. (1974) to more recent examples (e.g. Taylor et al., 1985), increased bronchial hyperresponsiveness has been found in COLD patients as well as smokers. This was associated with lower initial levels of pulmonary function, as well as more rapid annual decline in FEV_1 , and concurs with the thesis finding using between-measurement criteria. Taylor et al. (1985) concluded that increased bronchial activity in smokers is associated with accelerated annual decline in FEV_1 , although no test of this acceleration was reported in the study; nor is there an indication of the age of the group of smokers and nonsmokers studied. Anthonisen et al. (1986) studied

a much older group whose average age was 60.9 years. Unlike the other studies, they measured pre-bronchodilator change in FEV_1 . In this instance, patients with large bronchodilator responses demonstrated a decreased annual rate of decline of FEV_1 . This result could be explained by a more allometric type of decline expected from this older group, which would yield smaller FEV_1 declines.

Other Lung Function Measures

Becklake and Permutt (1979) stated that none of the longitudinal studies of pulmonary function have used a sufficient number of different pulmonary function tests over a long enough period of time to rank their effectiveness in predicting outcome. In this regard, the *VETERANS* data has the most potential to evaluate lung dysfunction. The longitudinal decline of FEV_1 , in addition to that of VC, MMF, RV, FCO and resting diffusing capacity, can all be observed for up to 25 years of follow-up with the *VETERANS* data.

The relatively low coefficient of variation found for within-subject measurements of FEV_1 is often used as a major reason for its use as a longitudinal decline measurement. On that basis, FVC measurements often have as low or lower variation within-subjects (Bates, 1989). However, apart from the possibility of less measurement error, the vital capacity may not be as responsive as FEV_1 to change with time.

The estimates of FVC and FEV_1 decline for the *COAL* workers were similar in magnitude and the order of the group estimates was the same. The random effects analysis of FVC for the *GRAIN* data as a more appropriate analysis for longitudinal investigation, also showed the FVC estimates of decline to be almost the same as that of FEV_1 for the two groups. The parallel decline in FEV_1 and FVC in the occupationally exposed miners and grain workers was a similar finding to that of

Howard (1970) on a group of industrial workers and Soutar and Hurley (1986) on a study of British miners. This parallel decline suggested to the latter group of authors that the effects of dust induced lung damage are different from those due to smoking. The parallel reduction of FEV_1 and FVC was hypothesized to be indicating damage at the level of the respiratory bronchial or alveolus. In the *VETERANS* study, vital capacity was a static volume measurement; although Toronto and Halifax showed similar parallel declines with FEV_1 , which may be indicative of "restrictive" lung disease, Winnipeg was found to have a lower annual VC decline in contrast to FEV_1 , which may be more indicative of expected changes resulting from "obstructive" lung disease processes.

Tests of damage to the lung parenchyma were found in the *VETERANS* study, where diffusing capacity and fractional carbon monoxide uptake were measured, as well as residual volume. The initial normality of the DLCO and FCO data in this study was used to define the sample, since it was considered that this criterion would exclude men who had already developed significant emphysema. Significant declines of FCO were observed for all the *VETERANS* groups. Winnipeg was the only group that had a significant negative diffusing capacity coefficient and the highest significantly positive residual volume coefficient. These measures were so variable that consistent significant estimates of change were not shown. Where residual volume has been measured in studies of occupationally exposed subjects, an increase of the value with time has been shown (Hull et al., 1975; Coates et al., 1983; Diem et al., 1982). The latter investigators also showed a decrement in diffusing capacity with time, as did those investigators studying subjects with chronic obstructive lung disease (Earle, 1969; Jones et al., 1967).

The level of MMF is thought to be a more sensitive test of airways obstruction, than the other spirometric variables (Dosman, 1980; McFadden et al., 1972). The small

airways are a physiologically "quiet zone" of the lungs, since considerable increased in resistance can occur in them before the FEV_1 is affected. MMF had the highest coefficient of variation in both the *VETERANS* and *GRAIN* data sets. The MMF annual decline estimates, unlike that of FEV_1 and FVC, were relatively high. In the *VETERANS* data MMF did not decline in a manner parallel to FEV_1 ; instead, Winnipeg and Halifax stood out as having comparatively greater declines in MMF. Winnipeg, the least polluted city, had a high percentage of continuous smokers and higher intensity of smoking in comparison to Toronto. Halifax had both a high percentage of continuous smokers and a polluted environment.

Based on a literature review of the studies in which more than two measurements were made on each individual, FVC was found to be the most common lung function measurement for which change estimates are available; this is not surprising since both measurements are obtained from the same forced expiratory volume curve. Generally, the FVC decline estimates were similar in magnitude to that found for FEV_1 . For example, Hughes et al. (1972) found for their "emphysematous" exsmokers, a change of -63.5 ml/yr in their smokers FEV_1 , compared to -53.1 ml/yr found for FVC. The limitations of two point analysis over a short time interval are typified by an extreme example of disparity between the two measures. In one such study Graham et al. (1981) reported that a group of granite workers showed an increase in FVC of 108 ml/yr in comparison to a 6 ml/yr decline in FEV_1 .

It can be concluded that, along with FEV_1 , the decline in FVC and MMF have the potential to discern the progression of different types of lung disorders. The magnitude of the different estimates of decline as well as their relationship with each other, provide information as to the type and degree of damage which may be occurring and the consistency of the effect. The complementary use of these measures in describing longitudinal decline is a promising area for further investigation. Additional

research is needed on the nonlinear behaviour of all three measures and whether one measure tends to show an earlier onset of accelerating decline or a greater degree of decline than the other measures under various conditions of exposure.

Nonlinearity Aspects of Decline

As exemplified by the individual plots given throughout this thesis, an individual's change in FEV_1 over time can follow a linear, quadratic, or allometric curve or a combination of these curves, or conform to none of these models. A sudden acceleration or precipitous "fall" in FEV_1 has been described in individuals by Howard (1967; 1969; 1970). Individuals which showed this pattern did not differ from the rest of the group and it was concluded that this can be a feature of all stages of the natural history of obstructive airways disease. In an attempt to define this type of extreme change, a "step" decline has been proposed as a fall between two consecutive observations of greater than twice the standard deviation of the regression (Bates, 1973; Howard, 1967). However, Berry (1974) showed that such step declines are possible merely by chance, and an increase in lung function is just as likely to occur. Hruby and Butler (1975) concluded that the finding of a stepwise change may be just a reflection of the natural variability of function rather than a significant deterioration.

Specific cohorts have been shown in a number of different circumstances to have a decline that is not linear with time. For example, an acceleration in FEV_1 decline was shown in asbestos workers after 10 to 15 years from first exposure (Siracusa et al., 1984). Schulzer et al.(1975) also emphasized a quadratic decline among grain exposed and smoking workers; Emergil and Sobol (1971) described an "exponential" analysis of bronchitic patients and showed significant nonlinear declines, with the percentage decline per year ranging from 6.3 to 8.7%.

