CORRECTING FOR THE EFFECT OF MEASUREMENT ERROR ON THE ASSOCIATION BETWEEN ENVIRONMENTAL EXPOSURE AND RESPIRATORY IMPAIRMENT

Leslie Michael Irwig

A thesis submitted to the Faculty of Medicine
University of the Witwatersrand, Johannesburg
for the degree of Doctor of Philosophy.

Johannesburg 1985

ABSTRACT

Exposure to environmental agents may cause respiratory impairment. The tests for respiratory impairment are prone to measurement error. The objective of this thesis is to examine how random error of measurement affects the association found between environmental exposure and respiratory impairment and to explore methods of correcting for these effects. This is done for two types of respiratory impairment, chronic airway obstruction as measured by the change in lung function test results over time, and pneumoconiosis detected on the chest radiograph.

The issues of random error of measurement in longitudinal lung function data are explored with reference to spirometric and peak flow test results in 433 men aged between 18 and 22 years at entry into the [illegible]cing industry. Lung function was measured initially and the change from initial measurement calculated over the subsequent 11 years. Analysis of the association between lung function change and environmental exposure should take into account the relationship between change in lung function and the initial level. When lung function change is regressed on the initial measurement and environmental exposure, random error of measurement in the lung function test results has several consequences:

1) The regression of lung function change on the initial lung function measurement is biased because of random error in the initial measurement.

2) If initial lung function and environmental exposure are associated, random measurement error in the initial lung function (the confounder) results in a biased regression of lung function change on environmental exposure.

3) The variance of lung function change explained by environmental exposure is underestimated if lung function change is measured with random error. Consequently statistical power, i.e. the likelihood of detecting a true association between environmental exposure and lung function change, is reduced.

4) If lung function change is dichotomized to obtain exposure-response relationships, random measurement error causes an underestimation of the slope of the exposure-response relationship.

With the exception of the effect on statistical power, all of the consequences of random error of measurement can be corrected at analysis stage. Several correction methods are suggested in the literature. The most attractive of them on theoretical and practical grounds is that using the reliability coefficient, defined as the ratio of the true (error-free) variance to the variance of the measured variable. This is easily estimated as the correlation between repeat measurements of the underlying true level. Correction for the effect of random measurement error requires multiplying the variance of lung function test measurements by the reliability coefficient. The corrected variance is then used to calculate regression or correlation coefficients.

The consequences of measurement error in radiographically-detected pneumoconiosis is explored with reference to the assessment of asbestosis by 3 independent readers in 1 680 asbestos miners and millers. Radiographically-detected pneumoconiosis is an example of a dichotomized variable, i.e. one that is analysed as a binary variable to obtain an exposure-response relationships but in fact has an underlying continuum which, in this case, is not measurable. The prevalence of abnormality and the exposure-response relationship may vary between observers because of differing random error and differences in the threshold implicitly chosen for categorizing radiographs, even in the presence of standardized criteria. The biserial correlation coefficient can be used to measure the correlation between environmental exposure and the continuum (including random error) underlying the dichotomized readings of an observer. Normal distribution theory can then be used to estimate exposure-response relationships at any required threshold for each reader. Biserial correlation coefficients and exposure-response relationships can be corrected for random observational error using the reliability coefficient, calculated as the tetrachoric correlation between repeat observations by the same or different readers.

The methods of correcting for error of measurement described in the thesis have wide applicability in the assessment of associations between environmental exposure and respiratory impairment. The concepts are easily transferable to other areas of epidemiological research.

DECLARATION

This thesis is my work. However, in the spirit of collaboration required for epidemiological research, many colleagues have helped me as indicated in the acknowledgements.

This thesis has not previously been submitted for a degree in any university. The work was conducted during my employment by the South African Medical Research Council.

Les Irving

.....

July 1985

DEDICATION

To my family; Judy, Kevin, Gary, André and Dani - for their love, music, and humour.

ACKNOWLEDGEMENTS

Thesis supervision

I thank:

- Professor Hennie Groeneveld, my statistical mentor, for the many stimulating hours spent together exploring and developing new ideas, and for providing high-level computing expertise.
- Professor Margaret Becklake, for her infectious enthusiasm, wide content expertise and editorial wisdom.

Contribution to the development of ideas

I thank:

- Mr Jan Pretorius, for introducing me over several years to the fascinating concepts surrounding measurement error.
- Dr Trevor Hastie, for discussion and contribution of ideas in Chapters 7 and 8.
- Dr Carl Lombard, for critical and careful comment on Chapters 3, 5 and 6.
- Professor D Krige, for discussion on methods of estimating the variance of measurement error (Chapter 5).
- Professor John Gear, Drs Duncan Saunders and Allen Herman, for comments on the clarity of verbal and written versions of the concepts described in the thesis.

Data collection

Although I was responsible for the collation of data on both data sets used in this thesis, I thank the following people for providing much of the basic information:

- The late Dr Helen van Doorn, who initiated the collection of longitudinal lung function data in miners.
- Dr Eugene Baskind, Mr E Olivier, and the staff of the lung function laboratory of the Medical Bureau for Occupational Diseases, for doing the lung function tests.
- Dr G K Sluis-Cremer, Dr Ray Glyn Thomas, and Professor Albert Solomon, for reading radiographs for pneumoconiosis.
- Dr Rudi du Toit, and the Cumulative Dust Exposure Record Committee for providing information on exposure to dust in mines.

Computing assistance

I thank Ms Magda de Beer, Ms April Butchart, and Ms Eva Hnizdo for provided computing assista e.

Organisational support

The South African Medical Research Council employed me during the time that the work on this thesis was done. My initial appointment was in the National Research Institute for Occupational Diseases, then under the directorship of Professor Ian Webster. After that, I was employed in the Institute for Biostatistics, under the Directorship of Dr Stephen Fellingham.

I thank Mrs Brenda Brooks for expertly typing the thesis.

PUBLICATIONS

Previous publications on a data set which is used in this thesis for the exploration of analytic techniques are:

Irwig L M, du Toit R S J, Sluis-Cremer G K, et al.
Risk of asbestosis in crocidolite and amosite mines in South Africa.
Ann N Y Acad Sci 1979; 330: 35-52.

Solomon A, Irwig L M, Sluis-Cremer G K, Thomas R G, du Toit R S J.
Thickening of pulmonary interlobar fissures: exposure-response relationship in crocidolite and amosite miners.
Br J Ind Med 1979; 36: 195-198.

Some of the concepts embodied in the thesis are dealt with in:

Irwig L M, Groeneveld H T, Pretorius J P G, and Hnizdo E.
Relative observer accuracy for dichotomized variables.
J Chronic Dis 1985; In press.

ABBREVIATIONS

For simplicity of presentation and reading, the following abbreviations are used for lung function tests in the thesis. The units of measurement of these tests are omitted in the tables.

FVC = Forced vital capacity (litres)

FEV = FEV_1
Forced expiratory volume in 1 second (litres)

FEF = $FEF_{25-75\%}$
Forced expiratory flow between 25% and 75% of Forced Vital Capacity (maximal midexpiratory flow) (litres per second)

$FEV\% = \frac{FEV}{FVC} \times 100$

PEFR = Peak expiratory flow rate (litres per minute)

LFT = Lung function tests; a collective term for any or all of the above

CONTENTS

	Page No
ABSTRACT	i
DECLARATION	iv
DEDICATION	v
ACKNOWLEDGEMENTS	vi
PUBLICATIONS	viii
ABBREVIATIONS	ix
LIST OF TABLES	xv
LIST OF FIGURES	xviii
CHAPTER 1 INTRODUCTION	1
1.1 EXPOSURE TO ENVIRONMENTAL AGENTS AND RESPIRATORY IMPAIRMENT	1
1.2 METHODS OF IMPROVING THE DETECTION OF ASSOCIATIONS BETWEEN EXPOSURE TO ENVIRONMENTAL AGENTS AND RESPIRATORY IMPAIRMENT	3
1.3 TESTS FOR CHRONIC NON-MALIGNANT RESPIRATORY DISEASE	3
1.3.1 Chronic obstructive pulmonary disease	3
1.3.2 Pneumoconiosis	5
1.4 INTERPRETATION OF TESTS WHICH MEASURE THE CONSEQUENCES OF ENVIRONMENTAL EXPOSURE ON THE RESPIRATORY TRACT	6
1.5 SOME IMPLICATIONS OF ERROR OF MEASUREMENT	7
1.5.1 Effect of random measurement error on the distribution of a variable	10
1.5.2 Effect of random measurement error on measures of association between two variables	11
1.5.3 The effect of random measurement error in a confounder	11
1.6 CONCLUSION	14
CHAPTER 2 ANALYSIS OF LONGITUDINAL LUNG FUNCTION DATA TO DETECT CONSEQUENCES OF ENVIRONMENTAL EXPOSURE: PRINCIPLES AND PRACTICE	
2.1 PRINCIPLES OF ANALYSIS OF LONGITUDINAL LUNG FUNCTION DATA	16
2.1.1 Is initial lung function a potential confounder of the association between environmental exposure and lung function change?	17

Page No

2.1.2 The relationship between lung function change, initial measurement, and exposure when lung function is measured with random error 21

2.2 THE PRACTICE OF LONGITUDINAL LUNG FUNCTION DATA ANALYSIS: REVIEW OF THE LITERATURE 22
2.2.1 Cross-sectional analysis of exposed and unexposed groups 23
2.2.2 Lung function change related to exposure 23
2.2.3 Lung function change regressed simultaneously on exposure and initial lung function 24
2.2.4 Lung function change regressed simultaneously on exposure and the mean of the initial and final measurement 24
2.2.5 The ratio of lung function change to level or the ratio of final to initial measurement in relationship to exposure 24
2.2.6 Regression of final lung function measurement on exposure after adjusting for initial lung function measurement 25

2.3 CONCLUSION 26

CHAPTER 3 METHODS OF CORRECTING FOR THE EFFECT OF RANDOM ERROR OF MEASUREMENT IN REGRESSION ANALYSIS OF LONGITUDINAL DATA 27

3.1 EFFECT OF RANDOM ERROR OF MEASUREMENT ON REGRESSION COEFFICIENTS 27
3.1.1 Effect of random measurement error when L_1 and X are not correlated 31
3.1.2 Effect of random measurement error when L_1 and X are correlated 32

3.2 THE CALCULATION OF REGRESSION COEFFICIENTS USING RELIABILITY COEFFICIENTS TO CORRECT FOR RANDOM ERROR OF MEASUREMENT 33
3.2.1 Correcting regression coefficients when X and L_1 are not correlated 34
3.2.2 Correcting regression coefficients when X and L_1 are correlated 35

3.3 OTHER METHODS OF CORRECTING FOR THE EFFECT OF RANDOM MEASUREMENT ERROR ON THE REGRESSION OF L_2-L_1 ON L_1 37
3.3.1 Substituting an independent measure of L_1 38
3.3.2 Substituting L_1 predicted from other variables 38
3.3.3 Substituting $(M_1+M_2)/2$ for L_1 39

3.4 CONCLUSION 42

		Page No
CHAPTER 4	THE LONGITUDINAL LUNG FUNCTION DATA SET: STUDY METHODS AND DATA STRUCTURE	43
4.1	STUDY METHODS	43
4.1.1	The data source	43
4.1.2	The sample	44
4.1.3	Measurement of lung function, height and weight	44
4.1.4	Measurement of smoking and occupational exposure	47
4.2	DATA STRUCTURE	49
4.2.1	Data for analysis of factors associated with long-term lung function change	50
4.2.2	Data for estimation of reliability coefficients	53
CHAPTER 5	ESTIMATING RANDOM ERROR OF MEASUREMENT IN LUNG FUNCTION DATA	54
5.1	CONCEPTS FOR ESTIMATING RELIABILITY COEFFICIENTS	54
5.1.1	Methods of estimating G	54
5.1.2	Components of variance included in the definition of random error of measurement	56
5.2	A SIMPLE ESTIMATION OF G IN THE VAN DOORN DATA SET	57
5.3	CHECKING ASSUMPTIONS FOR CORRECTING REGRESSION COEFFICIENTS USING RELIABILITY COEFFICIENTS OR INSTRUMENTAL VARIABLES	58
5.3.1	Assumptions for using G to correct regression coefficients	58
5.3.2	Assumptions for obtaining unbiased regression coefficients using instrumental variables	60
5.4	OTHER METHODS OF ESTIMATING RELIABILITY COEFFICIENTS	61
5.4.1	Concepts	61
5.4.2	Methods	63
5.4.3	Results	64
5.5	CONCLUSION	68
CHAPTER 6	THE ASSOCIATION OF LUNG FUNCTION CHANGE WITH INITIAL LUNG FUNCTION AND ENVIRONMENTAL EXPOSURE AFTER CORRECTING FOR RANDOM ERROR OF MEASUREMENT	70
6.1	CHARACTERISTICS OF THE DATA SET	72
6.1.1	Descriptive statistics	
6.1.2	Correlations of environmental exposure variables and initial lung function	77
6.1.3	Searching for other confounders	79
6.1.4	Conclusion	80

Page No

6.2 REGRESSION OF LUNG FUNCTION CHANGE ON INITIAL LUNG FUNCTION (ASSUMING NO ASSOCIATION BETWEEN INITIAL LUNG FUNCTION AND ENVIRONMENTAL EXPOSURE) 83
6.2.1 Methods 83
6.2.2 Results and discussion 84
6.2.3 Conclusion 88

6.3 REGRESSION OF LUNG FUNCTION CHANGE ON ENVIRONMENTAL EXPOSURE (ASSUMING THAT INITIAL LUNG FUNCTION IS NOT A CONFOUNDER) 91
6.3.1 Methods 91
6.3.2 Results 91
6.3.3 Discussion 93
6.3.4 Conclusion 96

6.4 CORRECTION FOR RANDOM ERROR OF MEASUREMENT IN MULTIPLE REGRESSION ANALYSIS (WHEN THE PREDICTORS INITIAL LUNG FUNCTION AND ENVIRONMENTAL EXPOSURE MAY BE CORRELATED) 97
6.4.1 Methods 97
6.4.2 Results and discussion 100
6.4.3 Conclusion 103

CHAPTER 7 THE CONSEQUENCES OF RANDOM ERROR OF MEASUREMENT IN A DEPENDENT VARIABLE: THE RELATIONSHIP BETWEEN LUNG FUNCTION CHANGE AND ENVIRONMENTAL EXPOSURE 104

7.1 RANDOM ERROR OF MEASUREMENT IN A CONTINUOUS DEPENDENT VARIABLE: EXPOSURE-EFFECT RELATIONSHIPS 104

7.2 RANDOM ERROR OF MEASUREMENT IN A DICHOTOMIZED DEPENDENT VARIABLE: EXPOSURE-RESPONSE RELATIONSHIPS 107
7.2.1 Concepts 107
7.2.2 Application 112

7.3 CONCLUSION 115

CHAPTER 8 RANDOM ERROR OF MEASUREMENT IN DICHOTOMIZED VARIABLES: EXPOSURE-RESPONSE RELATIONSHIPS FOR PNEUMOCONIOSIS 116

8.1 AN ILLUSTRATION OF THE CONSEQUENCES OF READER VARIABILITY ON EXPOSURE-RESPONSE RELATIONSHIPS FOR ASBESTOSIS 117
8.1.1 The Data Set 117
8.1.2 Results 119
8.1.3 Discussion 121

		Page No
8.2	METHODS OF ESTIMATING EXPOSURE-RESPONSE RELATIONSHIPS FOR DICHOTOMIZED RESPONSE VARIABLES	123
	8.2.1 Estimating ρ from dichotomized data	124
	8.2.2 Exposure-response relationships for each reader on the same threshold	127
	8.2.3 Exposure-response relationships corrected for random measurement error	129
8.3	AN EXAMPLE: ESTIMATING EXPOSURE-RESPONSE RELATIONSHIPS FOR RADIOGRAPHICALLY DETECTED ASBESTOSIS	131
8.4	CONCLUSION	139
CHAPTER 9	CONCLUSION	140
9.1	LUNG FUNCTION CHANGE AND EXPOSURE TO ENVIRONMENTAL AGENTS	140
	9.1.1 Summary: Effects of random error measurement and correction procedures	140
	9.1.2 Implications for future research and methodological development	141
9.2	RADIOGRAPHICALLY-DETECTED PNEUMOCONIOSIS AND ENVIRONMENTAL EXPOSURE	145
	9.2.1 Summary: Effect of error and correction procedures	145
	9.2.2 Implications for future research and methodological development	146
REFERENCES		148

LIST OF TABLES

		Page No
Table 3-1	Hypothetical data for X, M_1 and M_2 (all measured without error)	41
Table 4-1	Characteristics of 18-22 year-old men who remained in the mining industry or left	52
Table 5-1	Reliability coefficient ($\hat{G}$) and descriptive statistics for within-person differences in lung function tests one year apart (n = 331)	57
Table 5-2	Within-person variance by tertiles of lung funct.on level (n = 331)	59
Table 5-3	Within-person variance for transformed lung function measurements by tertiles of level	59
Table 5-4	Correlation matrix of within-person differences in lung function (n = 331)	61
Table 5-5	Reliability coefficients (G) for lung function measurements in the van Doorn data set	66
Table 6-1	Descriptive statistics for personal characteristics (n = 433)	72
Table 6-2	Descriptive statistics for lung function measurements	73
Table 6-3	Skewness and kurtosis of lung function measurements	76
Table 6-4	Frequency distribution for smoking	76
Table 6-5	Frequency distribution of the number of years worked in high-dust jobs on gold-mines	76
Table 6-6	Frequency distribution of cumulative dust exposure (Particles/ml x hours exposed/day x years) in underground gold-miners (n = 106)	78
Table 6-7	Correlations of exposure variables and initial lung function test results	79
Table 6-8	Correlation of potential confounders with lung function change	79
Table 6-9	Correlation of potential confounders with initial lung function, exposure variables and each other	81

		Page No
Table 6-10	The regression of change on initial measurement using various methods (n = 433)	85
Table 6-11	Regression of change (last-middle LFT) on middle and initial LFT (Method 5, n = 221)	87
Table 6-12	Regression of lung function change on environmental variables	92
Table 6-13	P-values (and sign, if negative) of partial association of lung function change separately on each type of dust measurement	93
Table 6-14	Cross-classification of height and average cigarette consumption	98
Table 6-15	Correlation induced between lung function and average cigarette consumption (n = 125)	98
Table 6-16	Regression coefficients of FEV change on initial FEV and cigarette consumption	100
Table 6-17	Sensitivity of correction of regression analyses using different values of the reliability coefficient (G)	102
Table 6-18	Consequences of error in both predictors	102
Table 8-1	Frequency distributions of parenchymal abnormality found by readers A, B, and C (n = 1 680)	118
Table 8-2	Frequency distribution of duration of asbestos exposure	119
Table 8-3	Effect of different thresholds in a standard bivariate normal distribution with a correlation of 0.4 (Theoretical calculations)	124
Table 8-4a	Association between normalized duration of asbestos exposure and parenchymal abnormality detected by reader A at various boundaries (n = 1 680)	132
Table 8-4b	Probabilities of differences between the biserial correlation in table 8-4a arising by chance	132
Table 8-5a	Biserial correlation coefficients of normalized duration of asbestos exposure and parenchymal abnormality (Boundary 1) for readers A, B, and C	133

		Page No
Table 8-5b	Probabilities of differences between the biserial correlations in table 8-5a arising by chance	133
Table 8-6	Estimates of exposure-response relationship for A from logistic regression on normalized exposure and biserial correlation, at boundary 1	135
Table 8-7	Prevalence of abnormality by exposure at a threshold set so that reader A finds 20% abnormal	137
Table 8-8	Tetrachoric correlations (and bootstrap standard errors in brackets) between readers A, B and C	138
Table 8-9	Estimated prevalence of true abnormality by exposure at the threshold set so that A finds 20% abnormal	138

LIST OF FIGURES

		Page No
Figure 1-1	Distribution of true values of a variable (solid line) and values measured with random error (dotted line)	10
Figure 1-2	Unaltered regression and increased residual variance when a dependent variable is measured with error	12
Figure 1-3	Attenuated regression coefficient when a predictor is measured with error	12
Figure 1-4a	No difference between the mean value of the dependent variable in exposed ($\bar{Y}_e$) and unexposed ($\bar{Y}_u$) groups after adjustment for a confounder (C)	13
Figure 1-4b	Underadjustment occurs when C is measured with error so that the adjusted $\bar{Y}_u$ and $\bar{Y}_e$ are found to differ	13
Figure 2-1	Lung function change with age in exposed and unexposed groups of men whose mean lung function differed at age 40. Exposure only starts after age 40	19
Figure 3-1	Graphical representation of alternative ways of expressing the model	29
Figure 4-1	Sources of data used in analysis	50
Figure 5-1	Correlation coefficients of lung function measurements by time between examinations	65
Figure 5-2	Fitting models to PEFR correlelogram data	67
Figure 6-1	Frequency distributions of initial lung function	73
Figure 6-2	Frequency distributions of lung function change	74
Figure 7-1	Exposure-response relationship from conditional distribution of Y on X	108
Figure 7-2	Exposure-response relationships for FEV abnormality (loss > 0.5 litre) by smoking, with and without correction for random measurement error	114
Figure 8-1	Exposure-response relationship for boundaries 1-4 (reader A)	120

Figure 8-2 Exposure-response relationship for readers A, B, C at boundary 1 120

Figure 8-3 Bivariate distribution of a continuous exposure variable and a continuous disease variable, demonstrating different thresholds 125

Figure 8-4 Logistic regression of response by reader A at boundary 1 on exposure (repeated from figure 8-1) and on normalized exposure 135

Figure 8-5 Exposure-response relationship for A and C at the same threshold, and the relationship corrected for measurement error using reader pair AB 137

CHAPTER 1 INTRODUCTION

This thesis is concerned with methods of analysing data to test the association between exposure to environmental agents and respiratory impairment which is assessed by tests prone to error of measurement. The purpose of this chapter is to:

- explain the reasons for concern about the risk of environmentally-induced respiratory disease;
- review methods to detect adverse effects of environmental exposure on respiratory health;
- examine some of the statistical effects of random error of measurement on data analysis;
- outline the scope of the thesis and define terms used in it.

1.1 EXPOSURE TO ENVIRONMENTAL AGENTS AND RESPIRATORY IMPAIRMENT

> 'I am the Lorax', he coughed and he whiffed.
> He sneezed and he snuffled. He snarggled. He sniffed.
> 'Once-ler!' he cried with a cruffulous croak.
> 'Once-ler! You're making such smogulous smoke!
> My poor Swomee-Swans ... why, they can't sing a note!
> No one can sing who has smog in his throat...'
>
> The Lorax, by Dr Seuss[1]

Scientific and public concern has been voiced about the possible effects of exposure to environmental agents on human health.[2,3,4,5] The word 'environmental' is used here and throughout this thesis to include both occupational and non-occupational exposure as well as exposure due to personal habits such as smoking. Detection and

quantification of the association between environmental exposure and chronic non-malignant respiratory disease is required for several reasons:

1) Chronic non-malignant respiratory disease is a common cause of morbidity and mortality in many countries.[6,7] There is evidence that its causes include exposure to certain environmental agents, notably cigarette smoking and dusts or fumes encountered in the work environment.[8,9] Large numbers of people may be exposed to potential hazards at work.[8]

2) New agents are introduced into the work or general environment at an ever-increasing rate. Pretesting in the laboratory before their introduction is concerned largely with the detection of carcinogenic activity.[2] Particular concern has been expressed about substitutes for known respiratory hazards such as tobacco or asbestos, materials classified as nuisance dusts, and agents which have been shown to affect airways in laboratory experiments.[10,11]

3) Exposure to combinations of agents may have adverse health effects which are greater than their individual effects.[12]

4) Where exposure to an agent is known to cause disease, the relationship needs to be quantified accurately enough to allow rational decisions about exposure limits.[13]

5) Where exposure to known hazards is unavoidable, e.g. in certain occupations, exposed populations should be monitored to ensure that their health status is acceptable and that a reduction in environmental concentrations of the hazardous agent has the desired effect.[11]

1.2 METHODS OF IMPROVING THE DETECTION OF ASSOCIATIONS BETWEEN EXPOSURE TO ENVIRONMENTAL AGENTS AND RESPIRATORY IMPAIRMENT

A wide variety of research designs have been suggested to aid the early rapid detection of environmental agents that are hazardous to the health of man.[14-18] Most of these have been concerned with the risk of cancer. To detect whether exposure to environmental agents have non-malignant effects on the respiratory tract, the major thrust has been the development and use of tests of abnormality which precede the development of overt disease.

1.3 TESTS FOR CHRONIC NON-MALIGNANT RESPIRATORY DISEASE

There are two major chronic non-malignant respiratory diseases which are known to be caused by exposure to environmental agents; chronic obstructive pulmonary disease and pneumoconiosis.

1.3.1 Chronic obstructive pulmonary disease

Chronic obstructive pulmonary disease includes emphysema, chronic bronchitis, and chronic airway obstruction.[9] Emphysema is defined on pathological grounds as enlarged airspaces in the lung beyond the terminal bronchioles with destructive changes in the walls. Chronic bronchitis is characterized by the presence of chronic mucus hypersecretion resulting in a chronic productive cough. The associated pathological findings in the large airways are hypertrophy of mucus glands and goblet cell metaplasia.[8] Chronic airway obstruction

(also called chronic airflow limitation) is defined as a decrease in airway patency, classically measured by a reduction in the rapidity wich which an individual can expire air from his lungs after a maximal inspiration. Chronic airway obstruction is therefore an abnormality that can be measured objectively during life. It forms the major focus of interest in studies of chronic obstructive pulmonary disease and will be the only manifestation which is considered in this thesis.

The major method of measuring airway obstruction is by the use of a forced expiratory manoeuvre recorded on a spirometer. The classic measure is the forced expiratory volume in the one second period after a maximal inspiration (FEV_1, referred to in this thesis as FEV). FEV reaches its maximum level at about 25 years of age and thereafter declines progressively.[9] Because of the large reserve of function, disability will only be experienced when the major proportion of function is lost. The age at which disability occurs (if ever) depends upon the peak level of function attained in early adulthood and the subsequent rate of loss. Loss of lung function will be determined by individual variability, by the effect of exposure to agents which increase the rate of loss, and by complex interactions between these two components. If one wishes to know whether exposure to an environmental agent causes chronic airway obstruction, this can be achieved by testing whether exposure increases the rate of loss of airway patency, even though none of the individuals in that sample would be considered to have obstructive airways disease.

This thesis also is concerned with several other lung function measurements. Three of them are derived from a forced expiratory manoeuvre performed on a spirometer. They are:

- the forced vital capacity (FVC);
- the forced expiratory flow in the middle half of expiration ($FEF_{25-75\%}$, represented in this thesis as FEF);
- the ratio of FEV to FVC expressed as a percentage (FEV%).

The other test used in this thesis is the peak expiratory flow rate (PEFR) as measured on the Wright Peak Flow Meter. This measures the maximal flow achieved during expiration, which occurs shortly after expiration has commenced. Current theories about the pathogenesis of airway obstruction indicate that the earliest manifestations of abnormality are found in small airways, i.e. those less than 2 mm in diameter.[19] FEV and PEFR are both measured during the early part of expiration, when flow is mostly dependent on the diameter of large airways. On the other hand, FEF measures flow later during expiration.

1.3.2 Pneumoconiosis

Pneumoconiosis is diffuse infiltrative disease of the lung parenchyma due to the inhalation of inorganic dust.[20] The basic pathology is interstitial fibrosis, resulting in restriction of lung movement. Pneumoconiosis is usually detected by the presence of small opacities in the lung fields on radiographs. The profusion of small opacities is conventionally quantified on a 12 point scale according to a standardized classification.[21]

1.4 INTERPRETATION OF TESTS WHICH MEASURE THE CONSEQUENCES OF ENVIRONMENTAL EXPOSURE ON THE RESPIRATORY TRACT

The American Thoracic Society has portrayed a series of consequences of exposure to an environmental agent.[22] They suggest:

- a proportion of the population carry a pollutant burden (or internal dose),
- a smaller proportion show some physiological or pathophysiological change and
- only a very small proportion have morbidity or die.

The lung function or radiographic tests of concern in this thesis are aimed at the detection of irreversible physiological or pathophysiological changes. The value of these tests depends on their ability to predict subsequent morbidity and mortality. The less their predictive ability, the less interpretable is the medical significance of results showing that test results differ between populations exposed or not exposed to an environmental agent. There is some evidence that radiographically detected early pneumoconiosis predicts subsequent morbidity and mortality.[23] Good evidence exists that spirometric lung function measurements predict an increased risk of death even when adjustment is made for other risk factors, including smoking.[24-27] The American Thoracic Society regards this evidence as so strong that they state: 'It should be emphasized that a small but statistically significant reduction in a population mean FEV_1 or $FEV_{0.75}$ is probably medically significant,

as such a difference may indicate an increase in the number of persons with severe respiratory impairment in the population'.[22] A reduction in mean FEV can be thought of as a pathophysiological change whereas the concomitant increase in the number of persons with severe respiratory impairment constitutes morbidity. This concept has been developed further in other areas of environmental health where a differentiation is made between 'effect' and 'response'.[28] 'Effect' refers to a change in the mean value of a continuous test variable and is used to derive exposure-effect relationships, e.g. by regression analysis. 'Response' in the case of a continuous test variable refers to the proportion of the population that is beyond a defined threshold or magnitude of effect, ideally chosen to correspond to the magnitude that is considered medically important. The procedure gives rise to an apparently binary variable (response absent or present). Exposure-response relationships are then commonly obtained by logistic regression analysis.

1.5 SOME IMPLICATIONS OF RANDOM ERROR OF MEASUREMENT

> 'Error is all around us and creeps in at the least opportunity'.
>
> Charles Nicolle,
> French bacteriologist 1866-1936[29]

It is well known that lung function tests and findings on chest radiography are prone to measurement error.[30-34] The term error is used here in its statistical sense to mean 'variation' rather than 'mistake'. In this thesis it refers to deviation of a test measure-

ment from the true test value and does not include the issue of whether a true test value is an erroneous measure of the disease process it aims to detect. This introductory section deals only with error of measurement in a continuous variable such as lung function test results. Measurement error in this variable can be thought of as systematic or random.

Systematic error is not dealt with in this thesis apart from the following brief indication of its importance. It occurs when the error distribution is shifted by an amount associated with another variable of interest. For example, a population exposed to an environmental agent may have their lung function tested on one spirometer while the control unexposed population is tested on another spirometer. Systematic error occurs if the two spirometers are calibrated differently, i.e. would give a different mean value with repeat measurement of one population. Note however that the variance of the measurements within a population is unaffected. Systematic error biases the association between environmental exposure and test results, an issue which has been raised in the recent literature on occupationally-induced respiratory disease. Vermont granite workers were found to have a large mean loss of FVC and FEV compared to that expected from previous studies of other populations.[35] Granite workers who were subsequently re-examined had a mean increase in FVC and FEV suggesting that the results of the previous study may have been incorrect.[36] In this case, specific technical reasons for systematic error were suggested, such as inadequate expiratory effort

and a leak in the spirometer. Even in the absence of an identifiable cause, systematic error can often be demonstrated in longitudinal lung function testing.[37,38] The problem is best overcome by the use of concurrently-measured controls in the study design.

Random error of measurement may be conceptualized as follows. A test on an individual has a 'true' value during a particular time period. A measurement on any occasion may differ from the 'true' value in a random fashion, sometimes over- and sometimes under-estimating it. The errors include short-term biological variation of the individual and variability in the test procedure, equipment, calibration or interpretation. In this thesis all of these biological and technical reasons for short-term variability around a 'true' (mean) value are included in the term 'random error of measurement', also referred to as 'random measurement error', 'random observational error', or simply 'random error'.

Random measurement error can have serious consequences on the results obtained from data analysis. The reasons for this are detailed in later chapters. However it is useful at this stage to summarize the major effects without recourse to statistical equations. Random error of measurement in a continuous variable affects its distribution and its association with another variable. After adjustment for a confounder, the relationship between two variables is biased if the confounder is measured with error.

1.5.1 Effect of random measurement error on the distribution of a variable

By definition, random error of measurement will not alter the mean value of a variable. However, the variance of the observations will be greater than that of the true values[39] (Figure 1-1).

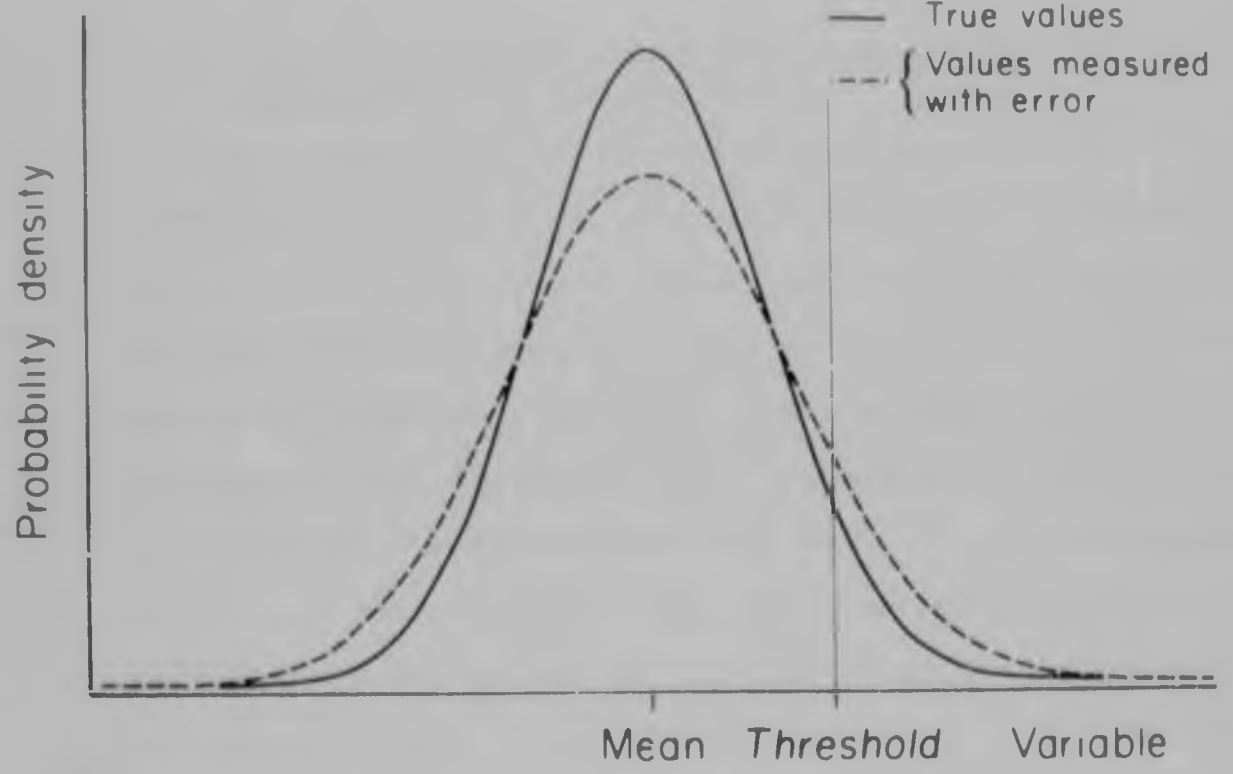

Figure 1-1 Distribution of true values of a variable (solid line) and values measured with random error (dotted line)

Two important consequences follow. First, the proportion of observed values beyond a specified threshold is greater than the true

proportion if the latter is less than 0.5 (Figure 1-1). This is important if we are considering the 'response to' rather than the 'effect of' an environmental exposure. Second, an extreme observation is more likely to contain a large error of measurement. In this case a repeat observation with independent random error of measurement will tend to be nearer the mean value of all measurements, a phenomenon known as regression to the mean.[40,41]

1.5.2 Effect of random measurement error on measures of association between two variables

The correlation between two variables and therefore the statistical significance of the association will be reduced or attenuated by random error of measurement in either variable.[42] The effect of random measurement error on the regression coefficient depends on whether the error occurs in the dependent variable or the predictor, and can be thought of as a consequence of the increased variance of the variable measured with error.[43] If the dependent variable is measured with error, the regression coefficient is unaffected although the residual variation around the regression line is increased (Figure 1-2). On the other hand, if the independent variable is measured with error, the regression coefficient will be attenuated (Figure 1-3).

1.5.3 The effect of random measurement error in a confounder

One often wishes to determine the partial association between a dependent variable and a predictor, i.e. the association after taking account of a confounder. A confounder is defined as a determinant of the dependent variable which is also associated with the predictor of

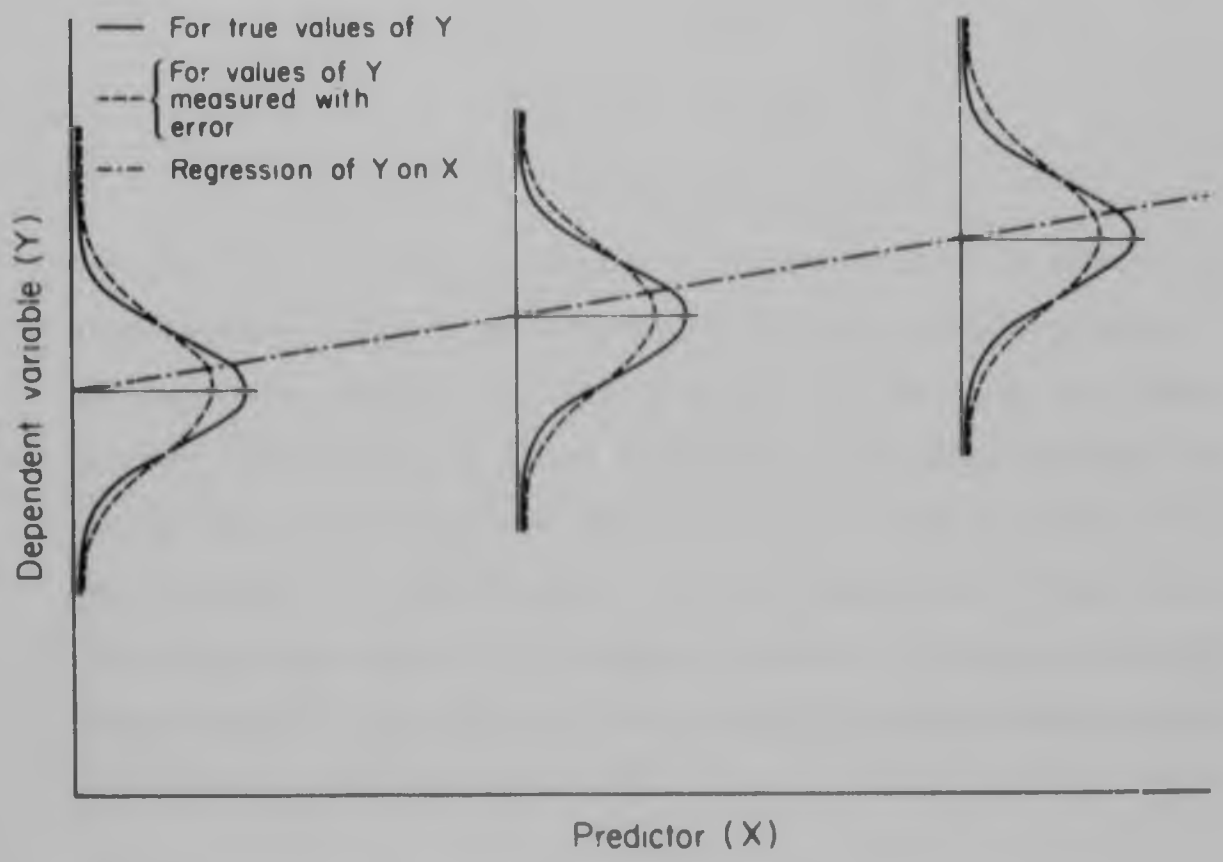

Figure 1-2 Unaltered regression and increased residual variance when a dependent variable is measured with error

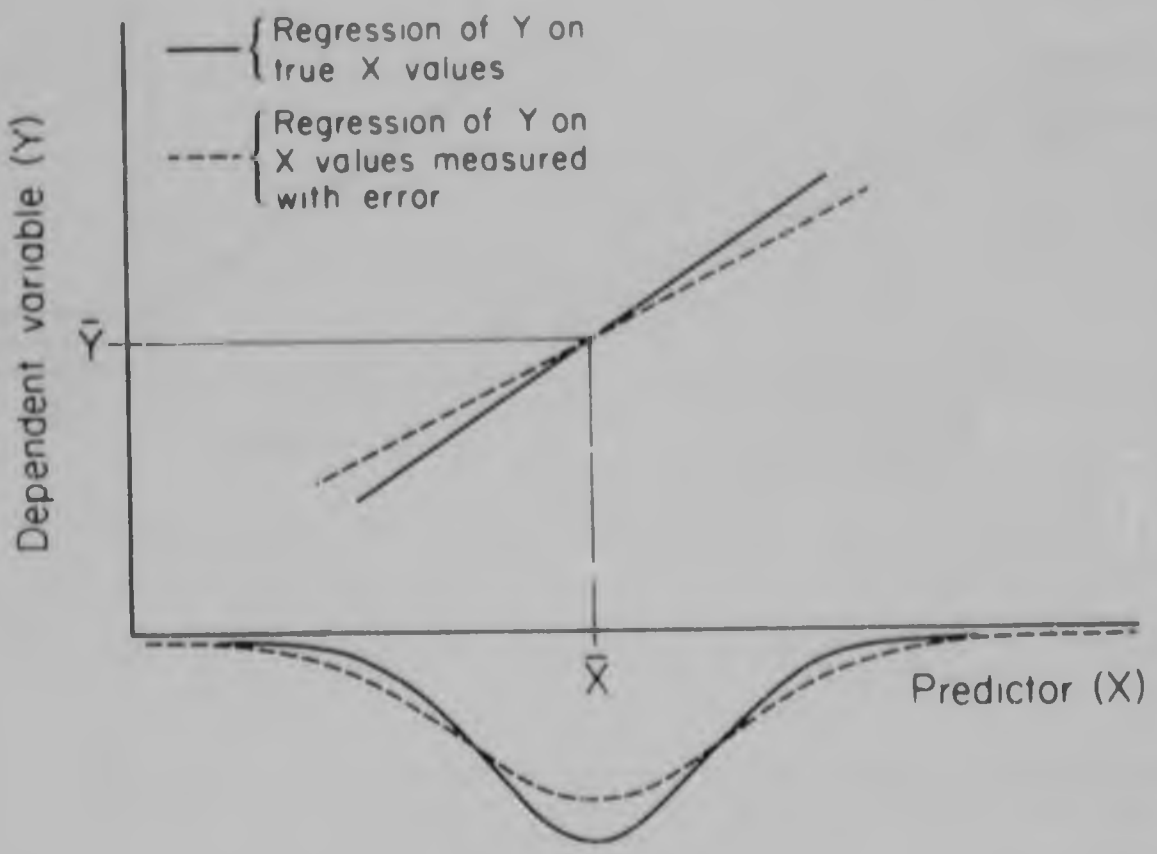

Figure 1-3 Attenuated regression coefficient when a predictor is measured with error

interest.[44] The term potential confounder is used to indicate a variable which should be considered a confounder if associated with the dependent variable and the predictor in the data set being analysed. The effect of random measurement error in a confounder is to attenuate its correlation with both of the other variables.[42] This results in overestimation of the association between the predictor of interest and the dependent variable. This effect can be conceptualised by considering the analysis of covariance example shown in Figure 1-4 (after Goldstein).[45]

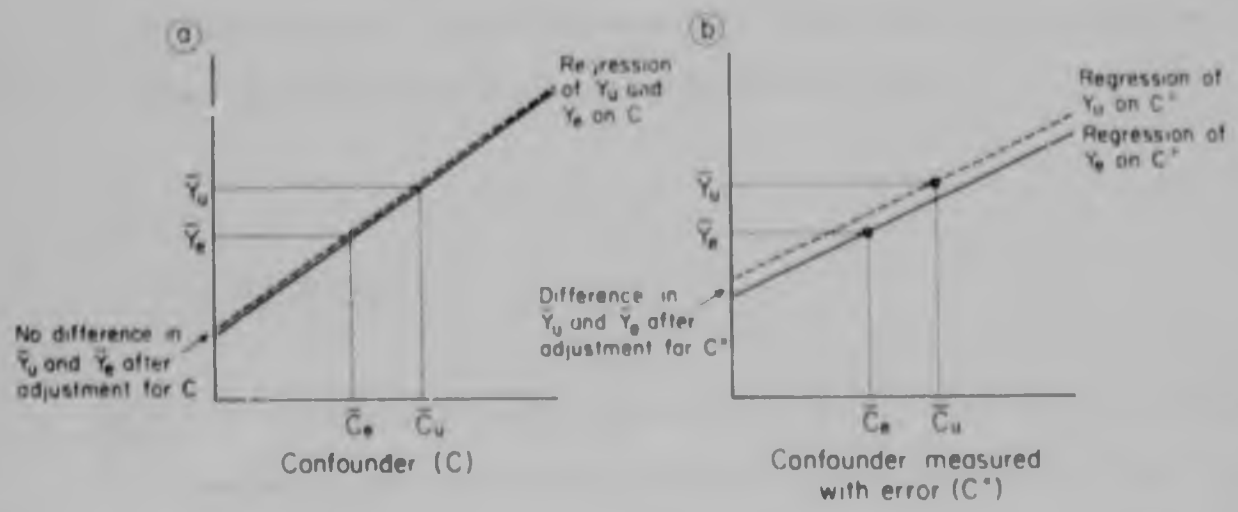

Figure 1-4 a. No difference between the mean value of the dependent variable in exposed ($\bar{Y}_e$) and unexposed ($\bar{Y}_u$) groups after adjustment for a confounder (C).

b. Underadjustment occurs when C is measured with error so that the adjusted $\bar{Y}_u$ and $\bar{Y}_e$ are found to differ.

Non-exposed and exposed populations have different mean values for the confounder represented on the X-axis. The populations also have different mean values for the dependent variable (Y-axis). After adjusting for the confounder using the true regression coefficient no effect of exposure is found (Figure 1-4a). The regression of the dependent variable on the confounder is attenuated when the confounder is measured with error. However, the attenuated regression still passes through the point defined by the mean for the dependent variable and the mean for the confounder in that exposure group. Adjustment for the confounder using the attenuated regression coefficient results in the finding that exposure has an effect on the dependent variable (Figure 1-4b). This result is spurious. It is due to underadjustment for the confounder because of its random error of measurement. This occurs even if the predictor of interest and the dependent variable are measured without error.

1.6 CONCLUSION

A conceptually sound approach to the detection of non-malignant consequences of environmental exposure on the respiratory tract is to examine the association between exposure and tests of early respiratory abnormality. However, these tests are known to be measured with error. Measurement error has known distortive effects on statistical associations. The purpose of this thesis is to examine the consequences of measurement error in the assessment of risks of

environmental exposure and to present analytic methods to avoid these consequences. This will be done for two study types in common use:-

1) Longitudinal assessment of lung function in relationship to smoking and occupational exposure (Chapters 2 to 7).

2) Multi-observer radiological assessment of pneumoconiosis in relationship to occupational exposure (Chapter 8).

CHAPTER 2 ANALYSIS OF LONGITUDINAL LUNG FUNCTION DATA TO DETECT CONSEQUENCES OF ENVIRONMENTAL EXPOSURE: PRINCIPLES AND PRACTICE

Random error of measurement distorts the association between variables (Section 1.5). The purpose of this chapter is to review whether these distortive effects are recognized and avoided in the longitudinal lung function literature. However, it is first necessary to explore some principles of longitudinal data analysis. Throughout the thesis, longitudinal data refers to measurement of the variable of interest on two occasions, though most of the concepts are transferable to data collected on many occasions.

2.1 PRINCIPLES OF ANALYSIS OF LONGITUDINAL LUNG FUNCTION DATA

The issues involved in longitudinal data analysis are best explored first using a hypothetical example. Consider a group of men whose exposure to a single environmental agent starts at the age of 40 years and in whom lung function is measured without error. The mean lung function of these men measured cross-sectionally (on only one occasion) at the age of 50 years may be poorer than that of a control group of men who have not been exposed to the environmental agent. This finding could arise because of an effect of exposure or because individuals who entered jobs involving exposure started with poorer lung function than those in the unexposed group. The latter possibility can be examined if lung function was also measured when the exposed and unexposed groups

were 40 years old. If the mean of the initial lung function measurements at 40 years of age does not differ between the two groups, it is reasonable to consider the difference in lung function at age 50 to be an effect of exposure between the ages of 40 and 50 years. An obvious analytic method is to calculate the change in each individual's lung function as the measurement at 50 years of age minus that at 40 years of age. Lung function change measurements exclude many of the determinants of variability (components of variance) of cross-sectional data.[46] They include only the effect of exposure and individual biological variability over the 10 year period. Differences in lung function change between exposed and unexposed groups can be attributed to the effect of exposure if one assumes that biological variability is equal in the two groups. Therefore, inferring causality from longitudinal data requires fewer assumptions than in the case of cross-sectional data. There is also another advantage of longitudinal data. Because it has fewer sources of variability, data on lung function change have a smaller variance than cross-sectional data. This improves statistical power and thereby aids attempts to detect small effects of exposure to environmental agents. The power achieved by longitudinal data exceeds that achieved by explaining the variance of cross-sectional lung function measurements by other determinants of variability, such as anthropometric characteristics.

2.1.1 Is initial lung function a potential confounder of the association between environmental exposure and lung function change?

Which variables should be considered confounders is a difficult

decision in any analysis,[44] and especially so in the case of longitudinal data.[47,48] Imagine a situation where the mean initial lung function differs between the groups of men subsequently exposed or not exposed to the environmental agent. Whether initial lung function should be considered a confounder of the association between exposure and lung function change is explored in the following two scenarios.

In the first scenario lung function is measured without error in two groups of men at 40 and 50 years of age. The one group of men is exposed to the environmental agent after their initial lung function measurement and the other group remains unexposed. The group which is subsequently exposed to the environmental agent has a poorer mean lung function at age 40 than that of the unexposed group. This could be due to a greater rate of change from 25 years of age (when their lung function was at its peak level) until 40 years of age. Continuation of this rate results in a greater loss of lung function in the exposed group, in the absence of any effect of exposure (Figure 2-1, line B). This phenomenon of lower lung function in middle age predicting subsequent loss is known as the horse-racing effect.[49] The rate of decline from 25 to 40 years of age should be regarded in this example as a confounder; it is a determinant of change from the age of 40 to 50 years and is associated with exposure. The confounder cannot be directly measured. However, on the assumption that the exposed and unexposed groups had similar mean lung function at age 25, we can use lung function at age 40 as a proxy measure of the loss from 25 to 40 years of age. Note that analysis ignoring the

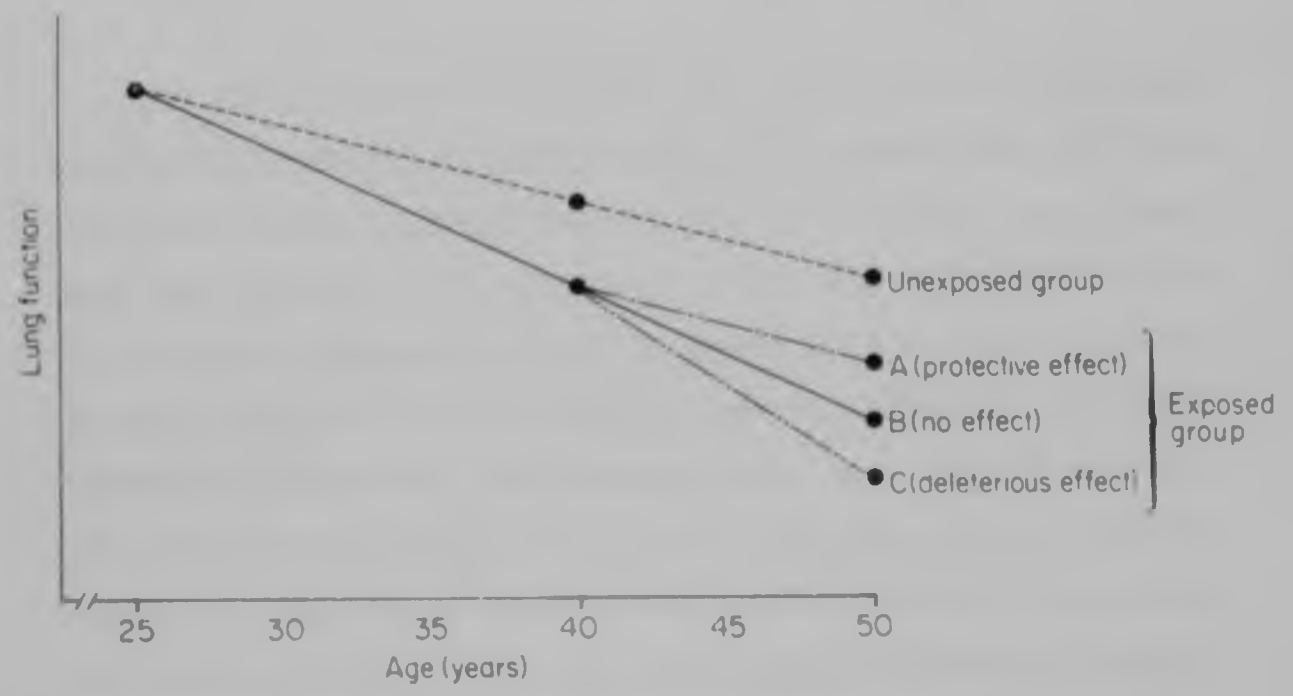

Figure 2-1 Lung function change with age in exposed and unexposed groups of men whose mean lung function differed at age 40. Exposure only starts after age 40.

confounder results in overestimation of the effect of exposure between 40 and 50 years of age. In Figure 2-1:

- Exposure would be found harmful when it has no effect (line B).
- Exposure would be found to have no effect when it actually protects men from lung function loss (line A).
- Exposure would correctly be considered deleterious, but the magnitude of effect overestimated (line C).

Conversely if the exposed group had better lung function than the unexposed group at 40 years of age, the effects of exposure on

subsequent lung function change would be underestimated.

The second scenario differs from the first only in that exposure to the environmental agent started at age 25 rather than after the initial lung function measurement at age 40. In this case, a lower mean lung function in the exposed men at age 40 could be entirely due to their prior exposure. Adjusting change for lung function at age 40 would essentially be adjusting for exposure. Therefore, while the unadjusted longitudinal data correctly show the effect of exposure, this would not be evident after adjustment for lung function at age 40. Clearly lung function at age 40 should not be considered a confounder in this case because it is not a determinant of subsequent decline. Rather it is a reflection of the effect of exposure on lung function change from 25 to 40 years which also causes the change from 40 to 50 years.

These two scenarios represent two extremes. In the first, where exposure only occurs after the initial lung function measurement, that measurement should be considered a confounder. In the second, where exposure occurs before the first lung function measurement, it should not be considered a confounder. Obviously, many real-life scenarios fall between the two extremes presented, depending on when the exposure occurred or other complexities not covered by the simple example and its attendant assumptions. Whether initial lung function should be considered a potential confounder needs to be considered separately for every study undertaken by considering what sources of

variation determine the initial lung function. If in doubt, it seems reasonable to regard initial lung function as a potential confounder because, in practice, exposure is never absolutely constant over time. Therefore, in the second scenario the effect of exposure should still be evident after adjustment for initial lung function, though the magnitude of effect will be underestimated. Adjustment for the initial lung function therefore constitutes a stringent test of causality.[49]

2.2 The relationship between lung function change, initial measurement and exposure when lung function is measured with random error

In Section 2.1.1, lung function was considered to be measured without error. However it is known that lung function measurement is prone to error.[30] The effects of random error of measurement in longitudinal data are analogous to those outlined in Section 1.5 and will be detailed further in chapter 3. However two points deserve special mention here.