Howard (1974) draws on individual observations to describe three phases in the process of chronic airways obstruction: the first phase is relatively unchanging FEV₁ values, followed by a second phase of linear decline in FEV₁ values until the third phase where a plateau in the response is shown, at about 0.75 liters of FEV₁. Postma et al. (1979) described various patients in their study as being representative of each of the three different patterns of decline. However the criteria used to distinguish linear and quadratic declines, R², is a poor discriminator, as examples of quadratic decline in the thesis data were found to have relatively high R² values.

In the present data sets, it was shown that individuals could have a significant linear decline in FEV₁, or a significant allometric decline, or a significant exponential decline, or a combination of these, over the entire follow-up period. It is possible that each of these analyses could be more important for different stages of aging and exposure. Over the entire adult span of life a smoothed curve for an individual could perhaps be described as a reversed sigmoidal shape, which would be more appropriately modelled by a logistic or Weibull function (Lawless, 1982). In adulthood after reaching their maxima, the FEV₁ values would show a gradual fall which becomes linear or even accelerates in a quadratic form, until the rate of decline lessens and a plateau is reached, with the achievement of a minimal volume of approximately 0.75 liters. Further investigation of the nonlinearity aspects of lung function decline is warranted to better model longitudinal changes with time.

Because longitudinal follow-up is conducted on a cross-sectional group of individuals, of various ages and each having various lifetime exposures and different constitutions, the population estimate of decline for these individuals would average the individual declines at various stages in their lives, therefore distorting the different patterns of individual decline. A more homogeneous group would be expected to show less variability for the chosen method of decline analysis. An evaluation of nonlinear

decline appears to have particular relevance to the analysis of data sets covering extended periods of observation of older populations and for those exposed to contaminated air, whether by smoking or from occupational sources. Distinguishing the type of decline, the timing and potential causes, would be of further interest for predicting disability due to lung function decline.

Conclusion

The use of different models in this exercise has emphasized the statement given by Tukey and Wilks back in 1965, who stated that "models must be used but must never be believed" (page 372). Two principles of modelling were stated by McCullagh and Nelder (1982):

"All models are wrong (although some are better than others)."

"Do not fall in love with one model, to the exclusion of alternatives."

Adding sophistication to the models may provide more information, but the simpler and more obvious approaches must not be neglected. A pertinent observation was made by Ware (1988) in summarizing the findings of a statistical conference on longitudinal analysis. He emphasized that as part of the progress in the growing sophistication of longitudinal models in the epidemiological setting, the most difficult part of a successful analysis may be the formulation of the right question in terms of the appropriate mathematical model.

Posing the following simple question, "Does group X show a greater average decline than group Y?", offers very little insight without elaboration of, first, of what is meant by "decline", second, a consideration of the length of time over which the decline coefficient is estimated, and third, an evaluation of the characteristics of each of the groups being analyzed.

How the decline in lung function with aging is described is very much a function of the type of model chosen to describe decline. The usual approach of describing decline of lung function, as a linear slope of change in FEV_1 with time, has the merit of being conceptionally simple. Use of quadratic and allometric coefficients are alternative approaches to describing change; they are still based upon linear models but describe either accelerating or decelerating curves respectively. The coefficient for the allometric model is interpreted as a percent change of the lung function measure with percent change in time. It is conceivable that two different groups would have the same linear decline with time but show contrasting curvature which would have relevance to the biological interpretation of the results.

The protocol of an ideal longitudinal study of pulmonary function decline with aging should take into account a number of different design factors. An attempt to ensure full compliance from all of the participants at each common measurement occasion would decrease the likelihood of imbalance in the data structure. Choosing the length of the longitudinal study depends on the question that is being posed, the age structure of the cohort, the ease of follow-up, expense considerations, and what type of lung function change one is expected to model. For example, a linear change of FEV_1 with time could be estimated with precision after only six years of follow-up. A more variable measure, such as maximal midexpiratory flow rate, would need further follow-up before the standard error between-individuals exceeded that within-individuals.

The analysis of longitudinal data should be preceded by a combination of exploratory and residual analysis techniques in order to evaluate the appropriateness of the chosen methods of analysis. The more "cross-sectional" approaches of using an overall regression of the data or its differences, can only serve as a preliminary evaluation of the order of the group estimates of decline. These methods cannot account for the longitudinal nature of the data, as between-individual variability is

ignored. While their estimated coefficients of decline were not greatly disparate from those of the longitudinal methods, inferences concerning the significance of the differences were nonconservative and misleading. However, for the analysis of severely imbalanced data, as was encountered in the *COAL* data, the differencing approach was particularly relevant to the profile analysis applications. Longitudinal modelling of the linear relationship of decline of lung function with time is most simply calculated by the end-point calculation of slope; but unless there is very little variability in the individual linear declines, the slope estimate obtained may be inaccurate. Yet the averaged estimates obtained by this simple approach were generally congruent with the findings from the individual regression analysis. A longitudinal OLS procedure in common use which takes into account all the points of follow-up on an individual basis, is the averaging of each individual's coefficient obtained from a regression analysis. Without access to software allowing generalized least squares analysis, a refinement to the regression procedure is to weight the slopes by the reciprocal of the standard error of the estimate or some other measure of variability. The more sophisticated approaches of generalized least squares, can be used to model the covariance structure between the measures. Of these, the random effects model has the further advantages of partitioning the sources of variability by describing the covariance structure in terms of the slope and intercept variances, their covariance, and within-subject variance. Due to the limitations of available software, the iterative techniques used for covariance modelling of unbalanced data each failed considered in this thesis were limited to longitudinal studies which are generally of shorter duration and less frequent in follow-up. Use of the more sophisticated statistical methods, such as the use of robust variance in a general estimating equation setting, use of Bayesian inference as well as further refinements in nonparametric modelling and boot-strap techniques, await further statistical developments, or application to real data. Pertinent issues which remain unresolved for longitudinal data analysis include nonlinear modelling

and the handling of "nonignorable" missing data. In addition, there is a definite need for the more advanced methods of statistical analysis of longitudinal data to be made available to epidemiologists in an understandable form that can be implemented without requiring intensive programming.

Longitudinal analysis is not an intuitively obvious procedure, particularly in the usual situation where the data has an unbalanced structure due to missing observations. Using multiple methods of analysis can provide evidence of consistency of effect and allow a firm basis for making conclusions of statistical significance and order of group comparisons. Along with FEV_1 decline, further information might result from analyses of the decline of forced vital capacity, maximal midexpiratory flow, and other lung function variables. Disentangling the biological explanation of differential rates of FEV_1 decline may well have to await the use of other criteria. It is of interest that Biernacki et al. (1989) have recently reported longitudinal concordance between the decline of diffusing capacity and the progressive loss of lung density as measured by a CAT scanner. If both of these methods can be taken to be measuring the progression of lung destruction ("emphysema"), then they could be compared to the synchronous rate of decline of the FEV_1 . Such methods may elucidate the complex factors affecting the rate of decline of FEV_1 , which might be described as a deceptively simple index.

REFERENCES.

- Allen GW, Sabin S. Comparison of direct and indirect measurement of airway resistance. *Am Rev Respir Dis* 1971;104:61-71.
- American Thoracic Society. Screening for adult respiratory disease. Official ATS statement. *Am Rev Respir Dis* 1983; 128:768-73.
- Ames RG, Trent RB. Respiratory impairment and symptoms as predictors of early retirement with disability in US underground coal miners. *Am J Pub Health* 1984;74:837-838.
- Anthonisen NR, Wright EC, Hodgkin JE, IPPB Trial Group. Prognosis in chronic obstructive pulmonary disease. *Am Rev Respir Dis* 1986;133:14-20.
- Asemota AE. The longitudinal method in mortality studies. In: Goldfarb N. ed. Applications of longitudinal research methods in fields of business, social sciences and biology. Hofstra University Yearbook of Business, 1979; 399-433.
- Ashford JR, Brown S, Morgan DC, et al. The pulmonary ventilatory function of coal miners in the United Kingdom. *Amer Rev Respir Dis* 1968; 97: 810-826.
- Attfield MD. Longitudinal decline in FEV1 in United States coalminers. *Thorax* 1985;40:132-137.
- Bailar JC, Mosteller F. Medical Uses of Statistics. Massachusetts: Nejm, 1986.
- Baltes PB, Nesselroade JR. History and rationale of longitudinal research. In: Nesselroade JG, Baltes PB. eds. Longitudinal research in the study of behavior and development. New York: Academic Press, 1979; 1-39.
- Bande J, Clement J, Van de Woestijne KP. The influence of smoking habits and body weight on VC and FEV1 in male air force personnel: a longitudinal and cross-sectional analysis. *Am Rev Respir Dis* 1980;122:781-790.
- Barter CE, Campbell AH, Tandon MK. Factors affecting the decline of FEV1 in chronic bronchitis. *Aust NZ J Med* 1974;4:339-345.
- Barter CE, Campbell AH. Relationship of constitutional factors and cigarette smoking to decrease in 1-second forced expiratory volume. *Am Rev Respir Dis* 1976;113:305-314.
- Bates DV. The fate of the chronic bronchitic: A report of the ten-year follow-up in the Canadian Dept. of Veteran's Affairs coordinated study of chronic bronchitis. *Am Rev Respir Dis* 1973;108:1043-1065.
- Bates DV, Pham QT, Chau N, Pivoteau C, Dechoux J, Sadoul P. A longitudinal study of pulmonary function in coal miners in Lorraine, France. *Am J Ind Med* 1985;8:21-32.
- Bates DV, Woolf CR, Paul GI. Chronic bronchitis. A report on the first two stages of the co-ordinated study of chronic bronchitis in the Department of Veterans Affairs, Canada. *Medical Services Journal, Canada* 1962;18:211-303.

- Bates DV, Gordon CA, Paul GI, Place REG, Snidal DP, Woolf CR. Chronic bronchitis. Report on the third and fourth stages of the co-ordinated study of chronic bronchitis in the Department of Veterans affairs, Canada. Medical Services Journal Canada 1966;22:5-59.
- Bates DV. The prevention of emphysema. Chest 1974;65:437-441.
- Bates DV. Chronic bronchitis and emphysema. The search for their natural history. In: Macklem PT, Permutt S, Eds. The Lung in the Transition Between Health and Disease. New York:Marcel Dekker, Inc. Ch. 1.
- Bates DV. Respiratory Function in Disease. 3rd Ed. Philadelphia: W.B. Saunders Company, 1989.
- Bates DV, Knott JMS, Christie RV. Respiratory function in emphysema in relation to prognosis. The Quarterly J Med 1956; 25:137-157.
- Beale EML, Little RJA. Missing values in multivariate analysis. J Royal Stat Soc Series B 1975; 37: 129-146.
- Beaty TH, Menkes HA, Cohen BH, Newill CA. Risk factors associated with longitudinal change in pulmonary function. Am Rev Respir Dis 1984;129:660-667.
- Beck GJ, Schachter EN, Maunder LR, Schilling SF. A prospective study of chronic lung disease in cotton textile workers. Ann Intern Med 1982;97:645-651.
- Beck GJ, Schachter EN. The evidence for chronic lung disease in cotton textile workers. The American Statistician 1983;37:404-412.
- Becklake MR. Grain dust and health: state of the art. In: Dosman JA, Cotton DJ, eds. Occupational pulmonary disease. Focus on grain dust and health. New York: Academic Press, 1980; 189-200.
- Becklake MR, Permutt S. Evaluation of tests of lung function for "screening" for early detection of chronic obstructive lung disease. In: Macklem Pt, Permutt S, Eds. The Lung in the Transition Between Health and Disease. New York:Marcel Dekker, Inc. Ch.16.
- Becklake MR. Concepts of normality applied to the measurement of lung function. Am J Med 1986; 80:1158-1164.
- Becklake MR. Epidemiologic studies in human populations. In: Witschi HP and Brain JD (Eds) Toxicology of Inhaled Materials, Vol.1. Berlin: Springer-Verlag, 1984.
- Berenson ML, Levine DM, Goldstein M. Intermediate Statistical Methods and Applications A Computer Package Approach. New Jersey: Prentice-Hall, 1983.
- Berk K. Computing for incomplete repeated measures. Biometrics 1987;43:385-398.
- Berkey CS, Ware JH, Dockery DW, Ferris Jr. BG, Speizer FE. Indoor air pollution and pulmonary function growth in preadolescent children. Am J Epidemiol 1986;123:250-260.
- Berry G, McKerrow CB, Molyneux MKB, Rossiter CE, Tombleson JBL. A study of the acute