1) The regression (or partial regression) of lung function change on initial lung function is biased. This effect is more complex than that outlined in chapter 1 because the first lung function is the predictor and forms part of the dependent variable, lung function change (defined as the second minus the first measurement). Because of this covariance, random measurement error results in a spurious negative association between lung function change and the first lung function.[40,50,51] Regression to the mean therefore causes a bias

opposite in direction to the positive association of the horse-racing effect described in Section 2.1.1.

2) Because of the incorrect partial regression of change on initial measurement, adjustment for the confounding effect of initial lung function results in a biased regression coefficient of lung function change on environmental exposure.

2.2 THE PRACTICE OF LONGITUDINAL LUNG FUNCTION DATA ANALYSIS: REVIEW OF THE LITERATURE

I have reviewed the literature on the association between exposure to environmental agents and lung function change to identify the analytic methods used. Each paper was assessed using the following questions:

1) Did the authors discuss whether initial lung function should be considered a confounder in their data?

2) Was initial lung function adjusted for, either explicitly or because the data were analysed in a way which implied a specific model of the relationship between change and initial level?

3) Did the analytic method take cognizance of the spurious negative association between change and initial level caused by random error of measurement?

In none of the papers was there any major discussion on whether initial lung function should be considered a confounder on conceptual

grounds (Question 1). In most papers exposure had commenced before the initial lung function measurement, although data were seldom given to quantify its duration. The analytic methods identified (Question 2) and whether the analysis took account of the effect of random error of measurement (Question 3) are discussed below. Many papers used more than one method to demonstrate whether their assessment of the effect of exposure was materially altered by the choice of analytic technique. These papers will be cited under only one of the methods used. The references are intended to be illustrative rather than exhaustive. As throughout the thesis, only two-occasion longitudinal studies are included.

2.2.1 Cross-sectional analysis of exposed and unexposed groups

Some papers presented lung function results on 2 occasions on the same men as if they arose from 2 separate groups.[52,53] This method does not make statistically efficient use of longitudinal data. However, it does at least demonstrate whether differences in lung function between exposure groups at the final examination were already present at the initial one.

2.2.2 Lung function change related to exposure

Examining the association of exposure and lung function change (difference between the final and initial measurement) is a common method of analysis.[35,36,54-59] The method assumes that initial lung function is not a confounder. This assumption is justified if the initial lung function is not a determinant of change or if the

mean initial measurement does not differ between exposure groups. None of the papers explicitly discussed whether initial lung function was a determinant of change. Many of the studies showed that initial measurements differed between exposure groups.

2.2.3 Lung function change regressed simultaneously on exposure and initial lung function

Lung function change was regressed simultaneously on initial measurement and exposure in three publications.[60-62] Though two of the papers[60,62] mention regression to the mean, they do not discuss whether the regression coefficients of change on exposure or initial measurement could be biased because lung function is measured with error.

2.2.4 Lung function change regressed simultaneously on exposure and the mean of the initial and final measurement

Kauffmann et al regressed lung function change on exposure and the mean of the initial and final height-ajusted lung function measurements.[63] This follows the suggestion of Oldham that the mean value provides a lung function level against which to regress change without obtaining spurious negative regressions because of random error of measurement.[50]

2.2.5 The ratio of lung function change to level or the ratio of final to initial measurement in relationship to exposure

In several papers the dependent variable was the ratio of change to the initial measurement[64,65] or to the mean of the initial and final measurement.[66,67] Other methods were the ratio of change to height^3[68] and of final to initial measurement.[69] Several of the

papers show different results for relative lung function change (the ratio method) and observed lung function change.[64,67] For example, Wilhelmsen found that smoking was more significantly associated with relative than absolute lung function change.[64] Though not made explicit in any of the publications, the ratio model specifies that the absolute magnitude of change is expected to be proportional to the initial level. In view of the horse-racing effect described in Section 2.1.1, this seems biologically unlikely.

2.2.6 Regression of final lung function measurement on exposure after adjusting for initial lung function measurement

Some authors[70] refer to a method suggested by Rosner.[71] He proposed a two-stage analytic procedure whereby the residuals of the final on the initial measurement are obtained and regressed on the exposure variable of interest. This method may be biased because the residual is calculated from the regression of final on initial measurement rather than the corresponding partial regression as required for the correct estimation of residuals. The problem is avoided by predicting the final lung function by the exposure variable, with the initial lung function as a covariate.[72] However neither method takes account of the effect of random error of measurement on the regression coefficient. Nor do they offer any advantage over regressing change on initial level, as will be shown in Chapter 3.

2.3 CONCLUSION

The literature review reveals a variety of methods used for analysing longitudinal (2-occasion) data. The conceptual issue of whether initial lung function is a potential confounder is not usually addressed. Different methods may result in different conclusions because differert explicit or implicit assumptions are made about the relationship between change and initial level. The only method which attempts to take account of the effect of random error of measurement is 2.2.4, in which the predictor used is the mean of the initial and final lung function measurements. In the next chapter I therefore present a variety of methods which could be used for obtaining unbiased regressions of change on exposure and initial lung function when lung function is measured with error. Thereafter, selected methods are applied as follows:

- In Chapter 4, the study design and methods which generated the data set for demonstration of the methods is described.
- In Chapter 5 the random measurement error in the data set is quantified as required for some of the methods given in Chapter 3.
- In Chapter 6 selected methods are applied.

The consequences of random measurement error in the dependent variable, lung function change, are explored in Chapter 7.

CHAPTER 3 METHODS OF CORRECTING FOR THE EFFECT OF RANDOM ERROR OF MEASUREMENT IN REGRESSION ANALYSIS OF LONGITUDINAL DATA

> Karl Friedrich Gauss's observatory at Göttingen was built about 1807. ... We look at the position of a star as it was determined then and now, and it seems to us that we are closer and closer to finding it precisely. But when we actually compare our individual observations today, we are astonished and chagrined to find them as scattered within themselves as ever. We had hoped that the human errors would disappear, and that we would ourselves have God's view. But it turns out that the errors cannot be taken out of the observations. ... Gauss pushed on to ask what the scatter of the errors tells us. He devised the Gaussian curve in which the scatter is summarised by the deviation, or spread, of the curve.
>
> J Bronowski[73]

3.1 EFFECT OF RANDOM ERROR OF MEASUREMENT ON REGRESSION COEFFICIENTS

Though often ignored in analyses of lung function data (Section 2.2) the consequences of random error of measurement are well described in the statistical and epidemiological literature.[39,42,74-77] Several papers deal specifically with random error in longitudinal data.[45,50,51,78-82] The effects of random error of measurement on regression coefficients are summarized in this chapter with particular reference to the relationship between lung function change and two predictors, namely initial lung function and exposure to an environmental agent. For the purpose of this chapter, initial lung function is considered a potential confounder of the relationship between environmental exposure and lung function change (Section 2.1.1).

The general linear additive model can be represented as:

$$L_2 - L_1 = \alpha + \beta_L L_1 + \beta_X X + \varepsilon \qquad (3:1)$$

where: L_1 = the true value or 'level' (i.e. measured without random error) on the first occasion

L_2 = the true value on a second occasion, for example ten years later

α = a constant

X = the measurement of exposure (without error) to an environmental agent between the first and second occasion

β = the regression coefficient for each of the independent variables as indicated by its subscript

ε = the residual term, i.e. a random deviation from the linear model

Note that change is measured as the second minus the first lung function measurement. A negative value therefore indicates lung function loss. The model can be used to examine the relationship between change (L_2-L_1) and an environmental agent (X) after taking into account any dependence of change on initial level (L_1). If there is exposure to more than one environmental agent this can be represented by the inclusion of further terms (i.e. X_1, X_2, X_3 ... X_n).

The model can also be expressed as:[78]

$$L_2 = \alpha + (1+\beta_L)L_1 + \beta_X X + \varepsilon \qquad (3:2)$$

The alternative representations of the relationship between lung function change and initial level in equations 3:1 and 3:2 are shown

graphically in figure 3-1. Change is not associated with initial level if the regression coefficient β_{L} is zero (Figure 3-1a), in which case the regression coefficient of L_2 on L_1 (i.e. $1+\beta_{L}$) will be 1 (Figure 3-1b).

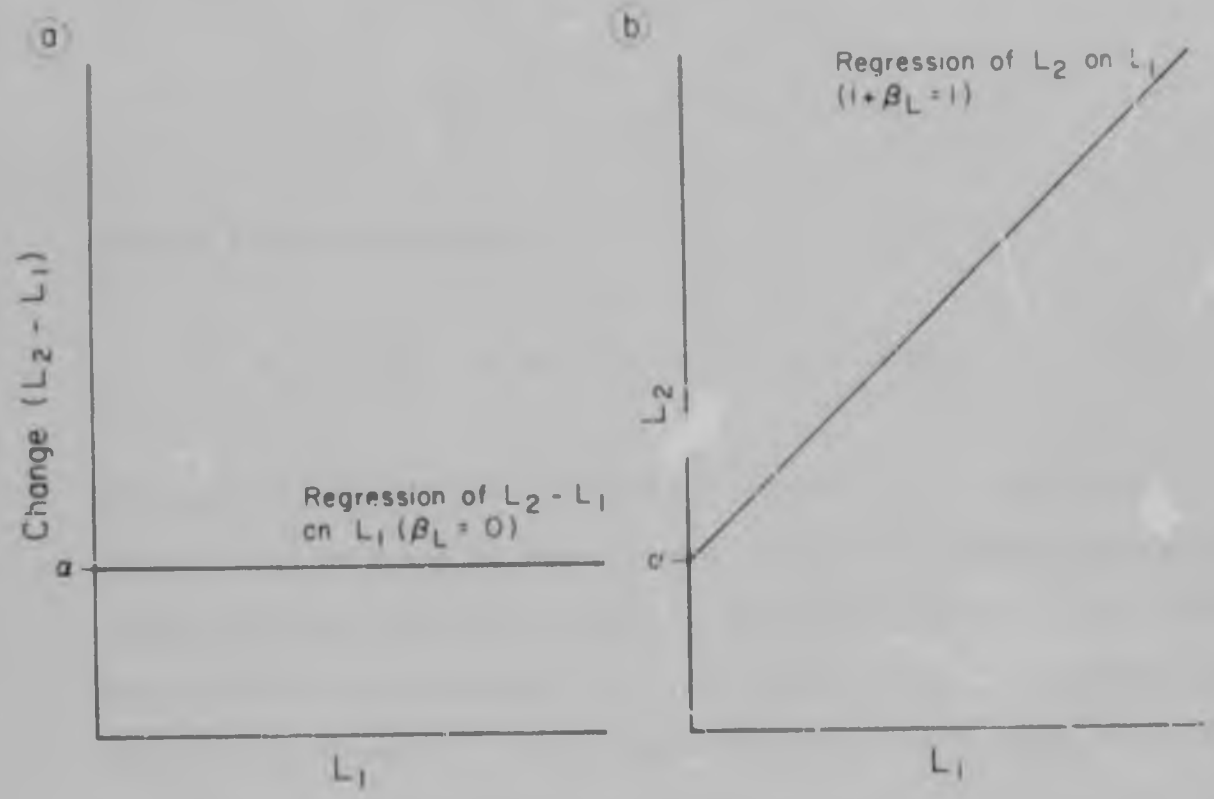

Figure 3-1 Graphical representation of alternative ways of expressing the model

In practice, L_i (i = 1 or 2) is usually observed with random error (δ_i) resulting in a measurement M_i, where:

$$M_i = L_i + \delta_i.$$

The model from equation 3:1 then becomes:[45]

$$M_2 - M_1 = \alpha + \beta_L M_1 + \beta_X X + \varepsilon - \delta_1 + \delta_2 - \beta_L \delta_1 \qquad (3:3)$$

alternatively expressed as:

$$M_2 = \alpha + (1+\beta_L)M_1 + \beta_X X + \varepsilon + \delta_2 - (1+\beta_L)\delta_1 \qquad (3:4)$$

The residual now consists of the sum of several terms, which will result in an increase in the standard errors of estimated regression coefficients and therefore a loss of statistical power. Some other implications of an increase in the variance of the residuals are explored in chapter 7. More important for present purposes is the fact that the error terms now include δ_1 and $\beta_L\delta_1$ which are not independent of M_1 and therefore violate the necessary assumptions of regression analysis.[73] To explore this further I shall use equation 3:4 as the preferred representation as it reduces longitudinal data analysis to a simple example of the errors-in-variables regression model.[76] The bias in the estimated coefficient $\widehat{1+\beta_L}$ when L_1 is measured with random error will be considered first for the case where L_1 and X are not correlated and then for the case where they are correlated.

3.1.1 Effect of random measurement error when L_1 and X are not correlated

If L_1 and X are not associated the coefficient $1+\beta_L$ is defined as $\sigma_{L_2L_1} / \sigma_{L_1}^2$. When M_i is used instead of L_i, the use of $\sigma_{M_1M_2} / \sigma_{M_1}^2$ leads to a biased estimate $\widetilde{1+\beta_L}$. As the random measurement error terms δ_1 and δ_2 are by definition uncorrelated, substituting M_i for L_i does not alter the expected covariance, i.e. $\sigma_{M_1M_2} = \sigma_{L_1L_2}$.[42] Therefore the expected value of $\widetilde{1+\beta_L}$ will be smaller than $1+\beta_L$, because $\sigma_{M_1}^2 > \sigma_{L_1}^2$. Assuming that L_1 and δ_1 are independent and normally distributed, $\sigma_{M_1}^2$ is equal to $\sigma_{L_1}^2 + \sigma_{\delta_1}^2$. The ratio of the variance of true values ($\sigma_{L_1}^2$) to that of measurements ($\sigma_{M_1}^2$) is referred to as the reliability or generalizability coefficient (G_{M_1}),[39,42,45,74,78,84] which may be written as:

$$G_{M_1} = \frac{\sigma_{L_1}^2}{\sigma_{M_1}^2} = \frac{\sigma_{M_1}^2 - \sigma_{\delta_1}^2}{\sigma_{M_1}^2} = 1 - \frac{\sigma_{\delta_1}^2}{\sigma_{M_1}^2} \qquad (3:5)$$

It follows that the expected value of $\widetilde{1+\beta_{L_1}}$ is equal to:

$$G_{M_1}(1+\beta_L) \qquad (3:6)$$

The implications of expression 3:6 for longitudinal data are easily demonstrated by a simple example in which change is not associated with initial level, i.e. $\beta_L = 0$ or $1+\beta_L = 1$. In the presence of measurement error, G is less than 1. Therefore, the expected value of $\widetilde{1+\beta_L}$ will be less than 1 and hence the expected value of $(\widetilde{1+\beta_L})-1$ will be less than zero. In other words, random error of measurement induces a spurious negative regression of change on initial level, the phenomenon known as regression to the mean.[40,41,50,51]

Although in this thesis I consider only error of measurement of lung function, the same principles apply to measurement error of the exposure variable. Therefore, by analogy with $1+\beta_L$, the estimate of β_X may be attenuated, i.e. its absolute value reduced, if X is measured with random error.

3.1.2 Effect of random measurement error when L_1 and X are correlated

In the example given above, in which L_1 and X are uncorrelated, the estimate of β_X is not biased by measurement error in L_1. On the other hand, if L_1 and X are correlated, random error of measurement in L_1 may result in a biased estimate for β_X as shown below. By analogy, β_L may be biased if X is measured with error. The true (partial) regression coefficient for X after taking into account the dependence of L_2 on L_1 can be shown as:(78)

$$\beta_{X|L_1} = \frac{\sigma_{L_1}^2\sigma_{XL_2} - \sigma_{XL_1}\sigma_{L_1L_2}}{\sigma_X^2\sigma_{L_1}^2 - \sigma_{XL_1}^2} \qquad (3:7)$$

The expected covariance terms are unaltered by random error of measurement, because of the absence of association between error and true values. However, if X or L_1 are measured with error, their variances are overestimated. As an example of how important this may be, envisage the situation where only L_1, and not X, is measured with error. $\sigma_{L_1}^2$ will be replaced in both the numerator and the

denominator of the equation by the larger variance $\sigma_{M_1}^2 = \sigma_{L_1}^2 / G_{M_1}$ A simple illustration of how this may bias $\tilde{\beta}_{X|L_1}$ is evident if one remembers that the sign of the regression coefficient is determined by the numerator of equation 3:7. This may truly be negative, but found to be positive if $\sigma_{L_1}^2\sigma_{XL_2}$ is inflated by replacing $\sigma_{L_1}^2$ by $\sigma_{M_1}^2$. This is an example of underadjustrent for a confounder because it is measured with error as first suggested in section 1.5.2. If only X is measured with error the denominator of equation 7 is increased with resultant attenuation of $\tilde{\beta}_{X|L_1}$. The effect of error in both X and L cannot easily be predicted. Note that $\sigma_{L_2}^2$ does not appear in the equation and therefore that error in the dependent variable does not bias the estimate of the regression coefficient, though it will increase its standard error.

3.2 THE CALCULATION OF REGRESSION COEFFICIENTS USING RELIABILITY COEFFICIENTS TO CORRECT FOR RANDOM ERROR OF MEASUREMENT

The effects of random error of measurement on regression coefficients can easily be corrected by replacing the estimated variance of measurements of explanatory variables (e.g. $\hat{\sigma}_{M_1}^2$) by the estimate of the true variance of the variable (e.g. $\hat{\sigma}_{L_1}^2$). This can be done if the reliability coefficient (G) for the measurements is known because (using the same example), $\hat{\sigma}_{L_1}^2 = G_{M_1}\hat{\sigma}_{M_1}^2$. The reliability coefficient, G, can easily be estimated from repeat measurements of the relevant variables, as will be shown in chapter 5. An equivalent method is to subtract the known random error variance ($\hat{\sigma}_\delta^2$) from $\hat{\sigma}_{M_1}^2$.(76)

3.2.1 Correcting regression coefficients when X and L_1 are not correlated

If X and L_1 are not correlated, an unbiased estimate of the regression coefficient for X and L_1 can be obtained. As might be expected from expression 3:6, the unbiased estimate $\widehat{1+\beta}_L$ is obtained by dividing the calculated regression coefficient by the relevant reliability coefficient[80] i.e.:

$$\widehat{1+\beta}_L = \frac{\widetilde{1+\beta}_L}{G_{M_1}} \qquad (3:8)$$

The corrected β_L is easily obtained by subtracting 1 from $\widehat{1+\beta}_L$. Note that the absolute value of the regression coefficient for a predictor will always be underestimated or attenuated because of measurement error. The correction of the estimate of the coefficient by dividing it by G must give a larger absolute value, the procedure being referred to as deattenuation. However, if $\widetilde{1+\beta}_L$ is less than 1, the 'attenuated' $\widetilde{\beta}_L$ will be negative with a greater absolute value than the 'deattenuated' estimate. The terms attenuation and deattenuation are therefore inappropriate for $\widetilde{\beta}_L$. The terms 'bias' and 'correction' are preferred for longitudinal data.

The regression coefficients for X and L estimated from samples can be corrected for random error of measurement by the method outlined above using estimates of G ($\hat{G}$) for that variable. The

variance of a corrected regression coefficient is estimated approximat·ly by:(51)

$$\hat{\sigma}_{\hat{\beta}}^2 = \hat{\beta}^2 \left[\frac{\hat{\sigma}_{\tilde{\beta}}^2}{\tilde{\beta}^2} + 2 \left(\frac{1-\hat{G}}{\hat{G}} \right)^2 \left(\frac{1}{n-1} + \frac{1}{n_1-1} \right) \right] \qquad (3:9)$$

where $\tilde{\beta}$ = biased estimate and $\hat{\beta}$ = corrected estimate of β (β may be β_X or $1+\beta_L$)

n = sample size for calculation of $\tilde{\beta}$

n_1 = number of pairs of measurements for calculation of $\hat{G}$

As n and n_1 approach infinity, the variance of the corrected regression coefficient decreases to $\hat{\sigma}_{\tilde{\beta}}^2/\hat{G}$. Therefore, $\hat{\sigma}_{\hat{\beta}}^2$ contains all the components of variance included in $\hat{\sigma}_{\tilde{\beta}}^2$ inflated by a factor of $1/\hat{G}$. Because $\hat{\beta} = \tilde{\beta}/\hat{G}$, tests of the hypothesis that $\beta = 0$ will be unaltered or found less significant after correction. However, in the special case where the β of interest is $1+\beta_L$, the hypothesis one wishes to test is whether change is independent of initial level, i.e. whether $\beta_L=0$ or $1+\beta_L=1$. If the biased estimate of β_L is negative, the corrected estimate will approach zero and its standard error will increase. Both of these effects make rejection of the hypothesis more likely.

3.2.2 Correcting regression coefficients when X and L_1 are correlated

When X and L_1 are correlated, estimates of partial regression coefficients can also be corrected. This is achieved by multiplying the variance of each variable by its reliability coefficient or subtracting the error variance for that variable. The corrected

variance estimates are then used in the equation for the calculation of partial coefficients, such as equation 3:7. Alternatively, the variance of each variable in a variance-covariance matrix can be corrected and the corrected matrix used to calculate regression coefficients.[76] The variance of estimated regression coefficients are also obtainable. For example, when L, but not X, is measured with error, approximate formulae[76] can be shown to be:

$$\sigma^2_{\hat{\beta}_L} = \sigma^2_{\widehat{1+\beta}_L} = \frac{\sigma_v^2}{n\sigma^2_{L_1}\hat{C}_{M_1}} + \left[\frac{1-\hat{C}_{M_1}}{\hat{C}_{M_1}}\right]^2 \left(\widehat{1+\beta}_L\right)^2 \left(\frac{1}{n} + \frac{2}{n_1-1}\right) \qquad (3:10)$$

$$\sigma^2_{\hat{\beta}_X} = \frac{\sigma_v^2}{n\sigma^2_X} \qquad (3:11)$$

where σ_v^2 = the variance of the residuals around the corrected regression line.

Obviously the method for correlated L_1 and X can also be used if these variables are uncorrelated. In this case the regression and partial regression coefficients are identical and the β estimates derived from the two correction procedures must be identical. The two approximate formulae for the calculation of the variance of the regression coefficients (Equations 11 and 10) can be shown to be equivalent (assuming n equal to n-1) if there is only one predictor, say L_1. This follows from the fact that:

$$\sigma_v^2 = \sigma^2_{M_2|M_1} + \sigma^2_{M_1}\left((\widehat{1+\beta}_L) - (\widehat{1+\beta}_L)\right)^2$$

$$= \sigma^2_{M_2|M_1} + \hat{\sigma}^2_{M_1}(1-\hat{G})^2(\widehat{1+\beta}_L)^2 \qquad (3:12)$$

where $\sigma^2_{M_2|M_1}$ = the residual or conditional variance of M_2 given M_1

Alternative formulae when G is estimated on a totally independent data set give slightly smaller values of $\theta_{g_L}^2$ and are not considered further in this thesis.(77)

The method of correcting regression coefficients for correlated predictors is the preferred method for general use for two main reasons:

1) A determinant of the dependent variable is a confounder unless the correlation between that variable and the predictor of interest is zero in that sample. Because the distortive effect of the confounder depends on the sample correlation coefficient, any deviation from zero may be important, even if not statistically significant.(85) This correlation is unlikely to be precisely zero. The easiest way of assessing whether confounding is important is to examine whether the partial regression coefficient is altered by including the potential confounder in the model.

2) The method is generalizable to include several explanatory variables (e.g. X_1, X_2, X_3 ... X_n) (each of which may be measured with error).

3.3 OTHER METHODS OF CORRECTING FOR THE EFFECT OF RANDOM MEASUREMENT ERROR ON THE REGRESSION OF L_2-L_1 ON L_1

Apart from the use of the reliability coefficient, the literature contains several other methods of correcting for the effect of

measurement error in longitudinal data.[81] All of them are based on the principle of substituting the predictor M_1 by some other measure of L_1. All are inferior to the methods using the reliability coefficient. One disadvantage is that there is no way of simultaneously correcting for measurement error in the exposure variable (X). Further disadvantages will be described separately for each method.

3.3.1 Substituting an independent measure of L_1[79]

One can obtain two measures of L_1 (M_1 and M_{1a}) on occasions close enough to ensure that they both measure the same underlying level, but for which the random error of measurement of M_1 and M_{1a} are independent. If the error of M_{1a} is δ_{1a}, equation 3:3 now becomes:

$$M_2 - M_1 = \alpha + \beta_L M_{1a} + \beta_X X + \varepsilon - \delta_1 + \delta_2 - \beta_L \delta_{1a} \qquad (3:13)$$

This solves part of the problem of the association of residual terms and explanatory variables as δ_1 is independent of M_{1a}. However, $\beta_L \delta_{1a}$ is still associated with M_{1a} and the estimate of β_L obtained from the regression analysis should be corrected using the G of M_{1a}. In contrast to the situation when M_2-M_1 is regressed on M_1, the absolute value of the β_L estimate will <u>increase</u> with correction, because M_2-M_1 is obtained independently from the predictor M_{1a}.

3.3.2 Substituting L_1 predicted from other variables[45]

Rather than using an independently measured M_{1a}, one can use an M_{1a} predicted by a set of instrumental variables. These are defined

as variables which are correlated with the predictor of interest but independent of its random error of measurement. For example, one can predict some lung function test results by height, and use the predicted lung function value as M_{1a} in equation 3:13. However, even if height is measured with negligible error, the method has several disadvantages. First, the predicted value excludes the component of variance due to biological variability unrelated to height. This component may have a different relationship to lung function change than that of the height-related component. Second, the variance of the predicted M_{1a} will be smaller than that of L_1. This results in a large standard error of the β_L estimate.

When X and L are uncorrelated, an unbiased regression coefficient can be calculated using an instrumental variable (H), as follows:[86]

$$\widehat{1+\beta_L} = \frac{\hat{\sigma}_{M_2H}}{\hat{\sigma}_{M_1H}} \qquad (3:14)$$

A related method is to calculate the regression coefficient using only those explanatory M_1 values which appear in low and high value groups selected using a grouping method which is independent of the errors of M_1.[86,87]

3.3.3 Substituting $(M_1+M_2)/2$ for L_1[50]

The regression of M_2-M_1 on $(M_1+M_2)/2$ is a method of longitudinal data analysis which is in use in the lung function literature.[63] The method is based on the following concept. If the variance of

random measurement error of a single observation is σ_δ^2, that of the mean of n observations in σ_δ^2/n, assuming independent errors on the two occasions.[39] The reliability coefficient (G) of the mean of n observations increases as n increases:

$$G = \frac{\sigma_L^2}{\sigma_L^2 + \frac{\sigma_\delta^2}{n}} \qquad (3:15)$$

This equation demonstrates the value of multiple measurement of a predictor. However, the method described uses the mean of M_1 and M_2 rather than repeat measurements at the time M_1 is measured.

Equation 3:3 then becomes:

$$M_2 - M_1 = \alpha + \beta_L\left(\frac{M_1+M_2}{2}\right) + \beta_X X + \varepsilon - \delta_1 + \delta_2 - \beta_L\left(\frac{\delta_1+\delta_2}{2}\right) \qquad (3:16)$$

The method partly takes account of the association between the predictor and the residual because the variance of the $\beta_L\left(\frac{\delta_1+\delta_2}{2}\right)$ term will be less than that of $\beta_L\delta_1$ alone. The predictor is also less associated with δ_1 but is now associated with δ_2. However, the most important defect of the model is that it is biologically incorrect. It incorporates a component of change in the $(M_1+M_2)/2$ term on which β_L is estimated.[81] This results in overestimation of α_L and the possibility of showing a spurious positive association between lung function change and initial level. β_X may then also be incorrectly estimated. For example, if X is not associated with M_1 but is truly associated with change, it will be associated with M_2 and therefore with $(M_1+M_2)/2$. The β_L estimate using the model in equation 3:16 will include part of the effect

of X. Therefore β_X will be underestimated. This is easily demonstrated by analysis of the hypothetical data in table 3-1. The data are free of random error of measurement and not intended to represent lung function data.