- and chronic changes in ventilatory capacity of workers in Lancashire cotton mills. *Brit J Ind Med* 1973;30:25-36.
- Berry G. Step decline in lung function. Letter to the editor. *Am Rev Respir Dis* 1974; 110: 203.
- Biernacki, W, Ryan M, MacNee W, Flenley DC. Can the quantitative CT scan detect progression of emphysema? *Am Rev Respir Dis* 1989; 149:A120.
- BMDP. BMDP Statistical Software. Berkeley: University of California Press, 1988.
- Bosse R, Sparrow D, Rose CL, Weiss ST. Longitudinal effect of age and smoking cessation on pulmonary function. *Am Rev Respir Dis* 1981;123:378-381.
- Bosse R, Sparrow D, Garvey AJ, Costa PT, Weiss ST, Rowe JW. Cigarette smoking, aging, and decline in pulmonary function. *Arch Environ Health* 1980;35:247-252.
- Bouhuys A, Zuskin E. Chronic respiratory disease in hemp workers. A follow-up study, 1967-1974. *Ann Intern Med* 1976;84:398-405.
- Brinkman GL, Block DL, Cress C. Effects of chronic bronchitis and occupation on pulmonary ventilation over an 11-year period. *JOM* 1972;14:615-620.
- Brown LK. Static lung volumes: In: Miller A Ed. *Pulmonary Function Tests in Clinical and Occupational Lung Disease*. Orlando, Grune & Stratton, 1986.
- Bryant E, Gillings D. Statistical analysis of longitudinal repeated measures designs. In: Sen PK. Ed. *Biostatistics: Statics in Biomedical, Public Health and Environmental Sciences*. Elsevier Science Pub., 1985.
- Bucca C, Avolio G, Rolla G, Maina A, Bugiani M, Arossa W, Spinaci S, Cacciabue M. A long term follow-up of patients with silicosis. *Bull Eur Physiopathol Respir* 1985;21:8A-9A.
- Buist AS, Vollmer WM, Johnson LR, Bernstein RS, McCamant LE. A four-year prospective study of the respiratory effects of volcanic ash from Mt. St. Helens. *Am Rev Respir Dis* 1986;133:526-534.
- Buist AS. Evaluation of lung function: concepts of normality. In: Simmons DH ed. *Pulmonology, Volume 4*. New York: John Wiley & Sons, Inc, 1986.
- Buist AS, Vollmer WM. The use of lung function tests in identifying factors that affect lung growth and aging. *Statistics in Medicine* 1988; 7:11-18.
- Buist AS, Vollmer WM. The use of lung function tests in identifying factors that affect lung growth and aging. *Stat Med* 1988; 7: 11-18.
- Burrows B, Cline MG, Knudson RJ, Taussig LM, Lebowitz MD. A descriptive analysis of the growth and decline of the FVC and FEV1. *Chest* 1983;83:717-724.
- Burrows B, Lebowitz MD, Camilli AE, Knudson RJ. Longitudinal changes in forced expiratory volume in one second in adults. Methodologic considerations and findings in healthy nonsmokers. *Am Rev Respir Dis* 1986;133:974-980.

- Burrows B, Earle RH. Course and prognosis of chronic obstructive lung disease. A prospective study of 200 patients. *New Engl J Med* 1969;280:397-404.
- Burrows B, Knudson RJ, Camilli AE, Lyle SK, Lebowitz MD. The "horse-racing effect" and predicting decline in forced expiratory volume in one second from screening spirometry. *Am Rev Respir Dis* 1987;136: 69-75.
- Burrows B, Strauss RH, Niden AH. Chronic obstructive lung disease. III. Interrelationships of pulmonary function data. *Am Rev Respir Dis* 1965 p 861-868.
- Burrows B, Bloom JW, Traver GA, Cline MG. The course and prognosis of different forms of chronic airways obstruction in a sample from the general population. *New Eng J Med* 1987;317:1309-1314.
- Camilli AE, Burrows B, Knudson RJ, Lyle SK, Lebowitz MD. Longitudinal changes in forced expiratory volume in one second in adults. Effects of smoking and smoking cessation. *Am Rev Respir Dis* 1987;135:794-799.
- Campbell AH, Barter CE, O'Connell JM, Huggins R. Factors affecting the decline of ventilatory function in chronic bronchitis. *Thorax* 1985;40:741-748.
- Causton, DR. *A Biologist's Basic Mathematics*. Edward Arnold, 1983.
- Chan-Yeung M, Schulzer M, MacLean L, Dorken E, Tan F, Lam S, Enarson D, Grzybowski S. A follow-up study of the grain elevator workers in the port of Vancouver. *Arch Environ Hlth* 1981;36:75-81.
- Chan-Yeung M, Schulzer M, MacLean L, Dorken E, Grzybowski S. Epidemiologic health survey of grain elevator workers in British Columbia. *Am Rev Respir Dis* 1980; 121:329-38.
- Chan-Yeung M, Enarson DA, MacLean L, Irving D. Longitudinal study of workers in an aluminum smelter. *Arch Env Hlth* 1989; 44: 134-139.
- Clement J, Van de Woestijne KP. Rapidly decreasing FEV or VC and development of chronic airflow obstruction. *Am Rev Respir Dis* 1982;125:553-558.
- Cochrane GM, Prieto F, Clark TJH. Intrasubject variability of maximal expiratory flow volume curve. *Thorax* 1977;32:171-6.
- Cochrane GM, Prieto F, Clark JH. Intrasubject variability of maximal expiratory flow volume curve. *Thorax* 1977; 32:171-6.
- Cole JW, Grizzle JE. Applications of multivariate analysis of variance to repeated measurements experiments. *Biometrics* 1966; 22: 810-828.
- Cole TJ. Linear and proportional regression models in the prediction of ventilatory function. *J R Statist Soc* 1975; 138: 297-338.
- Colquhoun D. *Lectures on Biostatistics*. Oxford: Clarendon Press, 1971.
- Cook NR, Ware JH. Design and analysis methods for longitudinal research. *Ann Rev Public Health* 1983;4:1-23.

- Cook NR. A general linear model approach to longitudinal data analyses. Doctoral dissertation. Biostatistics. Harvard School of Public Health, 1982.
- Corbin RP, Loveland M, Martin RR, Macklem PT. A four-year follow-up study of lung mechanics in smokers. *Am Rev Respir Dis* 1979;120:293-304.
- Cotes JE, Gilson JC, McKerrow CB, Oldham PD. A long-term follow-up of workers exposed to beryllium. *Br J Ind Med* 1983;40:13-21.
- Cotes JE. Lung Function Assessment and Application in Medicine. 3rd Ed. London: Blackwell Scientific Publ., 1979.
- Cotton DJ, Graham BL, Li KYR, Froh F, Barnett GD, Dosman JA. Effects of grain dust exposure and smoking on respiratory symptoms and lung function. *J Occup Med* 1983; 25:131-41.
- Dales RE, Hanley JA, Ernst P, Becklake MR. Computer modelling of measurement error in longitudinal lung function data. *J Chron Dis* 1987;40:769-773.
- Daniel C, Wood FS. Fitting Equations to Data Computer Analysis of Multifactor Data, 2nd ed. New York: John Wiley & Sons, 1980.
- Dawson A. Spirometry. In: Wilson AF. Pulmonary function testing indications and interpretations. Orlando: Grune & Stratton, 1985;9-31.
- Dechoux J, Pivoteau C, Wantz JM. Troubles fonctionnels respiratoires des mineurs des houillères du bassin de Lorraine à l'heure de la retraite. *Bull Eur Physiopath Respir* 1983;19:385-91.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc B* 1977;39:1-38.
- Diem JE. A statistical assessment of the scientific evidence relating cotton dust exposure to chronic lung disease. *The American Statistician* 1983;37:395-403.
- Diem JE, Jones RN, Hendrick DJ, Glindmeyer HW, Dharmarajan V, Butcher BT, Salvaggio JE, Weill H. Five-year longitudinal study of workers employed in a new toluene diisocyanate manufacturing plant. *Am Rev Respir Dis* 1982;126:420-428.
- Diem JE, Liukkonen JR. A comparative study of three methods for analyzing longitudinal pulmonary function data. Presented at the Workshop of Methods for Longitudinal Data Analysis in Epidemiological Clinical Studies, Bethesda, Maryland, 1986.
- Diem JE, Liukkonen JR. A comparative study of three methods for analysing longitudinal pulmonary function data. *Stat Med* 1988; 7: 19-28.
- Dockery DW, Ware JH, Ferris Jr BG, Glicksberg DS, Fay ME, Spiro III A, Speizer FE. Distribution of forced expiratory volume in one second and forced vital capacity in healthy, white, adult never-smokers in six U.S. cities. *Am Rev Respir Dis* 1985; 131:511-20.
- Donner A. Linear regression analysis with repeated measurements. *J Chron Dis* 1984;37:441-448.

- Dontas AS, Jacobs DR, Corcondilas A, Keys A, Hannan P. Longitudinal versus cross-sectional vital capacity changes and affecting factors. *J Gerontology* 1984;39:430-438.
- Dopico GA, Reddan W, Tsiatis A, Peters ME, Rankin J. Epidemiologic study of clinical and physiologic parameters in grain handlers of northern United States. *Am Rev Respir Dis* 1984; 130:759-65.
- Dosman JA. Chronic obstructive pulmonary disease and smoking in grain workers. In: Dosman JA, Cotton DJ, eds. *Occupational pulmonary disease. Focus on grain dust and health*. New York: Academic Press, 1980; 201-206.
- Eason G, Coles CW, Gettinby G. *Mathematics and Statistics for the Bio-sciences*. New York: John Wiley & Sons, 1980.
- Eisen EA, Robins JM, Greaves IA, Wegman DH. Selection effects of repeatability criteria applied to lung spirometry. *Am J Epidemiol* 1984;120:734-742.
- Eisen EA, Oliver LC, Christiani DC, Robins JM, Wegman DH. Effects of spirometry standards in two occupational cohorts. *Am Rev Respir Dis* 1985;132:120-124.
- Emirgil C, Sobol BJ, Varble A, Waldie J, Weinheimer B. Long-term course of chronic obstructive pulmonary disease. A new view of the mode of functional deterioration. *Am J Med* 1971;51:504-512.
- Enarson DA, Vedal S, Chan-Yeung M. Rapid decline in FEV1 in grain handlers. Relation to level of dust exposure. *Am Rev Respir Dis* 1985;132:814-817.
- Eriksson S, Lindell SE, Wiberg R. Effects of smoking and intermediate α -antitrypsin deficiency (PiMZ) on lung function. *Eur J Respir Dis* 1985;67:279-285.
- Fabsitz R, Feinleib M, Hubert H. Regression analysis of data with correlated errors: an example from the NHLBI twin study. *J Chron Dis* 1985;38:165-170.
- Fairclough L, Helms RW. Mixed effects model analyses of incomplete longitudinal pulmonary function measurements in children. University of North Carolina at Chapel Hill: Institute of Statistics Mimeo Series No. 1470, October 1984.
- Ferris Jr. BG, Puleo S, Chen HY. Mortality and morbidity in a pulp and a paper mill in the United States: a ten-year follow-up. *Br J Ind Med* 1979;36:127-134.
- Ferris Jr B.G. Air pollution. In: Macklem PT, Permutt S, Eds. *The Lung in the Transition Between Health and Disease*. New York: Marcel Dekker, Inc., 1979.
- Fletcher C, Peto R, Tinker C, Speizer FE. *The Natural History of Chronic Bronchitis and Emphysema*. Oxford:Oxford University Press, 1976.
- Fletcher CM, Oldham PD. Value of chemoprophylaxis and chemotherapy in early chronic bronchitis. *Brit Med J* 1966;1:1317-1322.
- Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Brit Med J* 1977;1:1645-8.
- Flood DFS, Blofeld RE, Bruce CF, Hewitt JJ, Juniper CP, Roberts DM. Lung function,

- atopy, specific hypersensitivity and smoking of workers in the enzyme detergent industry over 11 years. *Br J Ind Med* 1985;42:43-50.
- Fox AJ, Tombleson JBL, Watt A, Wilkie AG. A survey of respiratory disease in cotton operatives. Part 1. Symptoms and ventilation test results. *Br J Ind Med* 1973;30:42-47.
- Fox AJ, Collier PF. Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. *Brit J Prev Soc Med* 1976; 30:225-30.
- Glindmeyer HW, Diem JE, Jones RN, Weill H. Noncomparability of longitudinally and cross-sectionally determined annual change in spirometry. *Am Rev Respir Dis* 1982;125:544-548.
- Gold HJ. *Mathematical Modeling of Biological Systems - An Introductory Guidebook*. New York: John Wiley & Sons, 1977.
- Goldstein H. *The design and analysis of longitudinal studies. Their role in the measurement of change*. London: Academic Press, 1979.
- Goldstein H. Efficient statistical modelling of longitudinal data. *Ann Human Biol* 1986;13:129-141.
- Graham WGB, O'Grady RV, Dubuc B. Pulmonary function loss in Vermont granite workers. A long-term follow-up and critical reappraisal. *Am Rev Respir Dis* 1981;123:25-28.
- Greaves IA, Colebatch HJH. Observations on the pathogenesis of chronic airflow obstruction in smokers: implications for the detection of "early" lung disease. *Thorax* 1986;41:81-87.
- Greenland S. Modeling and variable selection in epidemiological analysis. *Am J Publ Hlth* 1989; 79: 340-349.
- Grizzle JE, Allen DM. Analysis of growth and dose-response curves. *Biometrics* 1969; 25: 357-381.
- Guire KE, Kowalski CJ. Mathematical description and representation of developmental change function on the intra- and interindividual levels. In: Nesselroade JG, Baltes PB. eds. *Longitudinal research in the study of behavior and development*. New York: Academic Press, 1979; 89-110.
- Hall DR, Lapp NL, Reger R, Seaton A. Small airways disease in coal miners. A longitudinal study. *Bull Physiopathol respir* 1975;11:863-877.
- Hanushek EA, Jackson JE. *Statistical Methods for Social Scientists*. New York: Academic Press, 1977.
- Healy MJR, Goldstein H. Regression to the mean. *Ann Hum Biol* 1978; 5:277-80.
- Heederik D, Brunekreef B. The occurrence and analysis of horse racing in longitudinal studies. *Bull Eur Physiopathol Respir* 1985;21:2A-3A.
- Higgins ITT. Relative risks of various tobacco usages for emphysema and/or chronic bronchitis. In: *Smoking and Health I. Modifying the risk for the smoker*. Washington: American Cancer Society, 1976.