Table 3-1 Hypothetical data for X, M_1 and M_2 (all measured without error)

X	M_1	M_2
0	40	40
1	40	45
0	40	60
1	40	65
0	40	80
1	40	85
0	100	100
1	100	105
0	100	120
1	100	125
0	100	140
1	100	145

The least squares regression equations calculated from this data set are:

$$M_2 - M_1 = 20 + 5.00X + 0.00L_1$$

(using the error-free M_1 as the measure of L_1)

and $$M_2 - M_1 = 8.97 + 4.66X + 0.14L_1$$

(using $(M_1+M_2)/2$ as the proxy of L_1).

Whether this effect is important in practice will be explored in chapter 6.

3.4 CONCLUSION

Random error of measurement in an explanatory variable may cause bias in estimates of regression coefficients for that variable and others in the model. Methods exist for obtaining unbiased coefficients. The best method is that which uses the reliability coefficient to correct the variance of measurements for the component of measurement error. This method is not in general use in the analysis of lung function change in relationship to initial level and exposure to environmental agents. It will be applied to a particular data set in chapter 6 and compared in practice with the other methods outlined above. But first, the origin of the data set is described (Chapter 4) and the reliability coefficient is estimated for lung function measurement in the data set (Chapter 5).

CHAPTER 4 THE LONGITUDINAL LUNG FUNCTION DATA SET: STUDY METHODS AND DATA STRUCTURE

4.1 STUDY METHODS

4.1.1 The data source

The longitudinal lung function data used in the next three chapters derive from the Medical Bureau for Occupational Diseases in Johannesburg. This statutory body is responsible for routine monitoring of the health of mineworkers. The predominant mining activity is gold-mining which occurs in the Southern Transvaal highveld, situated more than 1 500 metres above sea-level. The bulk of the mining is done at depths of between 1 and 2 kilometres.(88) Other major mines in this region include coal and platinum. All men who intend to work in dusty occupations must have a medical examination to ensure their fitness. More than 10% of applicants are rejected.(89) Examinations are repeated six months later and then annually. In the mid-1960s, the late Dr Helen van Doorn saw the opportunity of using these annual examinations for collecting lung function data which might be of value for surveillance or screening. She therefore arranged for white miners to have spirometric and peak exploratory flowrate measurements every few years, starting with their first routine examination six months after they were declared medically fit to work in the mining industry. The information collected is referred to as the van Doorn data set in this thesis.

4.1.2 The sample

The sample was assembled from men who had their first routine examination between 1964 and 1966. Sampling was done by clerical staff on the basis of whether lung function laboratory personnel and equipment were available to test them, and without knowledge of the men's medical examination results. This convenience sample of 1 974 men represented about 14% of all new miners.(89) Originally, an attempt was made to re-measure as many men as possible at each annual examination. This proved infeasible and lung function testing was attempted every 3-4 years. During 1976 and 1977, about 11 years after the initial examination, all men in the original sample were retested if they had not recently been examined and were still working in the mining industry in the region.

It is realized that the initial sampling method was far from ideal. Nevertheless, the study has a major advantage. It provides longitudinal lung function data over 11 years on a large sample of young men from the time of entry into the mining industry.

4.1.3 Measurement of lung function, height and weight

Spirometry was performed on one of four Godart water sealed spirometers using methods in reasonable accord with recommendations subsequently published by the American Thoracic Society.(90) The same instruments were used throughout the study period. The spirometers were of 9 litre capacity with breathing resistance of

about 4 mm H_2O at 100 litres/minute. The mouthpiece was connected by about half a metre of 25 mm tubing. Spirometers were checked daily for leaks and monthly for the accuracy of time and volume measurements. The timer was checked using a stopwatch and volume using 1 and 4 litre syringes over the range of movement of the bell. Spirometry was performed by any of the 9 technicians employed at the time. The subject was seated during the manoeuvre and a rubber mouthpiece and noseclip were used to avoid leakage of air. The forced expiration was recorded at 20 mm/sec and the recording only stopped after the individual had commenced his next inhalation. Forced expiration manoeuvres were repeated until three tracings were found acceptable by the technician. Criteria of acceptance were that the subject must have appeared to use maximal effort and that the expiratory curve was smooth with a reasonably prompt initiation. The Forced Vital Capacity (FVC), Forced Expiratory Volume in 1 second (FEV) and Forced Expiratory Flow in the middle half of expiration (FEF) were measured on each of the three acceptable tracings, or on a smaller number of tracings in those few instances where three could not be obtained. To obtain FEV the start of expiration was timed from a point determined by back extrapolation.[90] FEF was measured by dividing the FVC into quarters, measuring the volume of air expired in the middle half of the FVC and dividing it by the time taken for this volume to be expired. The numerically highest values of FVC, FEV, and FEF were accepted as the value for each man, irrespective of whether they came from the same or different tracings. All measurements were adjusted to body temperature and pressure, saturated

with water vapour. The techniques used for FVC and FEV were therefore similar to those subsequently recommended.[90] However FEF was not obtained as recommended from the tracing with the largest sum of FVC and FEV. The procedure used in the van Doorn data set tends to overestimate FEF by about 5%.[91]

Peak exploratory flow rate was measured on one of 5 calibrated Wright Peak Exploratory Flow Meters after spirometry had been completed.[92] Peak flow was measured with the subject standing. The highest of 3 blows judged acceptable by the technician was recorded as the value for each man.

One way distributions of lung function data were examined prior to analysis to identify extreme outliers, i.e. individuals whose change in lung function from one to any other occasion was more than 4 standard deviations away from the mean change over that time period. The original records of these individuals were checked for transcription or coding errors and corrected if necessary. If no such errors were found, these individuals were excluded from analysis. The proportion of the sample excluded for this reason or because of incomplete data on this or any other variable was less than 1%.

Note that in tables in the thesis the units of measurements are omitted. They are given below for reference.

Spirometric tests: FVC - litres

FEV - litres

FEF - litres per second

FEV%, - derived from (FEV/FVC) x 100

PEFR on the Wright Peak Flow Meter - litres per minute.

Height and weight were measured whenever lung function was tested. Height was measured to the nearest inch (in earlier years) or centimetres (in later years) on an anthropometer with the subject erect and the head in the Frankfort plane. The subjects did not wear shoes. Height measurements did not change significantly between the first and last examination, with the exception of 23 men less than 18 years of age who were subsequently excluded from all analyses. The mean of the first and last measurement was taken as the best estimate of height for each individual. Subjects were weighed on a standing scale in light indoor clothing to the nearest pound (in earlier years) or kilogram (in later years). The measurement was included on the computer file for each occasion at which lung function was measured. Height is given in centimetres and weight in kilograms in the tables in this thesis.

4.1.4 Measurement of smoking and occupational exposure

At each annual examination, men were asked by the examining doctor about their smoking habits and occupation during the previous year. Cigarette smokers were asked about the number smoked per day and pipe smokers about the number of ounces of tobacco smoked per week. Cigar

smoking was recorded as the number of small and large cigars per week and converted to the number of ounces of tobacco per week, assuming 5 large or 12 small cigars per ounce. The tobacco smoked in pipes and cigars were not differentiated from each other on the computer file. The number of ounces of non-cigarette tobacco consumed per week was converted to grams per day prior to analysis. The average cigarette and tobacco consumption per year was then calculated for each man by cumulating the amount recorded at each annual examination over the years between lung function tests and dividing by that number of years.

The occupational history consisted of information about the type of mine or other job in which the individual worked during each year. Underground gold miners were classified as working in one of 11 job-specific categories. In occupations in other mines and industries job-specific exposures are not known. Generally, dust levels are much lower in other mines than for high-dust gold mining jobs (R du Toit, Air Quality Section, Government Mining Engineer, Personal communication). For purpose of analysis two types of indices of dust exposure in mines were used:

1) The number of years spent in underground high dust jobs. Dust levels for these jobs were about twice as high as other underground or surface jobs.[93,94]

2) The calculated average annual dust exposure. For this calculation the following three dust measurements were used:

a) Particles per ml air as measured on the thermal precipitator and counted before acid treatment. This is a measure of all particles of any size irrespective of whether they are organic or inorganic.

b) Particles, per m³ air, of size 0.5-5 micron in diameter after acid treatment. This is a measure of inorganic particles of respirable size.

c) Particles, per ml air, as measured by the konimeter after acid treatment. This measures inorganic particles which, because of the sampling characteristics of the konimeter, are approximately between 0.25 and 5 microns in diameter. It therefore includes smaller particles than b).

Estimates of job-specific dust exposure in the 1960s were obtained from exposure information, using personal samplers, weighted by the number of hours spent in a shift in each job.[93,94] Overall mine dust measurements have decreased since 1960 by about 2.56% per annum (R du Toit; Personal communication). Job and year-specific dust exposure for each man was estimated taking account of this decrease. Average dust exposure over the time between lung function tests was calculated by summing annual exposures over the years and dividing by that number of years.

4.2 DATA STRUCTURE

There were 1 974 men who had initial lung function tests. Only 1 077 of them had 2 or more tests and therefore could be included in longitudinal analysis. Figure 4-1 shows how these data were used with the purpose of:

1) Analysing how long-term change in lung function is determined by initial level and environmental exposure.

2) Obtaining an estimate of the reliability coefficient (G) for correction of regression coefficients in 1).

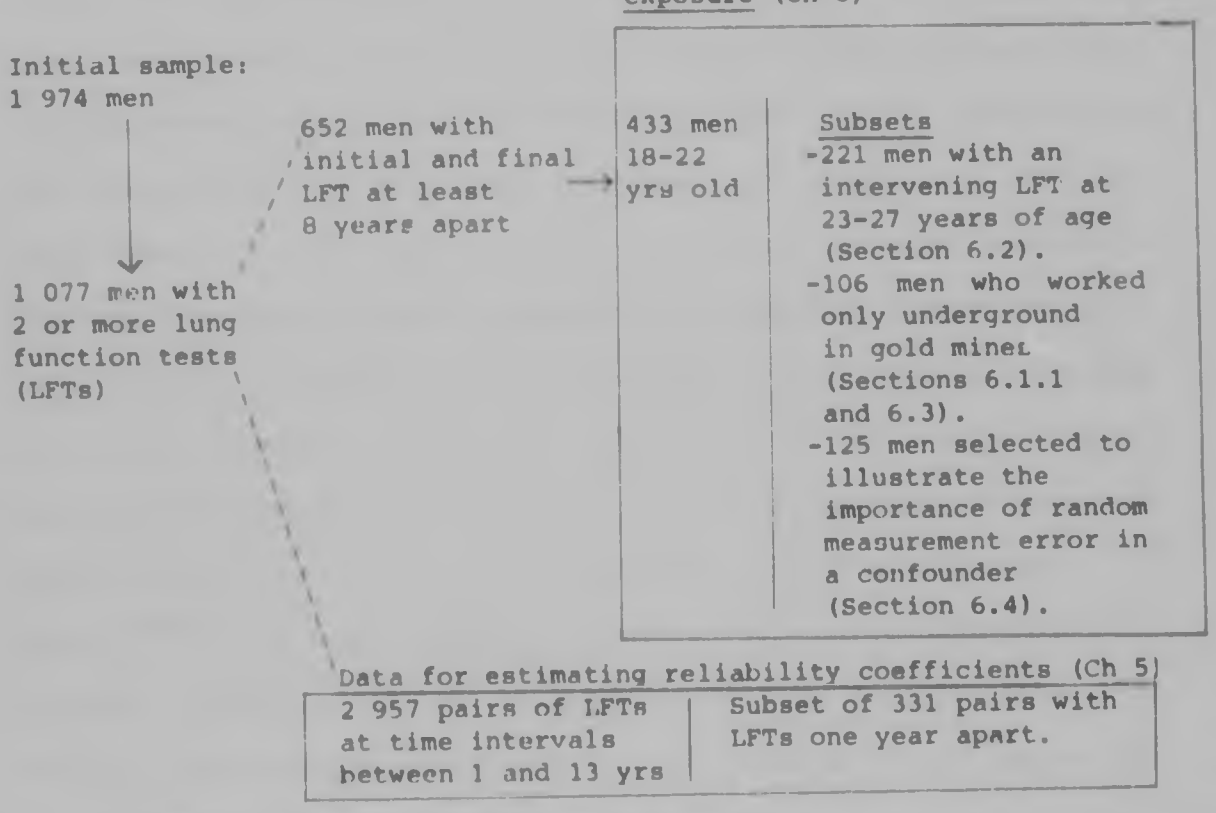

Figure 4-1 Sources of data used in analysis

4.2.1 Data for analysis of factors associated with long-term lung function change

There were 1 974 men in the initial sample. Only 652 of the 1 077 men with 2 or more tests remained in the mining industry long-term, defined as having their final lung function measurement at least 8 years

after the initial measurement. The average time between initial and final lung function tests on these men was 11.3 years with 79% of men having between 10 and 12 years of follow-up. Of the 652 men with initial and final examinations 8 or more years apart, 17% had no intervening lung function tests, 32% had one, 41% had two and 10% had three intervening tests. Even if linear change is assumed, one measurement approximately midway between the initial and final tests does not materially alter the estimate of the slope of linear change of lung function over time. Therefore only about half of the 652 men had useful data (2 or more tests) in addition to the initial and final test results. It was therefore felt that the most reasonable approach to analysis was to examine the change from initial to final lung function, ignoring intervening tests. This approach was made still more necessary by the age distribution of the sample. About two-thirds of the men were less than 23 years of age and the change in lung function over the subsequent decade was likely to be curvilinear[46,95] and require many data points for accurate characterization. The analyses of determinants of lung function change in chapter 6 was therefore restricted to individuals in the narrow age interval of the greatest biological interest, i.e. the 433 men aged between 18.00 and 22.99 years of age. Analyses were also done on various subsets of the data on these men, as shown in Figure 4-1. The reason for and definition of these subsets is presented in later chapters, as shown in the figure.

These 433 men with long-term follow-up represent only about one third of all 18-22 year-olds who had initial examinations. Concern about their representativeness is reduced by two findings. First, over half of the men who did not remain in the mining industry long enough to qualify for inclusion, had left within two years of their initial examination. Within so short a time it seems unlikely that their leaving was related to the development of respiratory impairment. Second, at the initial examination there were no major differences in the characteristics of the 433 men who remained in the industry and those who left (Table 4-1).

Table 4-1 Characteristics of 18-22 year-old men who remained in the mining industry or left

Characteristics	Mean (± SD) of characteristics in: Men in the industry N = 433	Men who left N = 841	p
Age	20.2 (1.2)	20.4 (1.3)	<0.01
Height	176.8 (5.9)	177.6 (6.3)	<0.05
Weight	70.7 (8.1)	71.0 (7.7)	NS
FVC	5.57 (0.61)	5.60 (0.63)	NS
FEV	4.68 (0.52)	4.70 (0.55)	NS
FEF	5.49 (1.23)	5.47 (1.23)	NS
FEV%	84.3 (6.4)	84.1 (6.2)	NS
PEFR	575 (67)	585 (67)	<0.05
Percentage of men who smoked	61	65	NS

NS Not significant ($p > 0.05$)

4.2.2 Data for estimation of reliability coefficients

There were 1 077 men who had two or more lung function tests. Because some men had more than two measurements, they generated 2 957 pairs of lung function measurements with time intervals ranging from 1 to 13 years. These data, in particular the subset of 331 men with examinations one year apart, were used to estimate the reliability coefficient of lung function data as is now described in the next chapter.

CHAPTER 5 ESTIMATING RELIABILITY COEFFICIENTS FOR LUNG FUNCTION MEASUREMENTS

As discussed in chapter 3, the estimated regression coefficients of lung function change on initial lung function and environmental exposure may be biased if these predictor variables are measured with random error. However, unbiased regression coefficients can be obtained by multiplying the estimated variance of each predictor by its reliability coefficient (G) and using these corrected variances to calculate regression coefficients. The purpose of this chapter is to obtain an estimate of G for each type of lung function measurement in the data set described in chapter 4. Unfortunately, reliability coefficients for variables measuring environmental exposure cannot be obtained from the available data.

5.1 CONCEPTS FOR ESTIMATING RELIABILITY COEFFICIENTS

5.1.1 Methods of estimating G

The reliability or generalizability coefficient (G) is defined as the ratio of the true variance of a variable to the variance of the variable measured with error (Section 3.1).[39,42,45,74,78,84]

$$G = \frac{\sigma_L^2}{\sigma_M^2} = \frac{\sigma_M^2 - \sigma_\delta^2}{\sigma_M^2} = 1 - \frac{\sigma_\delta^2}{\sigma_M^2} \qquad (5:1)$$

where L = true value for the variable

δ = random error of measurement

$M = L + \delta$

The σ_δ^2 required for calculation of G is easily estimated from repeat independent measurements (M_A and M_B) of the same true value, L. The <u>within-person variance</u> ($\sigma^2_{M_A-M_B}$) is twice the variance of the random error of measurement[96] because:

$$\sigma^2_{M_A-M_B} = \sigma^2_{M_A} + \sigma^2_{M_B} - 2\sigma_{M_A M_B} \qquad (5:2)$$

Now, because the variance of random measurement error is the same on both occasions, $\sigma^2_{M_A} = \sigma^2_{M_B} = \sigma^2_M$.

Assuming that δ is independent of L and that δ_A is independent of δ_B the covariance term is the same whether derived from M or L,[42] i.e:

$$\sigma_{M_A M_B} = \sigma_{L_A L_B} = \sigma_L^2$$

It follows that $$\sigma^2_{M_A M_B} = 2(\sigma_M^2 - \sigma_L^2)$$

$$= 2\sigma_\delta^2 \qquad (5:3)$$

Therefore, $\frac{1}{2}(\sigma^2_{M_A-M_B})$ can be substituted for σ_δ^2 in the formula for G (Equation 5:1).

A simple alternative is to use the correlation of M_A and M_B as the reliability coefficient[74] because:

$$\rho_{M_A M_B} = \frac{\sigma_{M_A M_B}}{\sigma_{M_A}\sigma_{M_B}}$$

which for reasons outlined under equation 5:2, simplifies to

$$\rho_{M_A M_B} = \frac{\sigma_L^2}{\sigma_M^2} = G \qquad (5:4)$$

An estimate of G ($\hat{G}$) can then be obtained from a sample correlation coefficient.

5.1.2 Components of variance included in the definition of random error of measurement

Sources of short-term variation of lung function measurements in an individual are well-documented in the literature.[97,98] First, there may be variations in the technique of measuring lung function. This includes variability of the equipment, the technician's instructions or methods of measurement, and the use of differing procedures for obtaining or selecting results. Reduction of the variability related to techniques is the objective of standardization.[90] Second, there may be short-term biological variation. Variability may be due to the subject's learning or fatigue, or bronchoconstriction induced by repeated forced test manoeuvres. Moreover, lung function may differ on two occasions several days or weeks apart. Some of this variation is known to be diurnal or seasonal, or due to acute responses to exposure to allergens or infectious agents. However, most of the short-term biological variability remains unexplained. Whatever its cause, the changes are reversible and randomly distributed around the mean level. As this section of the thesis is concerned with chronic irreversible airway obstruction, all of the above sources of variation, as they occur in the van Doorn data set, should be included in the variance of random error (σ_δ^2).[99] This is in contrast to much of the published literature on lung function variability, in which random error of measurement is often assessed on only one piece of equipment by one technician and sometimes on only one occasion.[30]

5.2 A SIMPLE ESTIMATION OF G IN THE VAN DOORN DATA SET

The simplest method of estimating the reliability coefficient is from a re-examination of lung function in a series of cases over a few weeks. This time interval is long enough to include all the necessary components of variance while short enough to avoid true changes in the lung function level. Unfortunately, repeat measurements within a few weeks were not available in the van Doorn study. However, 331 individuals had repeat measurements about 1 year apart (0.9-1.1 years). This time period may already include some variability attributable to a true change in level. However, the true variability is likely to be only a small component of the within-person variance over 1 year.[98] Estimates of the reliability coefficient ($\hat{G}$) on these 331 men varied between the different lung function tests, the highest being 0.929 for FVC and the lowest 0.786 for FEV% (Table 5-1).

Table 5-1 Reliability coefficient ($\hat{G}$) and descriptive statistics for within-person differences in lung function tests one year apart (n = 331)

Lung function	$\hat{G}$	Within-person differences (2nd-1st measurement) Mean	Variance	Skewness	Kurtosis†
FVC	0.929	-0.023	0.1012	-0.13	0.80**
FEV	0.899	-0.035	0.1348	-0.22	1.55***
FEF	0.836	-0.163	0.8733	-0.24	1.96***
FEV%	0.786	-0.29	26.54	0.18	2.95***
PEFR	0.818	-14.6	2161	-0.46***	0.94***

† Absence of kurtosis = 0

In all tables, * $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

5.3 CHECKING ASSUMPTIONS FOR CORRECTING REGRESSION COEFFICIENTS USING RELIABILITY COEFFICIENTS OR INSTRUMENTAL VARIABLES

5.3.1 Assumptions for using G to correct regression coefficients

Correction of regression coefficients using the reliability coefficient is based on the assumption that random measurement error is normally distributed and independent of the true level of lung function.[39,74,76] Though reasonably symmetrical, the distributions of within-person variance were kurtosed (Table 5-1). However, correction of regression coefficients is consistent if the error is not normally distributed though the variance of the regression coefficient may be slightly biased.[76] It is possible to assess whether the variance of random measurement error depends on level. To do so, level was defined as the mean of the repeat lung function test results, both of which are considered measurements of the same level. The within-person variances were compared between tertiles of level and the significance of differences between the variances in the upper and lower tertiles determined by the F-test.[100] The within-person variance was not significantly associated with level for FVC, FEV, and FEV%. For FEF there was more than a doubling of the variance from the lower to the upper tertile of level, whereas for PEFR the variance was about 50% greater in the lower compared to the upper tertile (Table 5-2).

Table 5-2 Within-person variance by tertiles of lung function level (n = 331)

Lung function	LFT level tertile			F-ratio	p
	Lower	Middle	Upper		
FVC	0.1075	0.1005	0.0966	1.11	NS
FEV	0.1150	0.1517	0.1372	1.19	NS
FEF	0.4878	0.8500	1.2550	2.57	<0.001
FEV%	23.16	29.21	26.08	1.13	NS
PEFR	2681	2126	1757	1.53	<0.025

NS Not significant ($p > 0.05$)

To avoid the association between within-person variance and level of FEF and PEFR, various simple transformations of the original lung function measurements were tried. Within-person variance and level were not significantly associated for $\sqrt{FEF}$ and for $PEFR^2$ (Table 5-3).

Table 5-3 Within-person variance for transformed lung function measurements by tertiles of level

Lung function	LFT level tertile			F-ratio	p
	Lower	Middle	Upper		
$\sqrt{FEF}$	0.03652	0.04190	0.04584	1.26	NS
$PEFR^2/10^6$	2.504	2.727	2.947	1.18	NS

The $\sqrt{FEF}$ transformation was therefore used in all further analyses in which corrections for random error of measurement were made. The reliability coefficient for $\sqrt{FEF}$ was 0.847. $PEFR^2$ was also used in further analyses. However, because the association between within-person variance and level was modest and all results using PEFR and $PEFR^2$ were similar, only those for PEFR will be presented in the thesis.

The within-person variance for each lung function was also examined in groups defined by other possibly important predictors. There was no important association with age over the relatively narrow age-range of the van Doorn study sample (Section 4.2). It therefore seems reasonable to estimate the random error variance irrespective of age and apply it to correct regression coefficients for the data on 18-22 year old men analysed in chapter 6. There was also no important association between the within-person variance of lung function and the calendar years in which the measurements were done. The variance of measurement error can therefore be assumed to be the same for initial measurements in the mid-'60s and final measurements on the same men in the mid-'70s when the consequences of measurement error in the dependent variable, lung function change are explored in chapter 7.

5.3.2 Assumptions for obtaining unbiased regression coefficients using instrumental variables

A variable only qualifies as an instrumental variable for a predictor if its error of measurement is independent of the error of the predictor (Section 3.3.2).[45] Whether lung function tests can be used as instrumental variables for each other was therefore tested by examining the correlation matrix of within-person differences. With the exception of that between FEV% and PEFR, all within-person differences were significantly correlated suggesting that lung function tests cannot be used as instrumental variables for each other (Table 5-4).

Table 5-4 Correlation matrix of within-person differences in lung function (n= 331)

	FVC	FEV	√FEF	FEV%
FEV	0.60	1		
√FEF	0.19	0.61	1	
FEV%	-0.12	0.71	0.58	1
PEER	0.13	0.16	0.12	0.09

$p < 0.05$ if $r \geq 0.11$

$p < 0.01$ if $r \geq 0.14$

Not surprisingly, the highest correlations were for those lung function measurements (FVC, FEV, √FEF and FEV%) derived from the same procedure, spirometry. However, lung function measured on different types of instruments, namely spirometers or peak exploratory flow meters, also showed small but statistically significant correlations. This presumably is a reflection of short-term biological rather than technical variability.

5.4 OTHER METHODS OF ESTIMATING RELIABILITY COEFFICIENTS

5.4.1 Concepts

Reliability coefficients for lung function may be underestimated using the simple method given in section 5.2 because a 1 year time period between examinations is likely to include some true variation in lung function level. An alternative method of calculating reliability including only that component of variance due to random

error of measurement has been proposed[96] and found applications in the biological literature.[74] The method consists of plotting correlation coefficients for lung function measurements against the time interval between the two measurements. An appropriate model is then fitted to this correlelogram (also called a correlogram). The model needs to take account of the form of the correlelogram which usually shows an asymptotic decrease with increasing time between measurements. An extrapolation of the model to time zero gives an estimate of the reliability coefficient. Alternative methods of applying this principle follow from the discussion of 5.1.1[96] First, the variances of differences between measurements could be plotted against each time interval and a model applied to this variogram to obtain an estimate of the within-person variance at time zero. The variance of random error of measurement, equivalent to half the within-person variance, can then be used to calculate the reliability coefficient. Second, the same procedure could be used, but plotting half the variance of differences between measurements as the dependent variable to directly obtain an estimate of the variance of random measurement error at time zero. This procedure is called a semi-variogram. Correlelogram, variogram and semi-variogram models should all result in similar estimates of the reliability coefficient. In the rest of this chapter only correlelograms will be used. In the van Doorn data set, correlelograms showed fewer outliers and models fitted better than for the other procedures.

5.4.2 Methods

Correlelograms were based on all possible pairs of lung function measurements in the van Doorn data set irrespective of the age of the individuals, how long they remained in the mining industry or when the measurements were done. For example, if a man had three measurements, this resulted in three possible pairs; the first and second measurement, the second and third, and the first and third. Men who had only one measurement were obviously excluded. Pairs were classified according to the number of years between the two lung function examinations. As for the simple estimation of G (Section 5.2), all pairs with measurements 0.9 to 1.1 year apart were grouped as measurements one year apart. Similarly, a 0.1 year interval was allowed around each of the 2, 3, 4 or 5 year measurements. From 6 years onwards all individuals were included within 6 months on either side of the exact year. The longest time interval was 13 years. The number of pairs remaining for analysis was 2 957, derived from 1 077 individuals.

Correlelograms were obtained by plotting the time-specific correlation coefficients against the mean time between examination for each group. Several regression models were then fitted to the data with the correlation coefficient as the dependent variable (y) and the time between examinations as the predictor (x). All regressions were weighted for the square root of the number of pairs of measurements which contributed to each correlation coefficient. The models used are listed below:

1) Linear regression.

2) Linear regression with more weight attached to the correlations at shorter time intervals. This was done because correlelograms were expected to be non-linear. In attempting to extrapolate a linear regression to time zero, it seemed reasonable to give greater weighting to those points closest to zero. The weighting factor used was 14 minus the time between examinations (in years), giving a weight of 13 for the measurements 1 year apart, 12 for that 2 years apart, ... and that 13 years apart.