- Hills M. *Statistics for Comparative Studies*. London: Chapman and Hall, 1974.
- Howard P. A long-term follow-up of respiratory symptoms and ventilatory function in a group of working men. *Br J Ind Med* 1970;27:326-333.
- Howard P. The changing face of chronic bronchitis with airways obstruction. *Br Med J* 1974;2:89-93.
- Howard P, Astin TW. Precipitous fall of the forced expiratory volume. *Thorax* 1969;24:492-5.
- Howard P. Evolution of the ventilatory capacity in chronic bronchitis. *Br Med J* 1967;3:392-6.
- Hruby J, Butler J. Variability of routine pulmonary function tests. *Thorax* 1975;30:548-553.
- Hughes JA, Hutchison DCS, Bellamy D, Dowd DE, Ryhan KC, Hugh-Jones P. The influence of cigarette smoking and its withdrawal on the annual change of lung function in pulmonary emphysema. *Quarterly J Med* 1982;51:115-124.
- Huhti E, Ikkala J. A 10-year follow up study of respiratory symptoms and ventilatory function in a middle-aged rural population. *Eur J Respir Dis* 1980;61:33-45.
- Hui SL. Curve fitting for repeated measurements made at irregular time-points. *Biometrics* 1984;40:691-697.
- Hui SL, Berger JO. Empirical bayes estimation of rates in longitudinal studies. *J Am Stat Assoc* 1983; 78:753-60.
- Hurley JF. Longitudinal studies of lung function in occupational groups: can we trust the answers? *Bull Eur Physiopathol Respir* 1985;21:1A-2A.
- Hurley JF, Soutar CA. Can exposure to coalmine dust cause a severe impairment of lung function? *Br J Ind Med* 1986;43:150-157.
- Irwig LM, Groeneveld HT, Becklake MR. Assessing the effect of exposure on lung function loss between two occasions: issues of confounding and measurement error. In: Hensley MJ, Saunders NA (Eds). *Clinical Epidemiology of Chronic Obstructive Pulmonary Disease*. New York: Marcel Dekker, Inc, 1989.
- Ishikawa S, Bowden DH, Fisher V, Wyatt JP. The "emphysema profile" in two mid-western cities in North America. *Arch Environ Health* 1969; 18:660-666.
- Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 1986;42:805-820.
- Johnston RN, McNeill RS, Smith DH, Legge JS, Fletcher F. Chronic bronchitis - measurements and observations over 10 years. *Thorax* 1976;31:25-29.
- Jones NL, Burrows B, Fletcher CM. Serial studies of 100 patients with chronic airway obstruction in London and Chicago. *Thorax* 1967;31:25-39.
- Kanner RE, Renzetti AD, Klauber MR, Smnith CB, Golden CA. Variables associated with changes in spirometry in patients with obstructive lung diseases. *Am J Med* 1979;67:44-50.

- Kanner RE, Renzetti Jr. AD, Klauber MR, Smith CB, Golden CA. Variables associated with changes in spirometry in patients with obstructive lung diseases. *Am J Med* 1979;67:44-50.
- Kauffmann F, Drovet D, Lellough J, Brille D. Twelve years spirometric changes among Paris area workers. *Int J Epidemiol* 1979;8:201-212.
- Kellie SE, Attfield MD, Hankinson JL, Castellan RM. Spirometry variability criteria - association with respiratory morbidity and mortality in a cohort of coal miners. *Am J Epidemiol* 1987;125:437-444.
- Kleinbaum DG, Kupper LL, Muller KE. *Applied Regression Analysis and Other Multivariable Methods*. Boston: PWS-Kent Publ Co., 1988.
- Knudson RJ, Slatin RC, Lebowitz MD, Burrows B. The maximal expiratory flow-volume curve: normal standards, variability, and the effects of age. *Am Rev Respir Dis* 1976; 113:587-600.
- 11 Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B. Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983;127:725-734.
- Krzyzanowski M. Changes of ventilatory capacity in an adult population during a five year period. *Bull Eur Physiopathol Respir* 1980;16:155-170.
- Kusiak R, Roos J. Analysis of pulmonary function data. *Can J Statistics* 1984; 12: 7-25.
- Kvalseth TO. Cautionary note about R². *Am Stat* 1985; 39:279-85.
- Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963-974.
- Laird N. Discussion. *Stat in Med* 1988; 7: 199-202.
- Laird N, Lange N, Stram D. Maximum likelihood computations with repeated measures: application of the EM algorithm. *Am Stat Assoc* 1987; 82: 97-105.
- Laird NM. Missing data in longitudinal studies. *Stat Med* 1988; 7:305-316.
- Laszlo G. Editorial. Standardised lung function testing. *Thorax* 1984;39:881-6.
- Lawless JF. *Statistical models and methods for lifetime data*. New York: John Wiley & Sons, 1982.
- Lawther PJ, Brooks AGF, Waller RE. Respiratory function measurements in a cohort of medical students: a ten-year follow-up. *Thorax* 1978;33:773-778.
- Lebowitz MD, Holberg CJ, Knudson RJ, Burrows B. Longitudinal study of pulmonary function development in childhood, adolescence and early adulthood (development of pulmonary function) *Am Rev Respir Dis* 1987 (in press)
- Lebowitz MD, Holberg CJ. Effects of parental smoking and other risk factors on the development of pulmonary function in children and adolescents. *Am J Epidemiol* 1988; 128: 589-597.