3) Linear regression using $\log_e$ of the correlation coefficient as the dependent variable. This was intended to introduce curvilinearity. It has some theoretical basis as it is the transformation which fits a first-order autoregressive (Markov) model. The model assumes that the value of the correlation depends only on the correlation for the immediately previous time interval.[74,101] This may be an over-simplification for complex biological data.

4) Asymptotic regression. This non-linear regression takes the form $y = a(1-bc^x)$.[102]

5) Polynomial regression of the form $y = a + bx + cx^2$. Unlike methods 3 and 4 this non-linear regression was not based on any prior expectation of the form of the correlelogram.

5.4.3 Results

The correlelogram data are plotted for each lung function test in Figure 5-1. As expected, the relationship appears curvilinear.

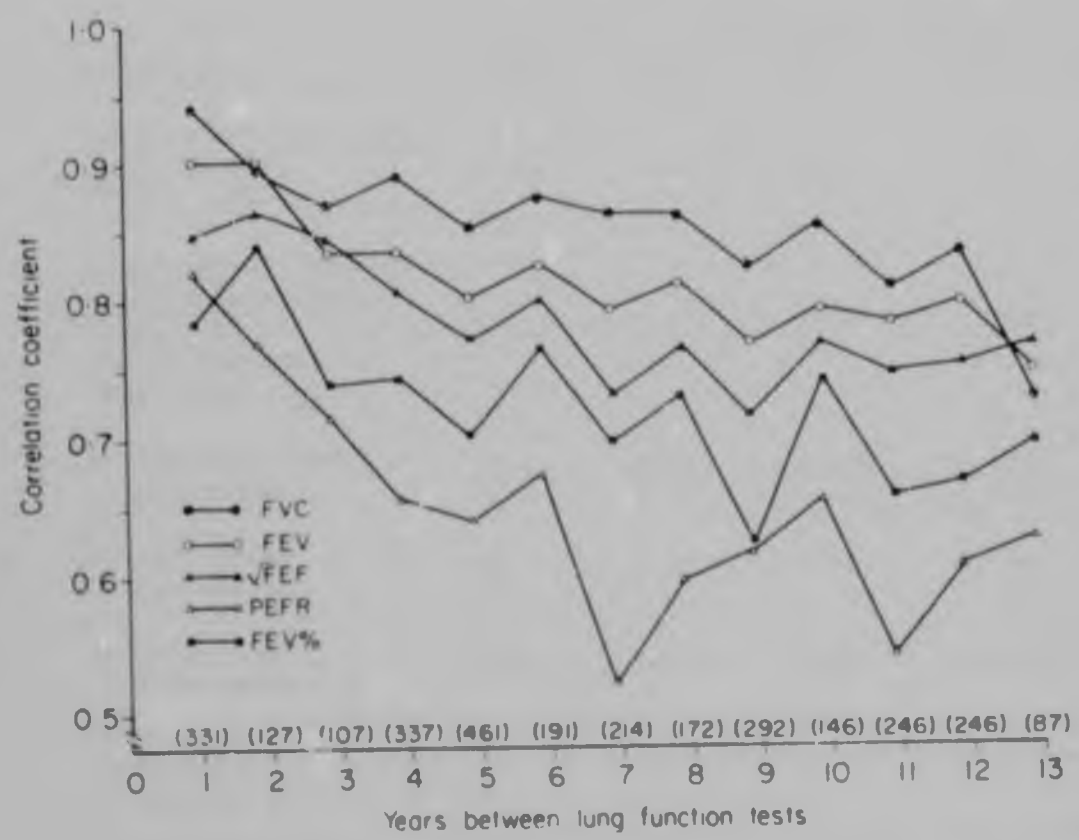

Figure 5-1 Correlation coefficients of lung function measurements by time between examinations
(Bracketed numbers above the X-axis are the number of pairs of measurements at each time interval)

The reliability coefficients estimated using the various models described above are given in table 5-5. Linear (Method 1) and $\log_e$ (Method 3) regressions gave estimates of G within 0.002 of each other. Only that for Method 1 is shown in the table. Except in the case of FEV%, models 1 and 3 gave estimates of G smaller than that obtained from measurements one year apart. The difference was greatest in the

Table 5-5 Reliability coefficients (G) for lung function measurements in the van Doorn data set

Procedure	Estimate of G for FVC	FEV	√FEF	FEV%	PEFR
Measurements one year apart (n=331) (from table 5-1)	0.929	0.899	0.847	0.786	0.818
Intercept of various correlation models: 1) and 3)† Linear regression (and using $\log_e$ y)	0.923	0.881	0.844	0.793	0.758
2) Weighted linear regression	0.921	0.894	0.860	0.804	0.792
4) Asymptotic regression	NA	0.947	0.901	0.834	0 937
5) Polynomial regression	0.915	0.922	0.897	0.829	0.867

† Numbering in accord with the description of the procedures in the methods sec[illegible] (5.4.2)

NA Not applic[illegible] Asymptotic regression analysis did not converge because the relationship is explained well by a straight line

case of PEFR. Weighted linear regression resulted in an estimate of G slightly lower than measure-ments one year apar[illegible] for FVC, FFV, and PEFR and slightly higher for √FEF and FEV%. With the exception of FVC, the non-linear regressions (asymptotic and polynomial) gave higher estimates of G than that obtained from any of the other methods. Similar results were obtained by each of the two non-linear methods (4 and 5) for √FEF and FEV%, whereas coefficients for the asymptotic regression were higher for FEV and especially for PEFR. For FVC, the estimates of G were very similar by any method, whereas those for the other lung functions were more variable. The

variability was greatest for PEFR, where the range of estimates was from 0.758 to 0.937. The observed correlations for PEFR and some fitted models are shown in Figure 5-2 to demonstrate the difficulties of extrapolating models to estimate G at the intercept.

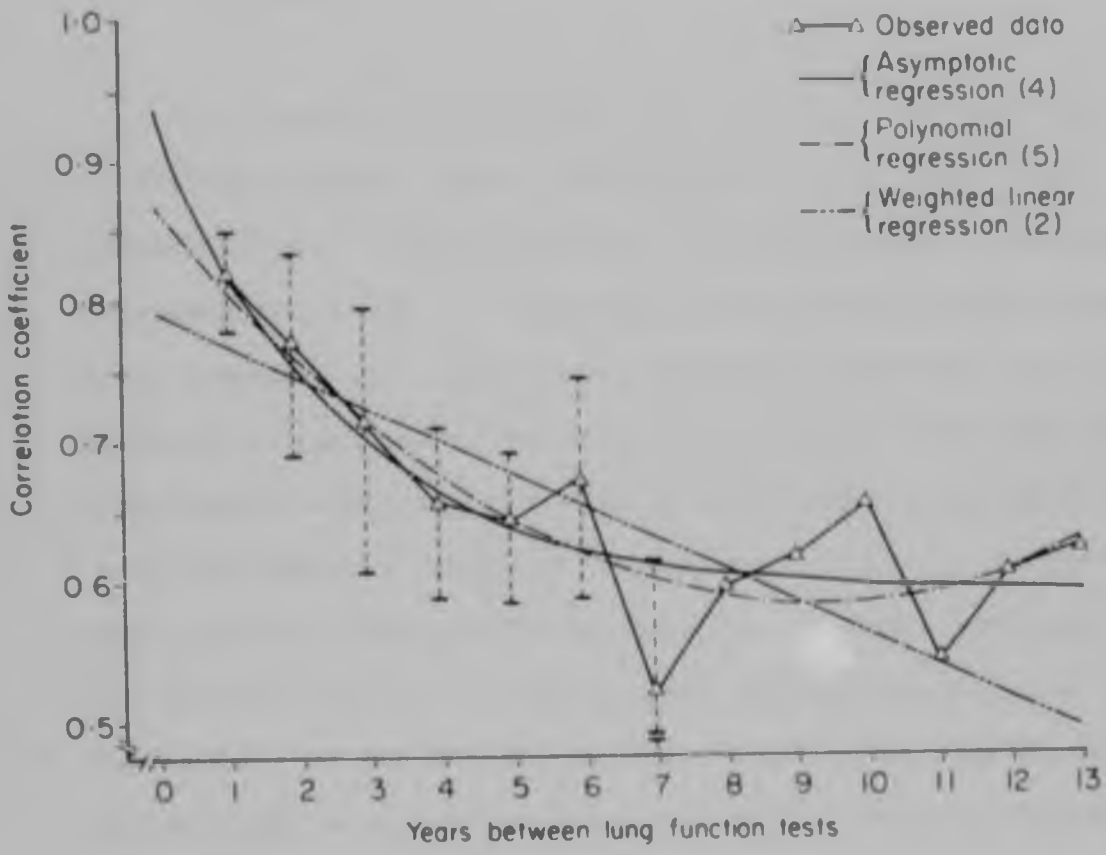

Figure 5-2 Fitting models to PEFR correlelogram data

The observed values suggest a curvilinear relationship for the correlelogram. The weighted linear regression model does not fit the data well for the correlation of measurements 1 year apart. Asymptotic and polynomial regressions fit the data well, both explaining

81% of the variance of the correlation coefficients. The fit at short time intervals is good for both methods, yet the estimates of G at the intercept vary considerably.

5.5 CONCLUSION

The reliability coefficient is equivalent to the correlation coefficient between repeat measurements of the same level. When examining chronic airway obstruction in relationship to environmental exposure, random error of lung function measurement should include both technical and short-term biological components of variance. Reliability coefficients can easily be estimated from lung function measurements within several weeks of each other. In the absence of such measurements, reliability coefficients for the van Doorn data were initially estimated from repeat measurements one year apart. Reliability was also estimated from various correlelogram models. The estimate of the reliability coefficient to be used in Chapter 6 is that based on measurements one year apart. This decision was made because:

a) Within-person variance measured over a few weeks or one year are very similar.[98]

b) The estimate is based on a known number of observations (n=331). This number can be used in the estimation of standard errors of regression coefficients corrected using the estimated G.

c) The estimate usually falls within the range of that obtained from correlelogram models.

G may be slightly underestimated using the one year repeat measurement data. Therefore some analyses will be repeated with alternative estimates of G when correcting regression coefficients in chapter 6.

CHAPTER 6 THE ASSOCIATION OF LUNG FUNCTION CHANGE WITH INITIAL LUNG FUNCTION AND ENVIRONMENTAL EXPOSURE AFTER CORRECTING FOR RANDOM ERROR OF MEASUREMENT

Chapters 3 to 7 of this thesis are concerned with methods for correctly identifying associations between exposure to environmental agents and lung function change. Conceptually, initial lung function should be considered a potential confounder of the relationship between environmental exposure and lung function change if the exposure (or at least a major part of it) occurred after the initial lung function measurement (Section 2.1.1). This is the case in the van Doorn data set for 18 to 22 year-old men who had initial lung function measurements at entry into the mining industry (Section 4.2.1). Lung function change was measured over a subsequent period of about 11 years with the intention of relating it to smoking histories and dust exposure in gold mines over that time period. Initial lung function was therefore used as a predictor in the regression analyses in the van Doorn data set. In studies of older men the initial FEV is sometimes adjusted for height. The purpose is to obtain a measure of the decline of FEV from its peak measurement (for which height acts as a proxy measure) at about 25 years of age to the age at which the FEV was initially measured.[103] The van Doorn study was done on men aged 18-22 years whose initial measurement must be near their peak lung function. The hypothesis under investigation is whether lung function change relates to peak level. Height was therefore not adjusted for in the analyses presented here. Analyses with FEV adjusted for height resulted in similar conclusions.

The relationship between lung function change and initial level is incorrectly estimated when lung function is measured with random error. Similarly, the confounding effect of initial lung function on the relationship between change and environmental exposure is incorrectly estimated (Section 3.1). Unbiased estimates are obtainable, if the reliability coefficient of the measurements can be estimated (Section 3.2).

In the 433 men between 18 and 22 years of age in the van Doorn data set it is of interest to determine:

1) Is lung function change associated with initial lung function after correcting for measurement error? This is of biological interest as longitudinal studies in young adults near the time when they reach their maximal lung function are uncommon. Furthermore, the result will indicate the need to consider initial lung function as a potential confounder in future studies of the effect of environmental exposure on lung function change in young men.

2) Is lung function change associated with smoking histories and mining experience?

Before answering these questions some descriptive statistics are given and the association examined between initial lung function, environmental variables and other potential confounders.

6.1 CHARACTERISTICS OF THE DATA SET

6.1.1 Descriptive statistics

Mean values and standard deviations for age, height, and weight at initial examination are shown in table 6-1, together with the mean change for age and weight between initial and final examinations.

Table 6-1 Descriptive statistics for personal characteristics (n=433)

Characteristic	At initial examination		Change (Final-initial measurement)	
	Mean	SD	Mean	SD
Age (years)	20.2	1.2	11.28	1.21
Height (cm) (Mean of initial and final measurements)	176.8	5.9	-	-
Weight (kg)	70.7	8.1	8.7	7.7

SD Standard deviation

Means and standard deviations for initial and final lung function measurements and lung function change are shown in table 6-2 and the distributions for initial lung function and change in figures 6-1 and 6-2 respectively. Although results for FEF are included in these tables, many subsequent analyses will use only $\sqrt{FEF}$ to ensure independence of level and the variance of random (Section 5.3.1) measurement error.

The means of the initial spirometric measurements were toward the upper end of the range expected for 20 year-old men from prediction equations.[95,104,105] The mean peak expiratory flow rate

Table 6-2 Descriptive statistics for lung function measurements

Lung function	Initial		Final		LFT change (final-initial)	
	Mean	SD	Mean	SD	Mean	SD
FVC	5.57	0.61	5.46	0.68	-0.11	0.42
FEV	4.68	0.52	4.34	0.60	-0.35	0.42
FEF	5.49	1.23	4.38	1.34	-1.12	1.00
√FEF	2.33	0.26	2.07	0.32	-0.26	0.23
FEV%	84.3	6.4	79.6	6.8	-4.7	5.5
PEFR	575	67	574	73	-2	65

SD Standard deviation

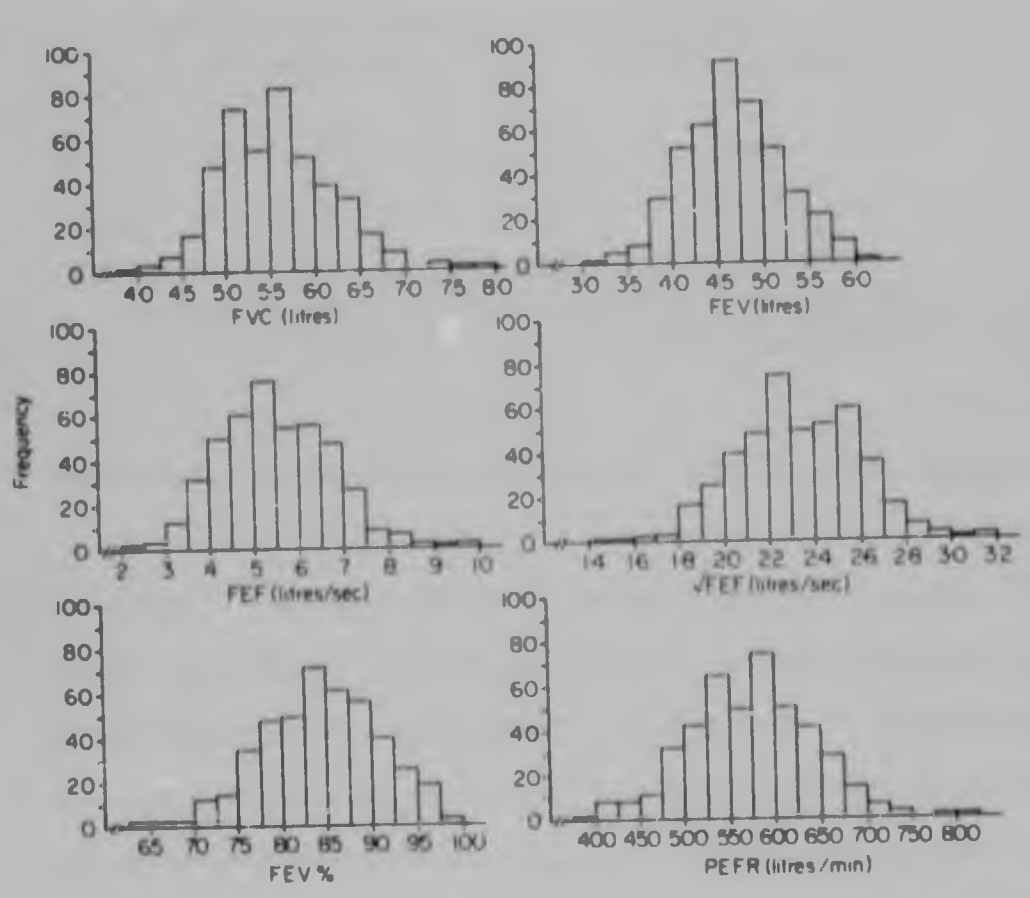

Figure 6-1 Frequency distributions of initial lung function

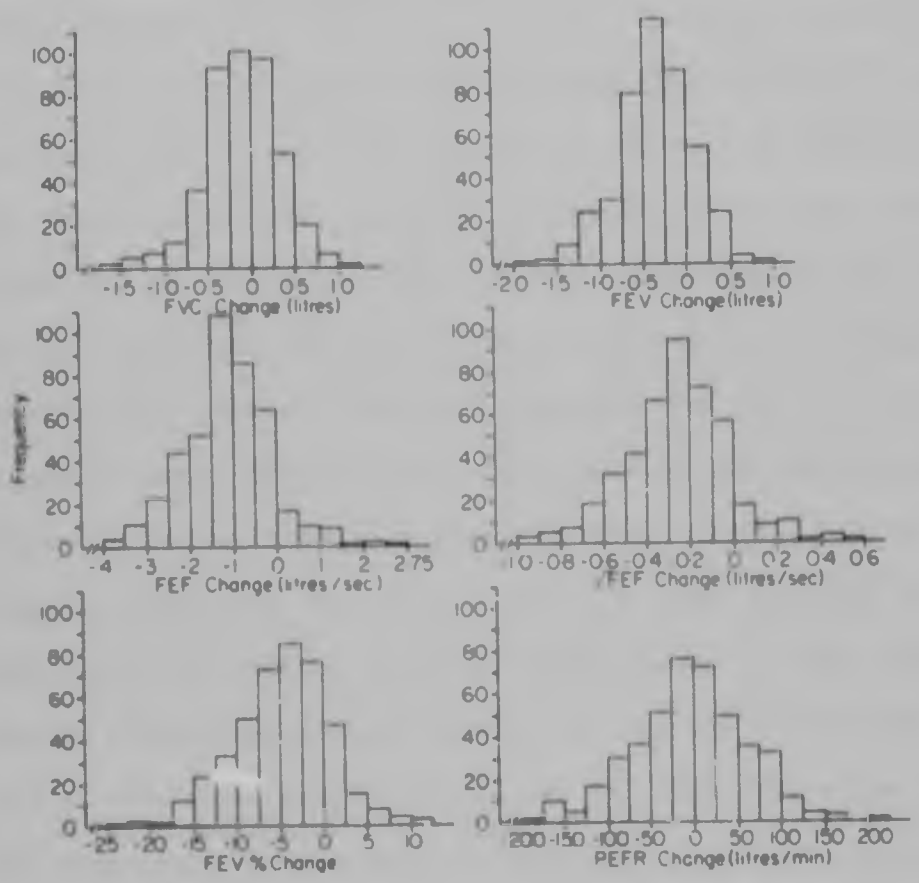

Figure 6-2 Frequency distributions of lung function change

was similar to that expected on the Wright flowmeter.[97,106,107] There was a slight increase in standard deviation from the initial to final lung function measurements, as expected on theoretical grounds and found in previous studies.[49] Lung function change (LFT change) is calculated as the final minus the initial measurement throughout the thesis. Change with a negative sign, as shown in table 6-2, therefore indicates a loss of function over the 11 years between the

initial and final lung function test. The mean change of all lung function measurements was within the range expected over 11 years from other studies.[46,54,56,57,60,61,63,67] However, except for FEF, the loss of lung function over time was significantly less in the 18 to 22 year olds than in men over 23 years of age in the van Doorn data set. The mean change over 11 years in the older men was -0.3? for FVC, -0.43 for FEV, -1.03 for FEF, -3.5 for FEV% and -32 for PEFR. The smaller mean loss in the 18 to 22 year-old men probably occurred because their annual change was positive for the first few years of follow-up until maximal adult lung function was reached, after which change became progressively mor negative. I will therefore not express change as an annual rate, as used in many publications, because it may imply a constant annual change to some readers. The standard deviation of lung function change was considerably less than that of the cross-sectional data, as expected from a consideration of the components of variance of cross-sectional and longitudinal data (Section 2.1). The distributions of initial lung function data did not show any major deviations from normality (Figure 6-1 and Table 6-3). Lung function change data were also distributed reasonably symmetrically around the mean, though somewhat kurtosed (Figure 6-2 and Table 6-3).

Descriptive data for smoking is shown in table 6-4, and for high-dust exposure in gold mines in table 6-5. The frequency distribution of the amount of time spent in high-dust occupations in gold-mines was very skewed, with only 9% of men having more than 8 years of high-dust exposure (Table 6-5).

Table 6-3 Skewness and kurtosis of lung function measurements

Lung function	Initial LFT		LFT Change	
	Skewness	Kurtosis†	Skewness	Kurtosis†
FVC	0.43	0.38	-0.43	1.00
FEV	0.08	-0.12	-0.27	0.47
FEF	0.33	0.20	0.21	0.94
√FEF	-0.01	0.02	0.00	0.78
FEV%	-0.28	-0.13	-0.15	0.46
PEFR	0.02	0.16	-0.14	0.17

†No kurtosis = 0

Table 6-4 Frequency distribution for smoking

Ave. no. cigarettes/day (Mean = 10.7)	No	%	Ave. g tobacco/day (Mean = 0.68)	No	%
None	87	20	None	340	79
1.0- 9.9	95	22	1.0- 2.9	47	11
10.0-14.9	102	24	3.0- 4.9	20	5
15.0-19.9	93	21	5.0- 6.9	10	2
20 or more	56	13	7.0 or more	16	4
Total	433	100	Total	433	100

Table 6 5 Frequency distribution of the number of years worked in high-dust jobs on gold mines

No of years (Mean = 2.16)	No	%
0	259	60
1- 2	54	12
3- 4	31	7
5- 6	23	5
7- 8	24	6
9-10	23	5
≥ 11	19	4
Total	433	99

Other jobs which men held were low-dust underground jobs on goldmines (mean 4.3 years), surface jobs on gold-mines (mean 1.1 years), jobs on coal, platinum or other mines (mean 1.1 years), or non-mining occupations (mean 2.9 years). Only 106 men had spent the full time period between lung function examinations either in underground jobs on gold mines, for which job-specific dust measurements were available, sales and clerical jobs for which it seemed reasonable to assume low dust exposure. Cumulative dust measurements, calculated as described in section 4.1.4, are shown for these 106 men in table 6-6. The correlation between the four indices of exposure on these 106 men ranged from 0.6 to 0.9.

6.1.2 Correlations of environmental exposure variables and initial lung function

Regression coefficients of lung function change on environmental exposure variables and on initial lung function obtained from separat linear regressions will not be biased if the sample correlation coefficient between anvironmental exposure and initial lung function variable is zero (Section 3.2.2). Product-moment correlation coefficients of the environmental exposure variables with each other and initial lung function tests were small (Table 6-7). Spearman correlation coefficients, calculated because of the very skew distribution of the exposure variables, gave similar results. The largest absolute value for a correlation was that between initial FEV and the number of years subsequently spent in high dust gold mining occupations ($r = -0.13$). By regression analysis, with height and weight as covariates (age and cigarette consumption were not significant predictors), initial FEV was 0.016 litre lower for each year subsequently spent in high-dust jobs.

Table 6-6 Frequency distribution of cumulative dust exposure (Particles/ml x hours exposed/day x years) in underground gold-miners (n = 106)

		Cumulative exposure					
No years in high dust jobs	No	Particles/ml (before acid)	No	Particles 0.5-5µm/ml (after acid)	No	Particles 0.25-5µm /ml (after acid)	No
0	49	< 2 000	12	< 100	4	< 600	1
1- 2	6	2 000-	38	100-	38	600-	48
3- 4	12	2 500-	10	150-	21	800-	12
5- 6	6	3 000-	8	200-	14	1 000-	18
7- 8	6	3 500-	12	250-	24	1 200-	26
9-10	13	4 000-	18	≥ 300	5	≥ 1 200	1
≥ 11	14	≥ 4 500	8				
Total	106		106		106		106

Table 6-7 Correlations of exposure variabl s and initial lung function test results

	Initial LFT					Exposure	
Exposure	FVC	FEV	√FEF1	FEV%	PEFR	Ave.g other tobacco/day	Years in high dust
Average cigarettes /day	-0.05	-0.08	-0.09	-0.05	-0.09	-0.10*	0.01
Average g other tobacco/day	0.03	-0.04	-0.07	-0.09	-0.05	1	0.06
Years in high dust	-0.10*	-0.13**	-0.08	-0.06	-0.02	0.06	1

* p <0.05

** p <0.01

6.1.3 Searching for other confounders

Several other variables might be confounders in longitudinal lung function data. The most important of these are age at initial examination, weight change and the time between examinations. The correlation between these variables and lung function change is shown in table 6-8.

Table 6-8 Correlation of potential confounders with lung function change

	LFT Change				
Potential confounder	FVC	FEV	√FEF	FEV%	PEFR
Age	-0.21***	-0.16***	-0.14**	0.02	-0.22***
Weight change	-0.15**	-0.14**	-0.04	-0.02	0.01
Time between examinations	-0.04	0.02	-0.02	0.08	0.01

**p <0 `1

***p <0.001

Age was significantly negatively correlated with all changes in lung function except for FEV%. This probably occurred because men aged 18-22 years at initial examination had not yet reached the age at which lung function started declining. Within this age range the older men would have more of the time between examinations falling beyond the age at which decline started and therefore have more negative lung function change. Weight change (presumably mostly obesity) was negatively correlated with change in FVC and FEV. Surprisingly the duration of time between the initial and final lung function tests was not predictive of change. This probably reflects the narrow range of times between examinations; 79% were between 10 and 12 years.

Table 6-9 shows the correlations between the variables being investigated for possible confounding and the variables whose prediction of lung function change are of prime interest. All these correlations were small (Table 6-9). Spearman correlation coefficients were of similar magnitude. The absence of any strong relationship of initial lung function with age is probably due to the narrow age range studied and the fact that 18-22 year old men have not yet commenced their adult lung function decline. Note that age was positively associated with FVC and PEFR, though only significantly so for the latter.

6.1.4 Conclusion

Correlations between initial lung function and environmental exposure variables were small and unlikely to result in biased

Table 6-9 Correlation of potential confounders with initial lung function, exposure variables and each other

Potential confounder	Initial LFT				
	FVC	FEV	√FEF	FEV%	PEFR
Age	0.06	-0.02	-0.06	-0.11*	0.10*
Weight change	0.06	0.04	0.01	-0.02	0.01
Time between examinations	0.00	-0.05	-0.06	-0.08	0.09

Potential confounder	Environmental exposure			Potential confounder	
	Average cigarettes /day	Average g other tobacco/day	Years high dust	Weight change	Time between examinations
Age	0.06	0.14**	0.00	-0.08	-0.06
Weight change	-0.04	-0.09	-0.03	1	0.07
Time between examinations	0.01	-0.05	0.04	0.07	1

* $p < 0.05$

** $p < 0.01$

estimates of regression coefficients obtained from separate regression analyses. Separate regression equations are therefore calculated for lung function change on initial lung function (with correction for random measurement error) in section 6.2 and on environmental exposure in section 6.3. Multiple regression procedures corrected for random error of measurement are dealt with in section 6.4.

Age was a determinant of lung function change and weight change a determinant of change in FVC and FEV. However these variables were not considered confounders because they were only weakly associated

with initial lung function and exposure variables. They were therefore not routinely adjusted for in subsequent analyses, though the effect of their inclusion on regression coefficients and significance tests was explored (Section 6.3.2).

6.2 REGRESSION OF LUNG FUNCTION CHANGE ON INITIAL LUNG FUNCTION (ASSUMING NO ASSOCIATION BETWEEN INITIAL LUNG FUNCTION AND ENVIRONMENTAL EXPOSURE)

6.2.1 Methods

The regression coefficient of lung function change on initial lung function was calculated for each lung function using several methods as follows:

1) The ordinary least squares (OLS) regression coefficient of lung function change on initial level was obtained without any correction for measurement error. (Note that estimates of regression coefficients are referred to as β rather than $\hat{\beta}$ in this chapter.)

2) The OLS coefficient obtained in 1) was corrected according to equation 3:8 using the estimate of the generalizability coefficient (referred to as G rather than $\hat{G}$ in this chapter) shown in table 5-1. Standard errors of the regression coefficient were calculated first using equation 3:9 and second by dividing the Method 1 OLS standard error by G.

3) FVC and FEV change were regressed on initial levels of FVC and FEV predicted from their regression on height, i.e. using height as an instrumental variable. Predicted values were not calculable for the other lung function tests as there were no appropriate instrumental variables.

4) Lung function change was regressed on the mean of initial and final lung function measurements.

5) Lung function change was regressed on a lung function measurement independent of those from which change was calculated. This method

required 3 separate lung function test results. This analysis therefore could not be done on the full sample of 433 men but only on a subset who had an additional ('middle') lung function test between their initial and final examinations. The criteria for sample selection were generated to maximise the sample size, while restricting the age range for the middle measurement. Men were only included if aged between 23 and 26.99 years at their middle lung function test, and if their final examination was at least 3.5 years later. Change was calculated as the final minus the middle lung function adjusted to the mean time between the two examinations (6.5 years). Change was regressed on middle lung function (ignoring the effect of error of measurement) and on the initial lung function test measured independently an average of 4.4 years earlier.

6.2.2 Results and discussion

Results of methods 1 to 4 are shown in table 6-10. Results for method 5 are given in table 6-11.

1) All of the OLS regression coefficients (Method 1) were significantly negative (Table 6-10, column 1).

2) The second column of the table shows the coefficients corrected for G. The values of G used were 0.929 for FVC, 0.899 for FEV, 0.847 for √FEF, 0.786 for FEV%, and 0.818 for PEFR (Table 5-1). Corrected regression coefficients did not differ significantly from zero, except for PEFR for which the coefficient remained significantly negative, and FEV%, which was only marginally significant. Standard errors of

Table 6-10 The regression of change on initial measurement using various methods (n = 433)

LFT		OLS Regression (Method 1)	Regression corrected using G (Method 2)	Regression on predicted initial LFT (Method 3)	Regression on (M1+M2)/2 (Method 4)
FVC	β	-0.109 (-0.16)†	-0.041	+0.076	+0.129
	SE	0.033	0.036 (0.036)††	0.052	0.033
	p	0.001	0.3	0.1	<0.001
FEV	β	-0.154 (-0.19)	-0.059	-0.032	+0.159
	SE	0.037	0.043 (0.041)	0.072	0.037
	p	<0.001	0.2	0.7	<0.001
√FEF	β	-0.128 (-0.15)	0.030	NA	+0.242
	SE	0.042	0.053 (0.050)		0.039
	p	0.002	0.6		<0.001
FEV%	β	-0.303 (-0.35)	-0.113	NA	+0.076
	SE	0.039	0.056 (0.050)		0.044
	p	<0.001	0.04		0.08
PEFR	β	-0.382 (-0.39)	-0.244	NA	+0.103
	SE	0.043	0.055 (0.053)		0.050
	p	<0.001	<0.001		0.04

β Regression coefficient estimate

SE Standard error

NA Not applicable

† Bracketed values in this column are correlations between lung function change and initial measurement

†† The first standard error is calculated from equation 3:9.
The bracketed standard error is calculated from SE (Method 1)/G.

the regression coefficients were increased by the correction procedure, but did not differ markedly whether calculated from equation 3:9 or simply by dividing the standard error of the OLS regression by G. The effect of the correction of the regression coefficient was very sensitive to the estimate of G used. For example, using the lowest estimate of G (0.881) from table 5-3 for FEV, the regression coefficient was -0.040 (not significant). Using the highest estimate of G (0.947), the regression coefficient was -0.107 ($p = 0.007$). Regression coefficients were not significantly different from zero for FVC and √FEF, irrespective of the estimate of G used from table 5-5. On the other hand, regression coefficients for FEV% and PEFR were significantly negative for all values of G.

3) The regression coefficients of FVC change on predicted FVC and FEV change on predicted FEV (the instrumental variable approach) were not significantly different from zero. The predicted variables had less variation than equivalent observations because only 40% of the variance of FVC and 27% of the variance of FEV were explained by height. The standard errors of the regression coefficients of change on predicted initial lung function were therefore wider than those of the corrected coefficients of change on initial values.

4) When change was regressed on the mean of the initial and final lung function results, the regression coefficients were positive in all cases and only failed to reach significance for FEV% ($p = 0.08$).

5) The results of analyses on the 221 men with lung function change (final-middle LFT) regressed on initial lung function are shown in the third column of table 6-11. For comparison the first column

shows the regression coefficient of change on middle measurement and the second column shows that coefficient corrected using G as was done on the second column of table 6-10.

Table 6-11 Regression of change (last-middle LFT) on middle and initial LFT (Method 5, n = 221)

LFT		On middle LFT (Method 1)	On middle LFT corrected by G (Method 2)	On independent initial LFT (Method 5)	Method 5 further corrected by G
FVC	β	-0.191	-0.129	-0.115	-0.124
	SE	0.034	0.038†	0.038	0.041†
	p	<0.001	0.001	0.002	0.002
FEV	β	-0.232	-0.146	-0.073	-0.081
	SE	0.038	0.044	0.048	0.053
	p	<0.001	0.001	0.1	0.1
√FEF	β	-0.215	-0.073	-0.026	-0.031
	SE	0.041	0.053	0.052	0.062
	p	<0.001	0.2	0.6	0.6
FEV¼	β	-0.300	-0.109	-0.072	-0.092
	SE	0.043	0.062	0.048	0.061
	p	<0.001	0.08	0.1	0.1
PEFR	β	-0.423	-0.295	-0.062	-0.076
	SE	0.054	0.069	0.068	0.083
	p	<0.001	<0.001	0.4	0.4

† SE's are obtained using equation 3:9

Note that the results cannot strictly be compared with those in table 6-10 because of the different starting age (23-27 years) from which change was measured and the shorter average time interval (6.5 years)

for the measurement of change. The regression coefficients of change on middle lung function were all highly significantly negative (Table 6-11, column 1). That for FVC, FEV, and PEFR remained significantly negative after correction using G (Table 6-11, column 2). When the same change measurements were regressed on independently measured initial lung function values, only that for FVC remained significantly negative. The results obtained in columns 2 and 3 were of similar significance except for FEV and PEFR. In the case of PEFR the discrepancy was marked. Obviously, the independent initial LFT used as the predictor in the regression analyses in column 3 is also prone to measurement error, for which the regression coefficients can be corrected. This requires no more than dividing the β estimates by G. (By contrast, note that the correction procedure, when the predictor measurement is also the one from which change is measured requires the estimate of $1+\beta$ to be divided by G.) The effect is to marginally increase the absolute magnitude of the coefficients without decreasing their p-values. P-values cannot decrease because the standard errors of the regression coefficients are also divided by G assuming very large sample sizes, or a quantity larger than G if sample sizes are smaller. Note that similar corrections could be applied to the results of method 4 (Table 6-10) using estimates of G calculated from equation 3:15.

6.2.3 Conclusion

a) The favoured method of correcting the regression of change in initial value for random error of measurement is by the use of the

generalizability coefficient (Method 2). The advantages over other methods are:

i) Unlike method 3, method 2 does not require instrumental variables, which are in fact not always available. Even when they are, the standard errors of the regression coefficients are very wide.

ii) Unlike method 4, method 2 is not expected, on theoretical grounds, to give biased results (Sections 3.2 and 3.3). The coefficients obtained in this data set suggest that the bias in method 4 is of sufficient magnitude to cause concern.

iii) Unlike method 5, which requires repeat initial measurement on the full sample, in method 2 a reliability coefficient (G) can be estimated from a subset or alternative sample. Furthermore the coefficients from method 5 are still minimally biased unless further corrected for the measurement error in the initial independent measurement.

iv) Finally, correction by the generalizability coefficient has conceptual appeal and is easily applied to other areas of epidemiological research.

b) Similar results were obtained by all valid methods (i.e. excluding methods 1 and 4 for FEV, FEF and FEV%. For each of these lung function tests there was no important association between change and initial lung function level. A similar result was obtained for FVC by methods 2 and 3, whereas method 5 yielded a significantly negative association. This may be due to the alternative sample

definition or sampling variability, as the regression coefficient for method 5 is still more negative than the uncorrected (and therefore negatively biased) OLS regression on the full sample of 433 men. It therefore seems reasonable to accept FVC change as not being significantly associated with initial FVC in men between 18 and 22 years of age. In the case of PEFR, correction using the estimated G results in a highly significantly negative coefficient (-0.244) whereas that of method 5 does not differ significantly from zero. In view of the uncertainty of the estimate of G for PEFR in this data set (Chapter 5.4.3) no conclusion can be reached about the true regression relationship. However correction using any of the estimates of G (Table 5-5) result in a negative regression. Therefore, with the possible exception of PEFR, it would be reasonable for investigators of the effect of exposure to environmental agents on lung function change in young men to consider that initial lung function is not a predictor of change. Change can then be directly regressed solely on the exposure variable of interest.

6.3 REGRESSION OF LUNG FUNCTION CHANGE ON ENVIRONMENTAL EXPOSURE (ASSUMING THAT INITIAL LUNG FUNCTION IS NOT A CONFOUNDER)

The association between exposure to environmental agents and lung function change in this data set have been explored without including initial lung function as a predictor in the regression equation because exposure and initial lung function have correlation coefficients close to zero (Section 6.1).

6.3.1 Methods

Lung function change was regressed simultaneously on the average annual cigarette and other tobacco (pipe and cigar) consumption and on the duration of work in high-dust gold-mining occupations. Regression analyses of lung function change on quantified dust exposure was also done on the sample of 106 full-time goldminers for whom dust exposures were estimated.

6.3.2 Results

The regression coefficients of lung function change on environmental exposure variables are shown in table 6-12. Repeat analyses including other significant covariates such as age and weight change (Section 6.1.3) had little influence on the coefficients or their significance.

Cigarette smoking was highly significantly associated with FVC, FEV, and FEF loss. Using these regression coefficients, men smoking 20 cigarettes per day lost on average 280 ml more FVC, 260 ml more FEV

and 420 ml/sec more FEF than non-smokers over the 11.3 year follow-up period. These lung function changes were linearly related to smoking; the addition of terms for non-linearity were not significant.

Table 6-12 Regression of lung function change on environmental variables

LFT Change	Regression coefficients for environmental variables		
	Average no. of cigarettes/day	Average g other tobacco/day	No. years high-dust gold-mining jobs
FVC	-0.013***	-0.029**	-0.01 (p = 0.07)
FEV	-0.014***	-0.026**	-0.003
FEF	-0.021***	-0.022	0.016
√FEF	-0.0055***	-0.0069	0.031
FEV%	-0.069* (p = 0.04)	-0.025	0.078
PEFR	-0.79* (p = 0.05)	1.01	-2.21* (p = 0.014)

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

Exact p-values are given in brackets if between 0.1 and 0.01.

Cigarette smoking was only very weakly associated with FEV% or PEFR loss. Pipe and cigar smoking was significantly associated with loss of FVC and FEV. No associations were found between lung function change and mining experience except in the case of PEFR. As estimated from the regression coefficient, miners who worked in high-dust gold-mining jobs for the full 11.3 year follow-up period lost 25 litres/minute more PEFR than men none of whose follow-up time was spent in these jobs. On the sample of 106 men for whom cumulative dust exposure on gold mines was estimable, PEFR change was again the only lung function measure which was significantly predicted by dust

exposure. However the level of significance was not markedly different to that for the crude measure of the number of years spent in high-dust jobs (Table 6-13).

Table 6-13 P-values (and sign, if negative) of partial association of lung function change separately on each type of dust measurement.

LFT Change	Dust measurement (particles/ml x hours exposed/day x years)			
	Years high dust	Particles/ml (before acid)	Particles 0.5-5μm/ml (after acid)	Particles 0.25-5μm/ml (after acid)
FVC	0.2 (-)	0.2 (-)	0.6 (-)	0.1 (-)
FEV	0.5	0.8	0.4	0.6
FEF	0.1	0.3	0.08	0.1
√FEF	0.1	0.4	0.1	0.2
FEV%	0.03	0.2	0.2	0.05
PEFR	0.04(-)	0.01(-)	0.01(-)	0.005(-)

Other items in each regression were cigarette smoking and smoking of other tobacco

6.3.3 Discussion

Cigarette smoking was related to loss of lung function using any of the test measurements. Smoking of pipe or cigar tobacco also resulted in loss of FVC and FEV. The loss of FEV and FEF is in accord with the known harmful effect of smoking. In this data set, the magnitude of effect of smoking on FVC was similar to that for FEV. This finding agrees with that of a number of other longitudinal studies.[38,57,60,64] The physiological explanation is unclear. However this may be an effect of increased air trapping in smokers during the forced test manoeuvre. It would be of interest in future studies to investigate whether the same phenomenon occurs with slow vital capacity measurements. The association between cigarette

smoking and FEV% barely reaches significance. This is not surprising in view of the similar magnitude of effect of cigarette smoking on both FEV and FVC. Because of this, and its relatively low reliability (Table 5-1), FEV% seems a poor measure for investigating possible effects of exposure to environmental agents on lung function.

The only significant association between indices of dust exposure and lung function was that with peak exploratory flow rate. This finding is all the more surprising in view of the fact that cigarette smoking, as a well known cause of airway obstruction, is only marginally significantly associated with PEFR. Conventional understanding of mechanisms of development of airway obstruction, based largely on studies of cigarette smoking, is that disease starts in the small airways. Early disease is therefore unlikely to be detectable by a test which is predominantly dependent on large airway calibre as is the case for peak exploratory flow rate. The finding in this study suggests the possibility that different environmental exposures may affect different parts of the bronchial tree. This suggestion has some support in the literature.[108,109] The differing responses may well depend upon differing characteristics of the agent to which individuals are exposed. If so, the medical importance of early abnormality in parts of the bronchial tree other than small airways is uncertain and requires further study. To date the major studies have largely been concerned with the natural history of airway obstruction caused by cigarette smoking. There was some suggestion in this data set that the duration of work in high-dust jobs on gold

mines may be associated with an increased loss of FVC (Table 6-12, $p = 0.07$). This could be explained by the onset of early restrictive lung disease, or possibly an effect on airways analogous to that for peak exploratory flow. The question may be raised why no association between dust exposure and change in spirometric measurements was found when exposure-response relationships between gold mining experience and spirometric measurements has previously been found for South African gold miners.[110] Moreover, the literature supports the view that silica exposure is likely to cause obstructive airways disease.[111] The lack of association between dust and lung function change in the van Doorn data set may well be a reflection of the relatively small range of dust exposure. Only a small proportion of men had spent the majority of their time in high-dust gold exposure. When not so occupied, they often worked in mining occupations in which they were exposed to dust, although this was known to be at a lower concentration than in the high dust occupations. However, much of the work experience was also outside of the mining industry and at levels of exposure which are unknown. For the 106 men who had always worked in the gold mines (or sales and clerical jobs assumed to be non-dusty), there was a relatively narrow range of calculated dust exposure (Table 6-6). Moreover, the cross-sectional study previously referred to[110] was done on older men in whom the effect of dust exposure had accumulated over several decades as opposed to the 11 year follow-up period in the van Doorn study.

6.3.4 Conclusion

a) Cigarette smoking was linearly associated with an accelerated loss of all the lung functions studied. Smoking of other tobacco (pipe and cigar) was significantly associated with loss of FVC and FEV.

b) Peak expiratory flow rate was significantly associated with indices of dust exposure on gold mines but only marginally significantly associated with cigarette smoking. This suggests the possibility of differential effects of exposure to environmental agents on different airways depending on the characteristics of the agent involved.

6.4 CORRECTION FOR RANDOM ERROR OF MEASUREMENT IN MULTIPLE REGRESSION ANALYSIS (WHEN THE PREDICTORS INITIAL LUNG FUNCTION AND ENVIRONMENTAL EXPOSURE MAY BE CORRELATED)

Methods for taking account of errors in variables in the multiple regression situation were described in Section 3.2.2. Analyses using these techniques were undertaken on the van Doorn data set. Not surprisingly, the results obtained were very similar to those obtained in sections 6.2 and 6.3 because environmental exposure and initial lung function were not strongly associated with each other. The purpose of this section is to illustrate the importance of correcting for measurement error in multiple regression analyses when the predictor variables are correlated. For this illustration only a simple example is necessary. The example I use is FEV change in relationship to cigarette smoking in a subset of the data. The subset was chosen in such a way that initial LFT and subsequent cigarette smoking were correlated.

6.4.1 Methods

Selection of subsample: An association between cigarette smoking and initial lung function in a subset of the van Doorn data could have been achieved by cross-classifying these two variables and choosing a sample from the diagonal of the cross-classification. However, this might result in a correlation between smoking and the error of lung function measurement. Because height and lung function are associated, selecting from the diagonal of a cross-classification of height and cigarette smoking should result in a correlation between

true values of initial lung function and smoking. The cross-classification of height and smoking on the van Doorn data set is shown in table 6-14.

Table 6-14 Cross-classification of height and average cigarette consumption

Average cigarettes/day)	Height (cm)										Tota
	-167	-170	-172	-174	-176	-178	-180	-182	-184	185+	
<0.001	(5	6	1)	7	7	21	13	11	4	12	87
0.001 - 5.24	3	(3	1	5)	9	6	7	2	5	2	43
5.25 - 9.10	3	4	(8	5	5)	8	3	3	2	1	42
9.11 - 11.35	6	2	7	(0	7	7)	8	3	0	4	44
11.36 - 13.30	3	4	3	4	(7	6	9)	4	2	1	43
13.31 - 15.68	3	3	5	2	4	(10	2	6)	5	4	44
15.69 - 17.70	3	1	5	4	9	8	(4	2	4)	2	42
17.71 - 20.01	1	4	2	5	4	12	7	(2	3	5)	45
>20.01	3	5	2	4	4	3	8	2	(5	7)	43
Total	30	32	34	36	56	81	61	35	30	38	433

The sample selected along the diagonal, indicated by brackets, consisted of 125 men and showed a moderate correlation with cigarette smoking and initial lung function (Table 6-15).

Table 6-15: Correlation induced between lung function and average cigarette consumption (n = 125)

	Initial FEV	FEV Change
FEV Change	-0.31	1.00
Average no of cigarettes/day	0.45	-0.18

The data were analysed using several methods.

1) FEV change was regressed simultaneously on average cigarette consumption and the initial FEV measurement without any correction for random error of measurement.

2) Corrected regression coefficients were obtained using the generalizability coefficient. This was done by multiplying the variance of the initial FEV measurement by its reliability coefficient and substituting it in the variance-covariance matrix of final on initial FEV and cigarette smoking. The corrected variance-covariance matrix was then used to calculate regression coefficients. The standard errors of the regression estimates were calculated using the method described by Warren et al[76] (Equations 3:10 and 3:11).

3) FEV change was regressed on cigarette consumption without including initial FEV in the model, on the basis of prior information (or the assumption) that initial FEV is not truly a predictor of FEV change.

For comparison two other methods commonly used in the lung function literature were also applied, even though I have already pointed out their inadequacies in section 2.2.

4) FEV change was regressed on average cigarette consumption and the mean of the initial and final FEV measurements, i.e. $(M_1+M_2)/2$.

5) The residuals of final on initial FEV were regressed on average cigarette consumption.

6.4.2 Results and discussion

Table 6-16 shows the results obtained using the various regression methods.

Table 6-16 Regression coefficients of FEV change on initial FEV and cigarette consumption

Variable		No correction for measurement error (Method 1)	Correction using G (Method 2)	Omitting initial FEV from model (Method 3)	$(M_1+M_2)/2$ as a proxy of M_1 (Method 4)	Regressing residuals (Method 5)
Initial FEV	β	-0.236	-0.127	NA	0.193	-0.258†
	SE	0.081	0.094		0.083	0.072
	p	0.004	0.2		0.02	0.0005
Cigarette Consumption	β	-0.0032	-0.0068	-0.0110	-0.0162	-0.0026
	SE	0.0059	0.0061	0.0054	0.0057	0.0052
	p	0.6	0.3	0.04	0.005	0.6

NA Not applicable
† Coefficient of change on initial FEV without cigarette consumption in the model

The effect of proposed corrections for measurement error on the regression coefficient of lung function change on initial lung function follow the same pattern as previously described (Section 6.2). The regression coefficients of change on cigarette consumption vary widely between methods. The uncorrected analysis (Method 1) resulted in a regression coefficient close to zero and not statistically significant. Correction by the generalizability coefficient (Method 2) resulted in an estimate of the effect of

cigarette smoking more than twice that obtained from the uncorrected analysis, and a smaller p-value. Initial lung function might be omitted as a predictor of lung function change in the model because of prior information (Method 3) or subsequent to demonstrating that it was not significant in method 2. Its omission resulted in a more negative regression coefficient for cigarette smoking than that obtained from method 2, and of the same order of magnitude as that on the full sample of 433 men (Table 6-12). Moreover, it was statistically significant. Correcting for measurement error by using the mean of the initial and final FEV as a measure of the initial level (Method 4) resulted in a slightly more negative coefficient for cigarette smoking than that of method 3, whereas the 2-stage procedure of method 5 resulted in a coefficient very close to zero. Obviously, biases in the opposite direction would have occurred with methods 1, 4, and 5 if the correlation between cigarette smoking and initial lung function was negative rather than positive as in the example used.

To determine how sensitive the correction procedure is to the value of the reliability coefficient, the analysis was repeated using values of G of 0.85 and 0.95. Both the regression coefficients for initial FEV and cigarette consumption were highly dependent on the value of G used (Table 6-17).

The amount of random error of measurement in cigarette smoking data was unknown. Table 6-18 shows the result if the reliability coefficient is assumed to be 0.90, and using the estimated reliability coefficient of 0.899 for initial FEV.

Table 6-17 Sensitivity of correction of regression analyses using different values of the reliability coefficient (G)

Variable		Correction using G of: 0.85	0.899 (from table 6-16)	0.95
Initial FEV	β	-0.060	-0.127	-0.185
	SE	0.105	0.094	0.087
	p	0.6	0.2	0.03
Cigarette consumption	β	-0.0090	-0.0068	-0.0049
	SE	0.0063	0.0061	0.0059
	p	0.2	0.3	0.4

Table 6-18 Consequences of error in both predictors

Variable		Correction using G of 0.899 for FEV and: Smoking assumed error-free (from table 6-16)	Smoking G assumed to be 0.9
Initial FEV	β	-0.127	-0.120
	SE	0.094	0.098
	p	0.2	0.2
Cigarette consumption	β	-0.0068	-0.0078
	SE	0.0061	0.0070
	p	0.3	0.3

Correction for the assumed random error in the smoking measurement increased the absolute magnitude of the regression coefficient for smoking and its standard error, leaving statistical significance

unaltered. The correction procedure also decreased the absolute magnitude of the regression coefficient for initial FEV and slightly increased its standard error.

6.4.3 Conclusion

Correction for random error of measurement in a confounder (initial lung function) has an important effect on the assessment of the relationship between a dependent variable (lung function change) and the predictor of interest (environmental exposure, in this case cigarette smoking). This occurs even with a reliability coefficient for the confounder as high as 0.9. Uncorrected OLS regression analysis, or other methods used in the literature (method 4 or 5) may result in biased estimates of the effect of environmental exposure. Correction for random error of measurement using the reliability coefficient (method 2) is suggested for future general use, with the caution that it can only be as good as the accuracy of the estimate of the reliability coefficient. In the particular case of FEV change data in young men, the assumption that the partial regression of lung function change on initial measurement is zero seems justified. Initial lung function is then best omitted from the analysis (method 3) and lung function change regressed only on the exposure variables.

CHAPTER 7 THE CONSEQUENCES OF RANDOM ERROR OF MEASUREMENT IN A DEPENDENT VARIABLE: THE RELATIONSHIP BETWEEN LUNG FUNCTION CHANGE AND ENVIRONMENTAL EXPOSURE

The consequences of measurement error in predictors in regression analyses were explored in chapters 3 and 6. The purpose of this chapter is to examine the consequences of measurement error in a dependent variable. For this purpose, I use the simple situation in which a dependent variable measured with error is related to one predictor variable, considered free of measurement error. Some of the concepts will be demonstrated by the examination of the relationship between FEV change and cigarette smoking in the van Doorn data set. The consequences of error will be explored for exposure-effect and exposure-response relationships, both of which are explained in section 1.4. To summarise, exposure-effect relationships examine how the mean value of the continuous test results relates to the predictor variable, for example by using regression analysis. Exposure-response relationships show how the proportion of individuals with abnormal test results relates to exposure, for example using logistic regression.[28]

7.1 RANDOM ERROR OF MEASUREMENT IN A CONTINUOUS DEPENDENT VARIABLE: EXPOSURE-EFFECT RELATIONSHIPS

It is important to note that the estimate of a regression coefficient of a dependent variable on a predictor is not biased when the dependent variable is measured with error. This is evident from the formulae for regression coefficients which contain no terms for

the variance of the dependent variable (Section 3.1). Therefore the regression coefficient estimate cannot be biased by inflation of the variance of the dependent variable due to measurement error.

On the other hand, the conditional or residual variance around the regression is increased if the dependent variable is measured with error. The reason is evident from consideration of the general model:

$$Y = \alpha + \beta X + \varepsilon \qquad (7:1)$$

where Y = dependent variable, measured without error

X = predictor (environmental exposure)

ε = residual variance

If Y is measured with random error δ so that $Y^* = Y + \delta$, then equation 7:1 becomes

$$Y^* = \alpha + \beta X + \varepsilon + \delta \qquad (7:2)$$

The residual variance increases from σ_ε^2 in equation 1 to $\sigma_\varepsilon^2+\sigma_\delta^2$ in equation 7:2. Therefore, when the dependent variable is measured with error, the standard error of an estimate of the regression coefficient (residual variance/$n\hat{\sigma}_X^2$) is <u>increased</u> and statistical power is compromised. This consequence of measurement error cannot be remedied during analysis and requires improved reliability of

measurement at the data collection stage, for example by better methods, quality control, or the use of repeat measurements.