- Lee J, Koh D. Adjustment of bias in the comparison of means. *Brit J Med* 1987; 44:430.
- Leuallen EC, Fowler WS. Maximal midexpiratory flow. *Amer Rev Tuberculosis* 1955; 72: 783.
- Little RJA, Rubin DB. On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *Amer Stat* 1983; 37:218-220.
- Little RJA. Commentary. *Stat Med* 1988; 7:347-355.
- Louis TA, Robins J, Dockery DW, Spiro III A, Ware JH. Explaining discrepancies between longitudinal and cross-sectional models. *J Chron Dis* 1986;39:831-839.
- Louis TA. General methods for analysing repeated measures. *Stat Med* 1988; 7:29-46.
- Love RG, Miller BG. Longitudinal study of lung function in coal miners. *Thorax* 1982;37:193-197.
- Lyons JP, Campbell H. Evolution of disability in coalworkers' pneumoconiosis. *Thorax* 1976;31:527-533.
- Marubini E. General statistical considerations in analysis. In: Johnston FE, Roche AF, Susanne C. Eds. *Human Physical Growth and Maturation*. New York: Plenum Press, 1980. Ch.6
- McCullagh P, Nelder JA. *Generalized Linear Models*. London: Chapman and Hall, 1983.
- McFadden ER Jr, Linden DA. A reduction in maximum mid-expiratory flow rate: a spirometric manifestation of small airway disease. *Am J Med* 1972; 52:725-37.
- Menkes HA, Beaty TH, Cohen BH, Weinmann G. Nitrogen washout and mortality. *Am Rev Respir Dis* 1985;132:115-9.
- Miller A. *Pulmonary Function Tests In Clinical and Occupational Lung Disease*. Orlando: Grune & Stratton, 1986
- Miller A, Thornton JC. The interpretation of spirometric measurements in epidemiologic surveys. *Environ Res* 1980;23:444-468.
- Miller A, Elliott JC, Thornton JC, Warshaw R, Geiger M, Anderson H. Comparison of spirometry performed on the same subjects by two teams using similar instruments: an investigation of variability in prevalence of impairment. *Environ Res* 1980;21:229-234.
- Miller WF, Johnson Jr. RL, Wu N. Relationships between maximal breathing capacity and timed expiratory capacities. *J Appl Physiol* 1959;14:510-6.
- Montgomery DC, Peck EA. *Introduction to Linear Regression Analysis*. New York: John Wiley & Sons, 1982.
- Morgan WKC, Seaton D. Pulmonary physiology - its application to the determination of respiratory impairment and disability in industrial lung disease. In: Morgan WKC, Seaton A, eds. *Occupational lung diseases*. 2nd ed. Philadelphia: W.B. Saunders Company, 1984; 18-76.
- Morris CN, Rolph JE. *Introduction to Data Analysis and Statistical Inference*. New Jersey:

Prentice-Hall, Inc, 1981.

- Morrison DF. The optimal spacing of repeated measurements. *Biometrics* 1970; 26: 281-90.
- Neter J, Wasserman W, Kutner MH. *Applied Linear Statistical Models*. 2nd Ed. Illinois: Irwin, 1985.
- Ng T, Chan S, Lam K. Radiological progression and lung function in silicosis: a ten year follow up study. *Br Med J* 1987;295:164-168.
- Olofsson J, Bake B, Svardsudd K, Skoogh BE. The single breath N_2 test predicts the rate of decline in FEV₁. The study of men born in 1913 and 1923. *Eur J Respir Dis* 1986;69:46-56.
- Palta M, Cook T. Some considerations in the analysis of rates of change in longitudinal studies. *Stat in Med* 1987; 6:599-611.
- Pare PD, Brooks LA, Bates J, Lawson LM, Nelems JMB, Wright JL, Hogg JC. Exponential analysis of the lung pressure-volume curve as a predictor of pulmonary emphysema. *Am Rev Respir Dis* 1982; 126:54-61.
- Pern PO, Love RG, Wightman AJA, Soutar CA. Characteristics of coalminers who have suffered excessive loss of lung function over 10 years. *Bull Eur Physiopathol Respir* 1984;20:487-493.
- Peto R, Speizer FE, Cochrane AL, Moore F, Fletcher CM, Tinker CM, Higgins ITT, Gray RG, Richards SM, Gilliland J, Norman-Smith B. The relevance in adults of air-flow obstruction, but not of mucus, hypersecretion, to mortality from chronic lung disease. *Am Rev Respir Dis* 1983;128:491-500.
- Petty TL, Pierson DJ, Dick NP, Hudson LD, Walker SH. Follow-up evaluation of a prevalence study for chronic bronchitis and chronic airway obstruction. *Am Rev Respir Dis* 1976;114:881-890.
- Pham QT, Benis AM, Mur JM, Sadoul P, Haluszka J. Follow-up study of construction workers with obstructive lung disease. *Scand J Respir Dis* 1977;58:215-226.
- Postma DS, Burema J, Gimeno F, May JF, Smit JM, Steenhuis EJ, Van der Weele LTh, Sluiter HJ. Prognosis in severe chronic obstructive pulmonary disease. *Am Rev Respir Dis* 1979;119:357-367.
- Poukkula A, Huhti E, Makarainen M. Chronic respiratory disease among workers in a pulp mill. A ten-year follow-up study. *Chest* 1982;81:285-289.
- Ratkowsky DA. *Nonlinear Regression Modeling. A Unified Practical Approach*. New York: Marcel Dekker, Inc., 1983.
- Ries AL, Clausen JL. Lung volumes. In: Wilson AF, ed. *Pulmonary function testing indications and interpretations*. Orlando: Grune & Stratton, 1985; 69-85.
- Rode A, Shephard RJ. Lung function in Canadian Inuit: a follow-up study. *Can Med Assoc J* 1984;131:741-744.

- Rom WN, Greaves W, Bang KM, Holthouser M, Campbell D, Bernstein R. An epidemiologic study of the respiratory effects of trona dust. *Arch Environ Hlth* 1983;38:86-92.
- Rosenstock L, Cullen M. *Clinical Occupational Medicine*. Philadelphia: W.B.Saunders Co.,1986.
- Rosner B. The analysis of longitudinal data in epidemiologic studies. *Scand J Respir Dis* 1976;57:309-322.
- Rosner B, Munoz A, Tager I, Speizer F, Weiss S. The use of an autoregressive model for the analysis of longitudinal data in epidemiologic studies. *Statistics in Medicine* 1985;4:457-467.
- Rosner B, Munoz A. Autoregressive modelling for the analysis of longitudinal data with unequally spaced examinations. *Stat Med* 1988; 7:59-72.
- Saric M, Kalacic I, Holetic A. Follow-up of ventilatory lung function in a group of cement workers *Br J Indust Med* 1976;33:18-24.
- Schachter EN. Small airways function: the tests and evaluation. In: Miller A. Ed. *Pulmonary Function Tests in Clinical and Occupational Lung Disease*. Orlando: Grune & Stratton Inc, 1986.
- Schachter EN, Doyle CA, Beck GJ. A prospective study of asthma in a rural community. *Chest* 1984;85:623-630.
- Schluchter MD. Analysis of incomplete multivariate data using linear models with structures covariance matrices. *Stat Med* 1988; 7:317-324.
- Schulzer M, Enarson DA, Chan-Yeung M. Analyzing cross-sectional and longitudinal lung-function measurements: the effects of age. *Can J Statistics* 1985;13:7-15.
- Schwertman NC, Heilbrun LK. A successive differences method for growth curves with missing data and random observation times. *J Am Stat Assoc* 1986; 81: 912-916.
- Segal MR, Weiss ST, Speizer FE, Tager IB. Smoothing methods for epidemiological analysis. *Stat in Med* 1988; 7:601-11.
- Shigeoka JW. Pulmonary function testing. In: Rom WN, ed. *Environmental and occupational medicine*. Boston: Little, Brown & Co, 1983; 99-112.
- Siracusa A, Cicioni C, Volpi R, Canalicchi P, Brugnamì G, Comodi AR, Abbritti G. Lung function among asbestos cement factory workers: cross-sectional and longitudinal study. *Am J Ind Med* 1984;315-325.
- Siracusa A, Forcina A, Volpi R. An 11-year follow up on lung function among PVC, cement and asbestos cement factory workers. Presented at V International Symposium, Epidemiology in occupational health, Los Angeles, September 1986.
- Snedecor GW, Cochran WG. *Statistical Methods*. Iowa: Iowa State University Press, 1972.
- Soutar CA, Hurley JF. Relation between dust exposure and lung function in miners and ex-miners. *Br J Ind Med* 1986;43:307-320.