Another consequence of measurement error in the dependent variable is attenuation of the amount of variation of the dependent variable explained by the predictor (ρ^2). The reason is as follows:

$$\rho^2 = \frac{\sigma_{YX}^2}{\sigma_Y^2 \sigma_X^2} \qquad (7:3)$$

If Y is measured with error δ as Y*, then

$$\rho^{*2} = \frac{\sigma_{YX}^2}{\sigma_{Y^*}^2 \sigma_X^2} \qquad (7:4)$$

As discussed in chapter 3, the covariance term (σ_{YX}) is unaltered by random error of measurement. Because $\sigma_{Y^*}^2 = \sigma_Y^2 + \sigma_\delta^2$, it follows that $\rho^{*2} < \rho^2$. The amount of variation in the dependent variable which is explained by the predictor is underestimated. By analogy with methods described in section 3.2, an unbiased estimate of ρ^2 can be obtained from measurements with random error if an estimate of the reliability coefficient (G) is available. To repeat the definition of G from section 3.1.1:

$$G_Y = \frac{\sigma_Y^2}{\sigma_{Y^*}^2} = \frac{\sigma_{Y^*}^2 - \sigma_\delta^2}{\sigma_{Y^*}^2} = 1 - \frac{\sigma_\delta^2}{\sigma_{Y^*}^2} \qquad (7:5)$$

An estimate of ρ^2 can be obtained using Y*, as follows:

$$\hat{\rho}^2 = \frac{\hat{\sigma}^2_{Y^*X}}{\hat{C}_Y \hat{\sigma}^2_{Y^*} \hat{\sigma}^2_X}$$

$$= \frac{\hat{\rho}^{*2}}{\hat{C}_Y} \qquad (7:6)$$

The effect of correction can be illustrated by a simple example. If $\hat{\rho}^{*2} = 0.3$ and $\hat{C}_Y = 0.5$, $\hat{\rho}^2 = 0.6$. In words, whereas the predictor explained only 30% of the variance of measurements, it explains 60% of the variability of the true dependent variable. Obviously, if measurement error also occurs in the exposure variable (X) it can be dealt with simultaneously using the reliability coefficient of X.

7.2 RANDOM ERROR OF MEASUREMENT IN A DICHOTOMIZED DEPENDENT VARIABLE: EXPOSURE-RESPONSE RELATIONSHIPS

7.2.1 Concepts

Exposure-response relationships show the proportion of individuals with abnormality as a function of the predictor variable, exposure. When a dependent variable is continuous, abnormal values are defined as those falling beyond a specified threshold value of the continuous variable. The apparently binary (normal or abnormal) variable derived from dichotomizing the continuum is referred to as a dichotomized variable.[112] The consequences of random measurement error for exposure-response relationships can be visualized by

considering a regression of the continuous dependent variable on a predictor with residuals normally and homoscedastically distributed around the regression line (Figure 7-1).

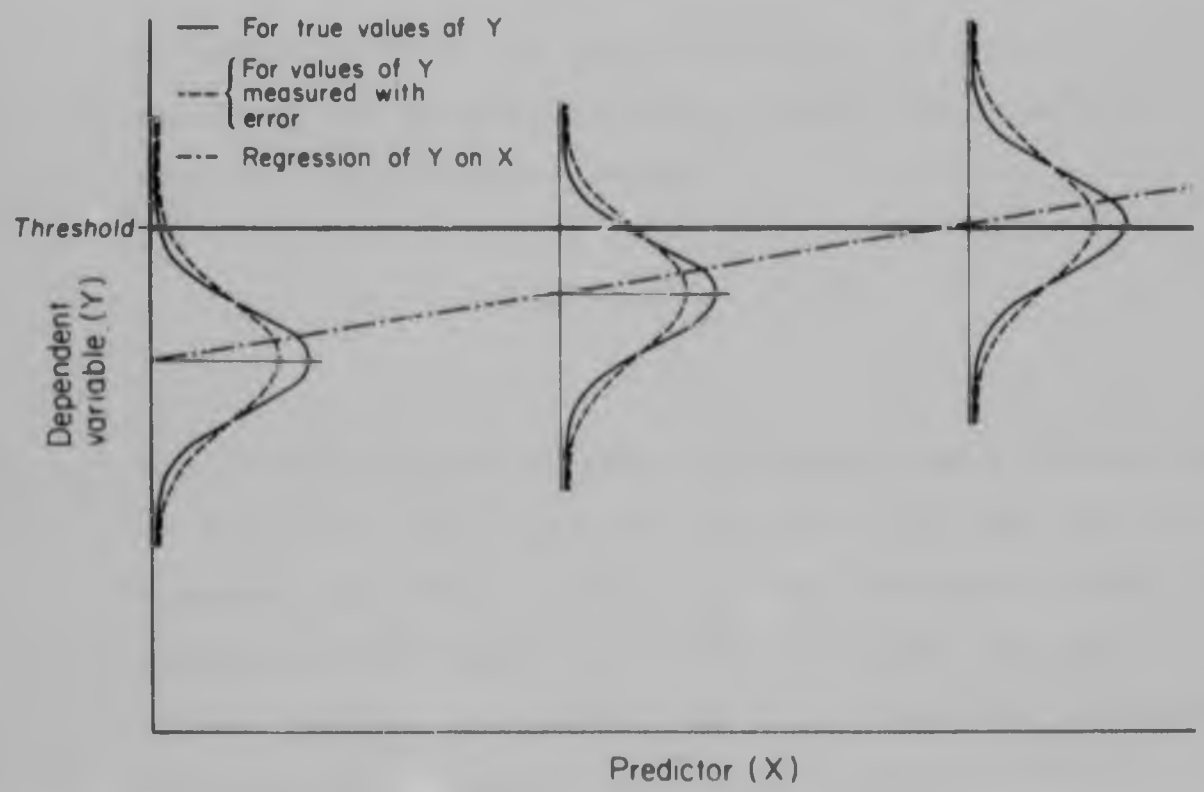

Figure 7-1 Exposure-response relationship from conditional distribution of Y on X

The proportion of individuals beyond a fixed threshold T depends on the value of the dependent variable predicted from the regression equation and the variance of the conditional distribution around that mean value (Figure 7-1, solid line distributions). The proportions can be calculated theoretically from normal distribution theory. The

threshold (T) can be expressed as a standard deviate conditional on X (θ_X) where:

$$\theta_X = \frac{T - (\alpha+\beta X)}{\sigma_{Y|X}} \qquad (7:7)$$

The proportion beyond the specified threshold (1-F(θ)) is then easily determined from the normal probability integral using tables or a suitable computer program, where:

$$1-F(\theta) = \int_{\theta}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}y^2} dy$$

Now, if the conditional variance is increased because of random error of measurement, the calculated proportions less than 50% will be increased and those greater than 50% decreased (dotted line distributions in figure 7-1). This will reduce the slope of the exposure-response relationship, as would also occur with random misclassification in a truly binary response measure.[113,114]

Methods of correcting exposure-response relationships for measurement error follow from a consideration of the statistical background to figure 7-1. If Y is measured as Y* with error:

$$\theta^*_X = \frac{T - (\alpha+\beta X)}{\sigma_{Y^*|X}} \qquad (7:8)$$

Now:
$$\sigma_{Y^*|X} = \sqrt{\sigma_{Y^*}^2(1-\rho^{*2})} \qquad (7:9)$$

For reasons discussed in section 7.1, $(1-\rho^{*2}) > (1-\rho^2)$. Therefore the absolute value of θ_X^* is less than $\theta_{\bar{X}}$ and the normal probability integral will be greater than that for θ_X when $\theta_X > 0$.

There are two methods of correcting for this consequence of measurement error, if G_Y can be estimated:

1) In equation 7:9, the variance estimate $\hat{\sigma}_{Y^*}^2$ can be corrected using equation 7:5 and ρ^{*2} can be corrected using equation 7:6, so that:

$$\hat{\sigma}_{Y|X} = \sqrt{\hat{\sigma}_{Y^*}^2 \hat{G}_Y \left(1-\frac{\hat{\rho}^{*2}}{\hat{G}_Y}\right)}$$

$$= \sqrt{\hat{\sigma}_{Y^*}^2 (\hat{G}_Y - \hat{\rho}^{*2})} \qquad (7:10)$$

$\hat{\sigma}_{Y|X}$ can then be used in equation 7:7 to obtain the corrected proportion beyond the threshold.

2) The <u>observed</u> conditional distribution can be corrected and the number of values of the corrected distribution which exceed the threshold counted. The corrected conditional distribution is obtained by applying a correction to each measurement. This is achieved by:

- expressing each measurement as a residual of the Y* value predicted from the regression equation,
- multiplying the value of each residual by the estimated reliability coefficient of the residuals $(\hat{\sigma}_{Y|X}/\hat{\sigma}_{Y^*|X})$,
- adding the Y* predicted from the regression equation to obtain the corrected value (Y_G).

Expressed as an equation:

$$Y_G = \left(Y^* - (\hat{\alpha}+\hat{\beta}X)\right)\left[\frac{\hat{\sigma}_{Y|X}}{\hat{\sigma}_{Y^*|X}}\right] + (\hat{\alpha}+\hat{\beta}X)$$

$$= (Y^* - \hat{\alpha}+\hat{\beta}X)\left[\sqrt{\frac{\hat{\sigma}_Y^2-\hat{\rho}^{*2}}{1-\hat{\rho}^{*2}}}\right] + (\hat{\alpha}+\hat{\beta}X) \qquad (7:11)$$

Note that I refer to the 'corrected' value as Y_G and <u>not</u> Y. It is important to realise that the procedure corrects the <u>overall</u> distribution but the 'corrected' values are <u>not</u> necessarily better measures of each individual Y. The proportion of Y_G values beyond the threshold at each X value can then be used directly in an appropriate model such as a probit or logistic regression.

The two procedures differ from correction procedures in which the variable measured with error is <u>truly binary</u>.[115-117] The latter correction procedures use two (or more) measurements of the dependent variable and require the assumption that the misclassification errors of the measurements are independent of each other. <u>In dichotomized data this assumption cannot be met</u> because the probability of misclassifying a case across the threshold is clearly dependent on the true value of the underlying continuous variable; the more extreme the true value, the smaller the probability of misclassification. Since this is so for all the measurements, misclassification errors must be correlated.

7.2.2 Application

The data set consisted of measurements on the 433 men in the van Doorn data set initially aged 18-22 years who were followed for 8 or more years (on average 11.3 years). The dependent variable of interest was FEV change between initial and final examination. The exposure variable was the average number of cigarettes smoked per day in the time interval between the two examinations. FEV change was shown to be linearly related to cigarette smoking (Section 6.3.2) with homoscedastic but somewhat kurtosed residuals. FEV change was by and large independent of initial FEV and other potential confounders which were therefore not included in the model (Sections 6.1 and 6.2). The threshold chosen for dichotomy was -0.5 litres. In other words, individuals were classified as abnormal if they lost more than 0.5 litres of FEV over the 11.3 year follow-up period.

The estimate of G for FEV change was obtained from:

$$\hat{G}_Y = \hat{G}_{M_2-M_1} = 1 - \frac{(1-\hat{G}_M)\left(\hat{\sigma}_{M_1}^2+\hat{\sigma}_{M_2}^2\right)}{\hat{\sigma}_{Y^*}^2} \qquad (7:12)$$

where $\hat{\sigma}_{M_1}^2$ = variance of initial FEV (calculated from table 6.2)

$\hat{\sigma}_{M_2}^2$ = variance of final FEV (calculated from table 6.2)

$\hat{\sigma}_{Y^*}^2$ = variance of FEV change (calculated from table 6.2)

$\hat{G}_M$ = reliability coefficient for cross-sectional FEV measurements = 0.899 (from table 5-1)

The reliability coefficient for FEV change was estimated as 0.63. Note that the estimated reliability coefficient for measurement of FEV change is much smaller than that for cross-sectional FEV measurements, for two main reasons. First, the error variance of a measurement of change includes the measurement error of both initial and final measurements. Second, the true variation of FEV change is smaller than that of cross-sectional measurements because there are fewer sources of true variance (Section 2.1). Two exposure-response relationships without correction for measurement error are shown in figure 7-2. Exposure-response relationship A was estimated from the expected proportion beyond the threshold at each level of exposure by applying the normal probability integral to the predicted mean and estimated conditional variance, as described in the explanation following figure 7-1. Exposure-response relationship B was obtained by logistic regression of abnormality (values beyond the threshold) on the exposure variable. Both methods showed similar results, though method B, based on observed data, gave slightly lower prevalences for each value of X than method A. This probably occurred because of deviations from normality of the residuals of FEV change measurements, given smoking. Figure 7-2 also shows exposure-response relationships after correction for measurement error using both methods described in section 7.2.1. Both correction methods resulted in steeper exposure-response relationships with smaller prevalences for non-smokers. For heavy smokers, in whom the prevalence of abnormality was near 50%, the

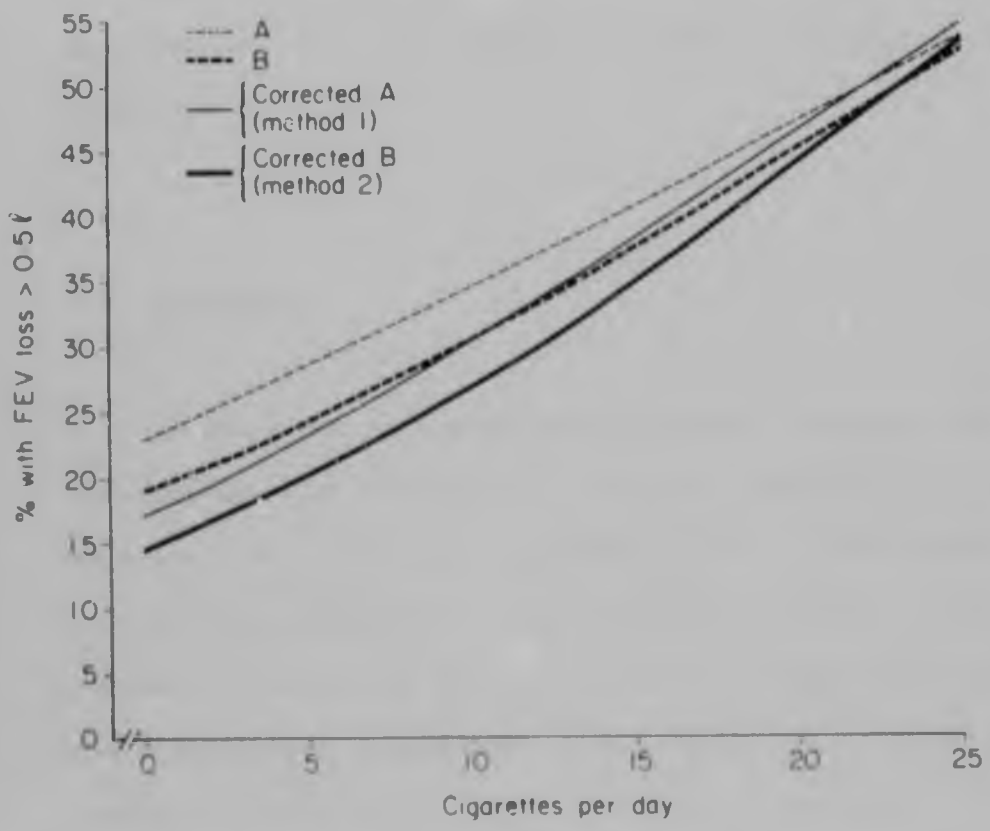

Figure 7-2 Exposure-response relationships for FEV abnormality (loss > 0.5 litre) by smoking, with and without correction for random measurement error

correction procedure, as expected, did not change the estimates appreciably. However for non-smokers, in whom abnormality was less common, the observed prevalences markedly exceeded the estimated true prevalences. Therefore, exposure-response relationships were much steeper after correction. The odds ratios for smokers of 25

cigarettes versus non-smokers were 3.2 for the correction method 1 and 3.6 for method 2 compared to 2.3 and 2.8 for their respective uncorrected methods, A and B. Analyses based on correcting distributions of FEV change in smoking categories gave similar results.

7.3 CONCLUSION

Random error in a continuous dependent variable does not affect its regression on a predictor. However, the amount of variation explained by a set of predictors (ρ^2) is attenuated. If the reliability coefficient of the dependent variable can be estimated, unbiased estimates of ρ^2 are calculable. When continuous variables are dichotomized, measurement error results in a flatter exposure-response relationship with a higher intercept. Though this consequence is similar to the effect of misclassification error in a binary dependent variable, it cannot be corrected using techniques appropriate for binary data. However, conditional distributions can be corrected if the reliability coefficient of the underlying continuous data can be estimated. An unbiased estimate of the exposure-response relationship is then easily calculable. The next chapter explores similar issues for dichotomized dependent variables for which the underlying continuous distributions are immeasurable.

CHAPTER 8 RANDOM ERROR OF MEASUREMENT IN DICHOTOMIZED VARIABLES: EXPOSURE-RESPONSE RELATIONSHIPS FOR PNEUMOCONIOSIS

Exposure-response relationships show the probability of a 'response' in relationship to a given amount of exposure to an environmental agent. In chapter 7 response was defined as a result beyond a specified threshold of a continuous test variable. This chapter deals with the the case where the underlying abnormality must be continuous but cannot be measured as such because of technical reasons or the limitations of human perception. An important environmental health example which will be considered in this chapter is pneumoconiosis detected on the chest radiograph (Section 1.3.2). The major radiographic feature of pneumoconiosis is the profusion of small opacities. By convention this is read on an 11 or 12-point categorical scale. Each category is defined by written criteria and reference radiographs.[21] A common method of analysis is to dichotomize data on the profusion of small opacities at one of the boundaries between categories and regard all of the categories below this boundary as normal and above it as abnormal. Exposure-response relationships are then obtained by examining the proportion of abnormal individuals at each level of exposure.

The literature abounds with examples of inter- and intra-observer variation in reading the profusion of small opacities on radiographs.[31-34] Not only do observers disagree about the presence of abnormality in individual radiographs; they also find very

different prevalences of abnormality in the same sample. For example, in a recent study the prevalence of pneumoconiosis found by several readers ranged from 7 to 23%.[31] Exposure-response relationships obtained by readers also differ markedly.[118] Rather than illustrating these points further by examples from the literature, I now demonstrate them on one particular data set. This data set is also subsequently used to illustrate methods of analysing response variables which appear binary but are truly dichotomized from an underlying immeasurable continuum and are measured with random error.

8.1 AN ILLUSTRATION OF THE CONSEQUENCES OF READER VARIABILITY ON EXPOSURE-RESPONSE RELATIONSHIPS FOR ASBESTOSIS

8.1.1 The Data Set

The data set consisted of all white and coloured workers employed on South African crocidolite and amosite asbestos mines on 30 November 1970, or who commenced employment between that date and 30 November 1975.[119] For each man information was available on the duration of work in asbestos mines. Chest radiographs coinciding with the last date of exposure were read by each of three experienced readers. Readers saw the radiographs independently of each other and without access to information on each man's exposure. Radiographs were read for the presence of small irregular opacities, according to the ILO U/C criteria with the reference X-rays on display.[21] Exposure or radiographic data were missing for less than 1% of the sample, leaving 1 680 men.

TABLE 8-1 Frequency distributions of parenchymal abnormality found by readers A, B, and C (n = 1 680)

International classification of profusion	Boundary	Reader A		Reader B		Reader C	
		No. in category	No. above boundary	No. in category	No. above boundary	No. in category	No. above boundary
0/0 (normal)		1246	-	1576	-	1386	-
	1						
0/1		149	434	56	104	114	294
	2						
1/0		79	285	20	48	106	180
	3						
1/1		130	206	10	28	54	74
	4						
1/2 or above		76	76	18	18	20	20

TABLE 8-1 Frequency distributions of parenchymal abnormality found by readers A, B, and C (n = 1 680)

International classification of profusion	Boundary	Reader A		Reader B		Reader C	
		No. in category	No. above boundary	No. in category	No. above boundary	No. in category	No. above boundary
0/0 (normal)		1246	-	1576	-	1386	-
	←----------1----------→						
0/1		149	434	56	104	114	294
	←----------2----------→						
1/0		79	285	20	48	106	180
	←----------3----------→						
1/1		130	206	10	28	54	74
	←----------4----------→						
1/2 or above		76	76	18	18	20	20

8.1.2 Results

The distribution of findings for each of the 3 readers (called A, B, C) is shown in table 8-1 in which several boundaries are defined. Boundary 1 is between profusion category 0/0 and 0/1, and boundaries 2, 3 and 4 between successively higher profusion categories. Readers found very different prevalences of abnormality in the same radiographs at equivalent boundaries. For example, at boundary 1 reader A found 434 (26%) of the radiographs abnormal, reader B 104 (6%) and reader C 294 (17%). The readers also perceived the 'width' of the categories differently. For example, reader A found many more abnormal than C in categories 0/1, 1/1, and 1/2 or above but fewer than C in category 1/0.

The distribution of the duration of asbestos exposure was extremely skewed to the left. This variable was categorized (T) and

Table 8-2 Frequency distribution of duration of asbestos exposure

Exposure (years) Category	Mean	Frequency No.	%
≤ 1.00	0.4	564	34
1.01- 3.00	1.9	346	21
3.01- 7.00	4.7	346	21
7.01-15.00	10.4	278	17
>15.00	20.8	146	9

exposure-response relationships obtained by plotting the category-specific prevalence of abnormality against the mean duration of exposure in each category. The exposure-response relationship at various boundaries of radiographic abnormality is shown for reader A in figure 8-1.

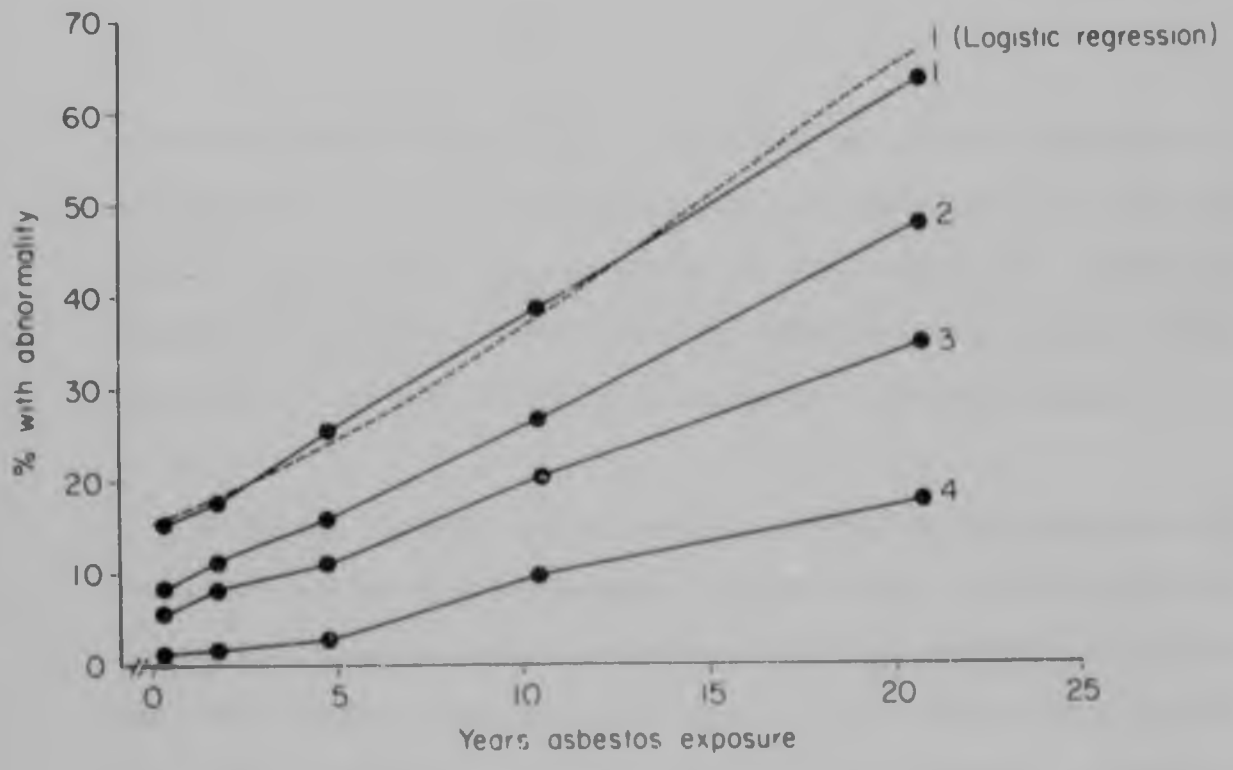

Figure 8-1 Exposure-response relationship for boundaries 1 to 4 (Reader A)

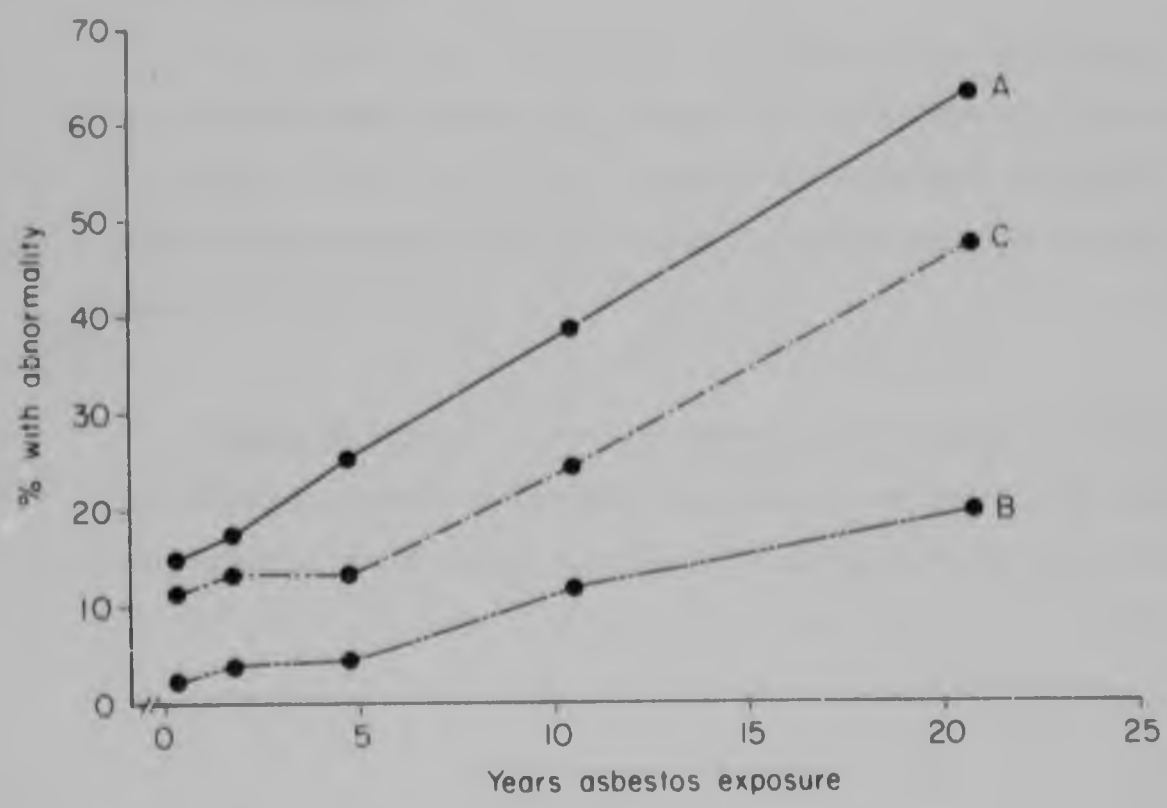

Figure 8-2 Exposure-response relationship for readers A, B, C at boundary 1

The more extreme the boundary chosen, the smaller the prevalence of abnormality in those individuals with negligible exposure and the smaller the increase in prevalence of abnormality with increasing duration of exposure. Note that the exposure-response relationship at boundary 1 is well described by a logistic regression model.

Different readers found markedly different exposure-response relationships at a fixed boundary (Figure 8-2). For example, at boundary 1, reader A found a prevalence of 15% abnormality in men with less than 1 year's exposure, increasing to 63% in those with a mean of 21 years' exposure. At the other extreme, reader B found a prevalence of abnormality of 2% in men exposed for less than 1 year and 20% in men with 21 years' exposure.

8.1.3 Discussion

The reasons for the marked differences in prevalence and exposure-response relationship between the three readers are similar to those which occur when measurable continuous variables are dichotomized, namely threshold choice and random error of measurement.[39,118,120-122]

Threshold choice. Even if boundaries are clearly defined and standardized, readers still may vary widely in whether they classify people above or below this boundary. This could be due to differences

in their visual perception[123] or decision attitude.[124-126] The threshold is therefore determined by the implicit decision made by a reader about the presence or absence of abnormality. The choice of different thresholds at a particular boundary is an obvious reason for readers to find different prevalences and exposure-response relationships.