- Sparrow D, Bosse R, Rosner B, Weiss ST. The effect of occupational exposure on pulmonary function. *Am Rev Respir Dis* 1982;125:319-322.
- SPSS Inc. SPSS-X Users Guide, 3rd Ed. Chicago: SPSS Inc, 1988.
- Statistics in Medicine. Vol 7. No.1/2, Jan.-Feb. 1988.
- Stebbing Jr JH. Chronic respiratory disease among nonsmokers in Hagerstown, Maryland. II. Problems in the estimation of pulmonary function values in epidemiological surveys. *Environ Res* 1971;4:163-192.
- Stebbing JH. Chronic respiratory disease among nonsmokers in Hagerstown, Maryland. II. Problems in the estimation of pulmonary function values in epidemiological surveys. *Environ Res* 1971; 4:163-92.
- Stigler SM. The History of Statistics. Cambridge: The Harvard University Press, 1986.
- Strope GL, Helms RW. A longitudinal study of spirometry in young black and young white children. *Am Rev Respir Dis* 1984;130:1100-1107.
- Tabona M, Chan-Yeung M, Enarson D, MacLean L, Schulzer M. Host factors affecting longitudinal decline in lung spirometry among grain elevator workers. *Chest* 1984;85:782-786.
- Tashkin DP, Clark VA, Coulson AH, Simmons M, Bourque LB, Reems C, kDetels R, Sayre JW, Rokaw SN. The UCLA population studies of chronic obstructive respiratory disease VIII. Effects of smoking cessation on lung function: A prospective study of a free-living population. *Am Rev Respir Dis* 1984;130:707-715.
- Taylor RG, Joyce H, Gross E, Holland F, Pride NB. Bronchial reactivity to inhaled histamine and annual rate of decline in FEV1 in male smokers and ex-smokers. *Thorax* 1985;40:9-16.
- Townsend MC, DuChene AG, Morgan J, Browner W. Cigarette smoking intervention in the multiple risk factor intervention trial: results and relationships to other outcomes. Chapter 6 - Pulmonary function in relation to cigarette smoking and smoking cessation. Unpublished manuscript, Pittsburgh: University of Pittsburgh, 1985.
- Townsend MC, Du Chene AG, Fallat RJ. The effects of underrecorded forced expirations on spirometric lung function indexes. *Am Rev Respir Dis* 1982;126:734-7.
- Townsend MC. The effects of leaks in spirometers on measurements of pulmonary function. The implications for epidemiologic studies. *J Occup Med* 1984;26:835-41.
- Tukey JW, Wilk MB. Data analysis and statistics: techniques and approaches. Proceedings of the Symposium on Information Processing in Sight Sensory Systems, November 1-3, 1965, California Institute of Technology, 1966.
- U.S. Dept of Health and Human Services. Epidemiology of respiratory disease: Task force report. Washington D.C.: National Institutes of Health Publication 81-2019, 1980.
- U.S. DHEW The health consequences of involuntary smoking. A report of the Surgeon General, 1986. Maryland: U.S. Department of Health and Human Services, 1986.

- U.S. Surgeon General. The health consequences of smoking. Chronic obstructive lung disease. U.S. Dept of Health and Human Services DHHS(PHS) 84-50205., Washington, 1984.
- Vacek PM, Mickey RM, Bell DY. A two-stage random effects model for pulmonary function changes in sarcoidosis patients. Paper presented at the 1987 Meeting of the American Statistical Association, August 20-23, San Francisco.
- Van 'T Hof MA. Some statistical and methodological aspects in the study of growth and development with a special emphasis on mixed longitudinal designs. Stichting Studentenpers Nijmegen, 1977.
- Van der Lende R, Kok TJ, Reig RP, Quanjer PH, Schouten JP. Decreases in VC and FEV1 with time: indicators for effects of smoking and air pollution. *Bull Physiopathol Respir* 1981;17:775-792.
- Van der Lende R, Kok T, Peset R, Quanjer PH, Schouten JP, Orie NGM. Longterm exposure to air pollution and decline in VC and FEV1. *Chest* 1981;80:suppl 1:23S-26S.
- Veney JD, Kaluzny AD. Evaluation and decision making for Health Services Programs. New Jersey: Prentice-Hall, Inc, 1984.
- Vollmer WM, Johnson LR, Buist AS. Relationship of response to a bronchodilator and decline in forced expiratory volume in one second in population studies. *Am Rev Respir Dis* 1985;132:1186-1193.
- Vollmer WM, Johnson LR, McCamant LE, Buist AS. Methodologic issues in the analysis of lung function data. *J Chron Dis* 1987; 40:1013-23.
- Vollmer Wm. Comparing change in longitudinal studies: adjusting for initial value. *J Clin Epidemiol* 1988;41:651-657.
- Vonesh EF, Carter RL. Efficient inference for random-coefficient growth curve models with unbalanced data. *Biometrics* 1987; 43:617-28.
- Ware J. Discussion. *Stat in Med* 1988; 7: 199-202.
- Ware JH, De Gruttola V. Multivariate linear models for longitudinal data: a bootstrap study of the GLS estimator. In: Sen PK. Ed. *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences*. Elsevier Science Publishers, 1985.
- Ware JH. Linear models for the analysis of longitudinal studies. *The American Statistician* 1985;39:95-101.
- Waternaux C, Laird NM, Ware JH. Methods for analysis of longitudinal data: blood lead concentrations and cognitive development. *J Am Stat Assoc* 1989; 84: 33-41.
- Wegman DH, Peters JM, Pagnotto L, Fine LJ. Chronic pulmonary function loss from exposure to toluene diisocyanate. *Br J Ind Med* 1977;34:196-200.
- Weiss St, Speizer FE. Increased levels of airways responsiveness as a risk factor for development of chronic obstructive lung disease, What are the issues? *Chest* 1984; 86:3-4.

- West JB. Pulmonary pathophysiology – The Essentials. Baltimore: The Williams & Wilkins Company, 1977.
- Whittemore AS. Air pollution and respiratory disease. *Ann Rev Public Health* 1981;2:397–429.
- Wilhelmsen L, Orha I, Tibblin G. Decrease in ventilatory capacity between ages of 50 and 54 in representative sample of Swedish men. *Br Med J* 1969;3:553–556.
- Winer BJ. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.
- Wu MC, Bailey K. Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Stat Med* 1988; 7:337–346.
- Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121–130.
- Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988; 44: 1049–1060.