Random error of measurement. At a fixed threshold, random measurement error in a dichotomized variable results in an increased prevalence of abnormality (if the true prevalence of abnormality is less than 0.5).[39] As discussed in section 7.2, this is because the variance of the underlying continuous variable is increased by random error, thereby increasing the proportion of individuals beyond the threshold. Random error therefore results in a flatter exposure-response relationship with a larger intercept.

Analytic techniques commonly used to obtain exposure-response relationships for pneumoconiosis treat the response measure as if it were truly binary. For example, reader's findings are often combined by a decision rule, for example the majority of several readers find abnormality more severe than a specified boundary.[119] This approach does not take cognizance of the fact that the readers' findings may differ because of their threshold choice or random error. Furthermore, as pointed out in section 7.2.1, procedures to correct for the effect of misclassification in binary data are not appropriate for dichotomized data. The purpose of the next section of this

chapter therefore is to describe analytic techniques to:

- test the relative strength of association between exposure and the readers' radiographic abnormality independent of threshold (Section 8.2.1);
- calculate exposure-response relationships for a reader at any specified threshold (Section 8.2.2);
- correct the exposure-response relationship for random error (Section [illegible].

8.2 METHODS OF ESTIMATING EXPOSURE-RESPONSE RELATIONSHIPS FOR DICHOTOMIZED RESPONSE VARIABLES

In section 7.2 it was shown that exposure-response relationships can be obtained by expressing the threshold (T) as a standardized normal deviate of the conditional distribution of Y on X. The normal probability integral beyond T measures the proportion of cases beyond the threshold if the conditional distribution is normal. From equations 7:7 and 7:9, T can be expressed as a standard normal deviate conditional on X (θ_X) as follows:

$$\theta_X = \frac{T - (\alpha+\beta X)}{\sqrt{\sigma_Y^2(1-\rho^2)}} \qquad (8:1)$$

If X and Y are standardized to have distributions with means equal to 0 and variances equal to 1, then

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2} = \sigma_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \rho \qquad (8:2)$$

chapter therefore is to describe analytic techniques to:

- test the relative strength of association between exposure and the readers' radiographic abnormality independent of threshold (Section 8.2.1),
- calculate exposure-response relationships for a reader at any specified threshold (Section 8.2.2).
- correct the exposure-response relationship for random error (Section 8.2.3).

8.2 METHODS OF ESTIMATING EXPOSURE-RESPONSE RELATIONSHIPS FOR DICHOTOMIZED RESPONSE VARIABLES

In section 7.2 it was shown that exposure-response relationships can be obtained by expressing the threshold (T) as a standardized normal deviate of the conditional distribution of Y on X. The normal probability integral beyond T measures the proportion of cases beyond the threshold if the conditional distribution is normal. From equations 7:7 and 7:9, T can be expressed as a standard normal deviate conditional on X (θ_X) as follows:

$$\theta_X = \frac{T - (\alpha+\beta X)}{\sqrt{\sigma_Y^2(1-\rho^2)}} \qquad (8:1)$$

If X and Y are standardized to have distributions with means equal to 0 and variances equal to 1, then

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2} = \sigma_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \rho \qquad (8:2)$$

An estimate of θ_x can then be expressed as:

$$\hat{\theta}_x = \frac{T-\hat{\rho}x}{\sqrt{1-\hat{\rho}^2}} \qquad (8:3)$$

If ρ can be estimated from the available data, it is then possible to estimate exposure-response relationships at any threshold which is expressed as a standard normal deviate.

8.2.1 Estimating ρ from dichotomized data

The correlation coefficient ρ is not correctly measured by the product moment correlation coefficient between the dichotomized response variable and the continuous exposure variable.[112,127] This is easily demonstrated by considering a standard bivariate normal distribution with a correlation of 0.4 and thresholds defined to give various prevalences of abnormality (Figure 8-3).

As the threshold is changed, the product-moment correlation decreases from a maximum of 0.32 at 50% prevalence to 0.19 at 5% prevalence (Table 8-3).

Table 8-3 Effect of different thresholds in a standard bivariate normal distribution with a correlation of 0.4 (Theoretical calculations)

% prevalence of abnormality (above threshold)	Product moment correlation	Logistic regression coefficient
50	0.32	0.72
10	0.23	0.90
5	0.19	1.01

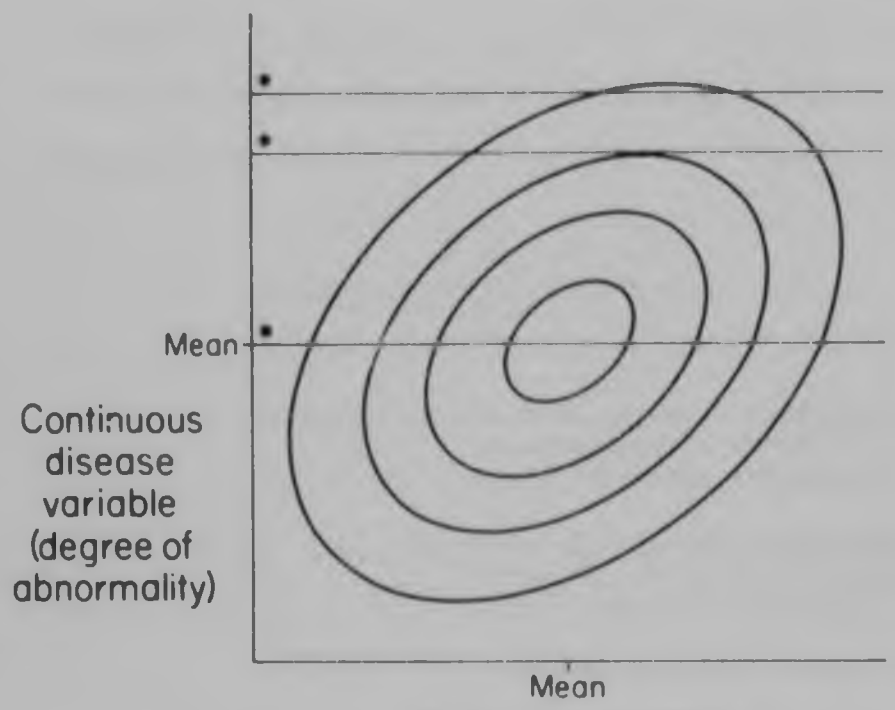

Figure 8-3 Bivariate distribution of a continuous exposure variable and a continuous disease variable, demonstrating different thresholds

The true correlation of 0.4 is underestimated at all thresholds, this phenomenon being more marked the more extreme the threshold. As might be expected, logistic regression coefficients also vary with changes in threshold (Table 8-3).

The correct measure of the correlation between a continuous (exposure) and dichotomized (response) variable is Pearson's biserial correlation coefficient (ρ_{bis})[112,127] which is estimated as follows from independent observations assuming an underlying bivariate normal distribution:

$$\hat{\rho}_{bis} = \frac{\bar{x}_L - \bar{x}_U}{\hat{\sigma}_X} \cdot \frac{\hat{P}_L(1-\hat{P}_L)}{f(\hat{T})} \qquad (8:4)$$

where $\bar{x}_L$ = mean of exposure measurements for observations of the disease variable lying below the threshold

$\bar{x}_U$ = mean of exposure measurements for observations of the disease variable lying above the threshold

$\hat{\sigma}_X$ = standard deviation of the cause variable

$\hat{P}_L$ = proportion of individuals below the threshold

$f(\hat{T}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\hat{T}^2}$ (Probability density function for the normal distribution)

where $\hat{T}$ is the $\hat{P}_L$.100'th percentile of the standard normal distribution

Pearson's biserial correlation coefficient provides a measure of the actual correlation between the exposure and response variables (including random observational error), on the assumption that they have a bivariate normal distribution or have been transformed to one. It is not possible to carry out an actual transformation to bivariate normality because the underlying continuous observations of the response variable are not known. However, since the dichotomization of this variable is invariant under any monotonic transformation, it is sufficient to assume that an appropriate transformation exists.

Its application would result in normality of the marginal distribution of the exposure variable. It is therefore only necessary to check the observations of this variable for approximate normality or transform them to normality. For any particular observer, the random error variance is assumed to be independent of his choice of threshold.

Approximate standard errors for sample biserial correlations can be calculated or obtained from tables.[128] However, when all the observers have examined the same cases, paired standard errors are required if one wishes to determine whether they differ from each other. In the absence of analytic methods, paired standard errors can be estimated by the bootstrap technique.[129,130]

8.2.2 Exposure-response relationships for each reader at the same threshold

To obtain exposure-response relationships for each of several readers at the same threshold, it is necessary to specify a threshold of radiographic abnormality beyond which a response is considered present. A threshold can be chosen at the standard deviation value required to make the overall prevalence a specific percentage. The biserial correlation for each reader together with the arbitrarily chosen threshold can be used to calculate the exposure-response relationship. This is accomplished using equation 8:3 at several exposure values.

Exposure-response relationships can be found for several observers at a particular fixed threshold. A fixed threshold on the underlying

continuum of measurement may be expressed as a different standardized deviate for different readers. For example, an observer who has more random error of measurement will find the underlying continuous variable to have a wider standard deviation and therefore, the standardized deviate will be smaller (Section 7.2.1). Because, in this chapter, the underlying continuum is immeasurable, an indirect method must be used to obtain the standardized deviate for each of several readers at the same true threshold. This method can best be illustrated by considering two observers A and B. Imagine that the threshold has been chosen to give A a specific prevalence of abnormality (e.g. 20%) for which the standardized deviate is T_A (0.84 for 20% prevalence).

$$\text{Now} \quad \sigma_Y^2 = G_A\sigma_A^2 = G_B\sigma_B^2 \qquad (8:5)$$

where σ_Y^2 = variance of the underlying variable measured without error

σ_A^2 and σ_B^2 = variance of the underlying variable measured with error by A or B as indicated by the subscript

G_A and G_B = Reliability coefficient for A or B as indicated by the subscript

Therefore

$$\sigma_B = \sigma_A \sqrt{\frac{G_A}{G_B}} \qquad (8:6)$$

Even if the reliability coefficient for A and B are unknown, it is clear from equation 7:6 that

$$\rho_{YX} = \frac{\rho_{AX}}{\sqrt{G_A}} = \frac{\rho_{BX}}{\sqrt{G_B}} \qquad (8:7)$$

where X is the exposure variable.

Therefore

$$\sqrt{\frac{G_A}{G_B}} = \frac{\rho_{AX}}{\rho_{BX}} \qquad (8:8)$$

The specified threshold for observer A can hence be expressed in standardized normal deviates for B as:

$$T_B = T_A \frac{\rho_{AX}}{\rho_{BX}} \qquad (8:9)$$

Substituting estimates of ρ_{AX} and ρ_{BX} obtained through the use of the biserial correlation coefficient in equation 8:9, one obtains T_B which is then the same true threshold as T_A.

Biserial correlation coefficients and thresholds for each reader can then be used in equation 8:3 to estimate exposure-response relationships for each reader at the same true threshold.

8.2.3 Exposure-response relationships corrected for random measurement error

The estimation of an exposure-response relationship corrected for measurement error is most easily understood by considering the attenuating effect of error on the correlation coefficient (Equation 8:7). If the reliability coefficient is known, the correlation between exposure and the continuum underlying the response measurement

can be corrected for random measurement error using equation 8:7. The reliability coefficient is equivalent to the correlation between repeat observations of the same underlying variable if the measurement errors of the observations are independent on the two occasions (Section 5.1). Assuming bivariate normality, the correlation in the underlying distribution of two dichotomized variables is measured by the tetrachoric correlation.[112,127] An estimate of the tetrachoric correlation ($\hat{\rho}_{tet}$) is derived from the following likelihood equation:

$$L\left(\hat{T}_A, \hat{T}_B, \hat{\rho}_{tet}\right) = d/n$$

based on a four-fold table with cell frequencies a, b, c, d and total frequency n, where:

$\hat{T}_A$ is such that the normal probability integral $F(\hat{T}_A) = (a+c)/n$

and $\hat{T}_B$ is such that the normal probability integral $F(\hat{T}_B) = (a+b)/n$

Though the tetrachoric correlation can be estimated from several approximate formulae, it is best obtained as a maximum likelihood estimate based on an iterative process. The estimate of the tetrachoric correlation can then be used to deattenuate the biserial correlation between exposure and the continuum underlying the response measurements before calculating exposure-response relationships. If observations are made by two observers, on the same set of cases, their tetrachoric correlation can also be used for correcting the biserial correlation using the following method. From equation 8:7 it follows that:

$$\rho_{YX} = \sqrt{\frac{\rho_{AX}\rho_{BX}}{G_A G_B}} \qquad (8:10)$$

Now because $\frac{\rho_{AB}}{\sqrt{G_A G_B}} = 1$ (8:11)

it follows that $\rho_{AB} = \sqrt{G_A G_B}$ (8:12)

Substituting in equation 8:10

$$\rho_{YX} = \sqrt{\frac{\rho_{AX}\rho_{BX}}{\rho_{AB}}} \qquad (8\cdot13)$$

An unbiased estimate of the correlation between exposure and the true continuum of disease can therefore be obtained from equation 8:14 if estimates of ρ_{AX}, ρ_{BX} and ρ_{AB} are available. The unbiased ρ_{YX} can be used to estimate the unbiased exposure-response relationship from equation 8:3.

8.3 AN EXAMPLE: ESTIMATING EXPOSURE-RESPONSE RELATIONSHIPS FOR RADIOGRAPHICALLY DETECTED ASBESTOSIS

The data set used was described in section 8.1.1. The skewed distribution of the duration of asbestos exposure in the data set was transformed to normality by ranking and assigning quantiles of the standard normal distribution to the ranked series. Biserial correlations have been calculated for each reader at different boundaries and between readers at a fixed boundary. Probability estimates are based on bootstrap standard errors using between 150

and 300 replications, taking into account that all of the biserial correlations derive from the same set of cases.

Table 8-4 shows the product-moment and biserial correlation coefficients at each boundary for reader A, who found the highest prevalence of abnormality at each boundary and was therefore expected to show least chance variability of the biserial estimates.[128]

Table 8-4a Association between normalized duration of asbestos exposure and parenchymal abnormality detected by reader A at various boundaries (n = 1 680)

Boundary	Product-moment correlation	Biserial correlation	Standard error
1	0.29	0.40	0.03
2	0.28	0.41	0.03
3	0.23	0.37	0.04
4	0.20	0.42	0.06

Table 8-4b Probabilities of differences between the biserial correlation in table 8-4a arising by chance†

Boundaries ↓ →	2	3	4
1	0.60	0.47	0.62
2	-	0.17	0.80
3	-	-	0.28

† based on Fisher's Z-test using standard errors of the differences derived by bootstrap replications, taking into account that all biserial correlations were calculated from the same data set.

As expected from the theoretical example, the product-moment correlation coefficient decreased from boundary 1 to 4 (Table 8-4a). Biserial correlation coefficients varied between 0.37 and 0.42 (Table 8-4a) with standard errors increasing with more extreme boundaries for which a smaller number of individuals were classified as abnormal. Biserial correlations did not differ significantly from each other (Table 8-4b). For reader B and reader C, biserial correlation coefficients at the 4 boundaries did not show any trends with change in threshold but were more variable than for A, probably due to their small number of abnormal readings. The similarity of biserial correlations for one reader at different boundaries suggests that it is reasonable to assume that the random error variance for a reader is independent of his choice of threshold.

Table 8-5a Biserial correlation coefficients of normalized duration of asbestos exposure and parenchymal abnormality (Boundary 1) for readers A, B, and C

Reader	No. abnormal	Biserial correlation	Standard error
A	434	0.396	0.028
B	104	0.382	0.054
C	294	0.324	0.040

Table 8-5b Probabilities of differences between the biserial correlations in Table 8-5a arising by chance†

	Reader pair		
	AB	AC	BC
Probability	0.79	0.07	0.30

† based on Fisher's Z-test using standard errors of the differences derived by bootstrap replications, taking into account that all readers examined the same radiographs

Table 8-5 shows the biserial correlation between the duration of asbestos exposure and the findings of readers A, B, and C using boundary 1 to maximize the number of abnormals and thereby reduce the standard error of the estimated correlation.[128] Reader C gave the lowest value for the biserial correlation, i.e. 0.32 compared to 0.40 for A, ($p = 0.07$) while the biserial correlations for A and B were similar.

The differences between prevalences and exposure-response relationships initially observed between the readers A and B in section 8.1 are therefore attributable to a different choice of threshold. On the other hand C has a smaller biserial correlation than A. A reader's biserial correlation was presumably attenuated by random error of measurement. If it is assumed that readers have no other method of determining exposure, that reader whose findings correlate most closely with the duration of asbestos exposure (in this case A) is likely to have less observational error than the reader with an appreciably smaller correlation with exposure (reader C). This technique could be used to identify readers with the greatest accuracy, irrespective of threshold choice and without assumptions about reader errors being independent.[131]

Exposure-response relationships calculated from the biserial correlation coefficient to give A an overall prevalence of 26% (i.e. the same as that actually found at boundary 1) agreed closely with logistic regression estimates using the same normalized exposure measurements (Table 8-6). However, neither exposure-response relationship corresponded closely with that based on logistic regression using untransformed exposure data (Figure 8-4).

Table 8-6 Estimates of exposure-response relationship for A from logistic regression on normalized exposure and biserial correlation, at boundary 1

Duration of exposure		Prevalence from	
SD on normalized data	Years	logistic regression	biserial correlation
-1.5	0.0	9.4	8.8
-1.0	0.3	13.0	12.7
-0.5	0.9	17.7	17.8
0.0	2.5	23.6	23.9
0.5	5.5	30.9	31.2
1.0	10.6	39.1	39.0
1.5	16.8	48.1	47.6
..)	22.9	57.2	56.2

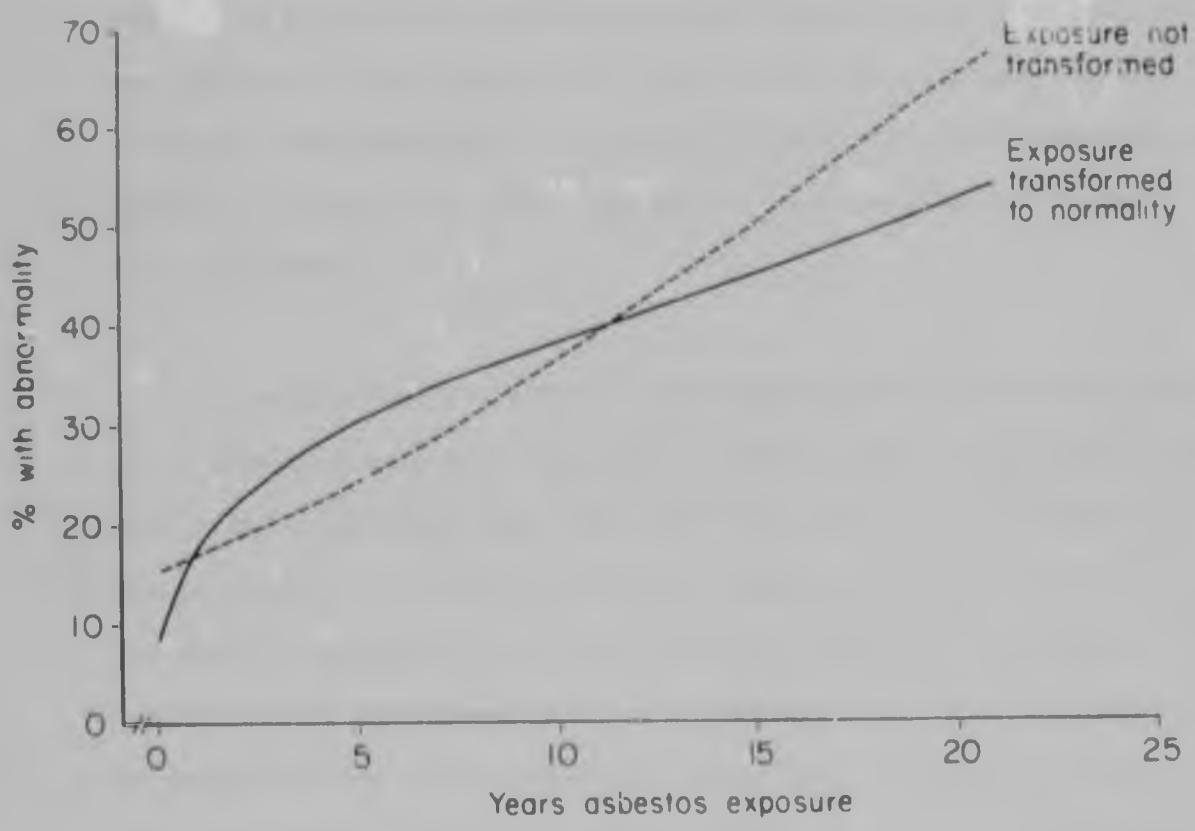

Figure 8-4 Logistic regression of response by reader A at boundary 1 on exposure (repeated from figure 8-1) and on normalized exposure

The unusual shape of the exposure-response relationship using the biserial correlation is probably a consequence of normalizing the very skewed exposure distribution. It could be avoided in future studies by selecting the sample in such a way that exposure is distributed approximately normally.

Exposure-response relationships were calculated from the biserial correlation for each reader at the same threshold which gave reader A an overall prevalence of 20%. These are shown in table 8-7 for all readers and figure 8-5 for readers A and C. The exposure-response relationships for A and B were very similar. Reader C shows a prevalence at low exposure that is almost twice that for A, while the prevalences at high exposure are near 50% and very similar for all readers. The odds ratio of abnormality after 23 years of exposure as against no exposure was therefore almost twice as high for reader A than for reader C.

Tetrachoric correlations for the reader pairs are shown in table 8-8. The corrected biserial exposure-effect correlation coefficients were 0.463 using reader pair AB, 0.442 using AC and 0.428 using BC. Theoretically, all of these corrected biserial estimates should be the same. Differences were small and could be due to sampling variability or non-independence of reader errors. The corrected exposure-response relationships are shown for all reader pairs in table 8-9 and for reader pair AB in figure 8-5. Using any reader pair, the prevalence at low exposure is smaller and the exposure-

Table 8-7 Prevalence of abnormality by exposure at a threshold set so that reader A finds 20% abnormal

Duration of exposure (years)	Prevalence at fixed threshold for: Reader A (overall prev 20.0%)	Reader B (overall prev 20.6%)	Reader C (overall prev 24.5%)
0	5.9	6.7	10.7
0.3	8.9	9.9	14.2
0.9	12.9	13.8	18.4
2.5	18.0	18.9	23.3
5.5	24.2	25.1	28.8
10.6	31.4	31.9	34.8
16.8	39.4	39.7	41.6
22.9	47.9	48.0	48.3

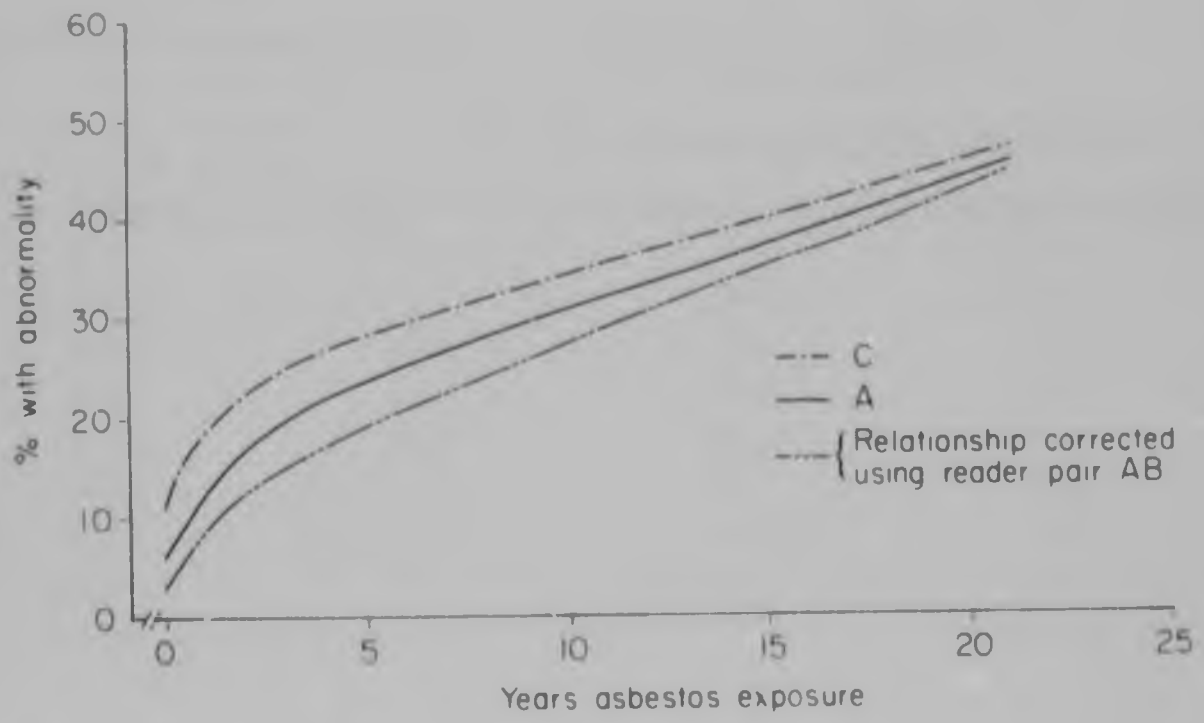

Figure 8-5 Exposure-response relationship for A and C at the same threshold, and the relationship corrected for measurement error using reader pair AB

response relationship steeper than for any of the readers individually. Though prevalences near 50% (at high exposure) are largely unaltered, those at no exposure are about half the prevalences found by each reader individually.

Table 8-8 Tetrachoric correlations (and bootstrap standard errors in brackets) between readers A, B and C

Readers → ↓	B	C
A	0.705(0.037)	0.656(0.031)
B	-	0.675(0.045)

Table 8-9 Estimated prevalence of true abnormality by exposure at the threshold set so that A finds 20% abnormal

Duration of exposure (years)	True prevalence estimated from reader pairs		
	AB (overall prev 16.3%)	AC (overall prev 17.4%)	BC (overall prev 18.1%)
0	2.9	3.7	4.3
0.3	5.1	6.2	6.9
0.9	8.5	9.8	10.7
2.5	13.4	14.7	15.7
5.5	19.8	21.1	22.1
10.6	27.8	29.0	29.8
16.8	37.2	37.9	38.4
22.9	47.4	47.5	47.8

8.4 CONCLUSION

Radiographically detected asbestosis is an example of a variable which is dichotomized for the purpose of analysing its relationship to exposure but which has an underlying immeasurable continuous distribution. Determining exposure-response relationships for dichotomized variables is beset by difficulties arising from random error in the response observations and observer differences in threshold choice. These difficulties can be overcome by using the biserial correlation coefficient as a measure of association between exposure and response. Exposure-response relationships can then be calculated at a specified threshold for each reader using normal distribution theory. The exposure-response relationship can also be corrected for readers' random error of measurement, using the tetrachoric correlation of repeat measurements as the reliability coefficient.

CHAPTER 9 CONCLUSION

This chapter briefly summarizes how the association between environmental exposure and respiratory impairment is affected by measurement error and how these effects can be corrected during data analysis. Some of its implications for future research and methodological development are then explored. This is done for the two important measures of respiratory impairment dealt with in the thesis, namely lung function change (Chapters 2 to 7) and radiologically-detected pneumoconiosis (Chapter 8).

9.1 LUNG FUNCTION CHANGE AND EXPOSURE TO ENVIRONMENTAL AGENTS

9.1.1 Summary: Effects of random error of measurement and correction procedures

Lung function is measured with random error. This has several consequences when lung function change is regressed on initial lung function and environmental exposure.

1) The regression of lung function change on initial lung function is biased because of random error in the initial measurement.

2) If initial lung function and environmental exposure are associated, random error in the initial lung function (the confounder) results in a biased regression of lung function change on environmental exposure.

3) The variance of lung function change explained by environmental

exposure is underestimated if lung function change is measured with random error. Consequently statistical power, i.e. the likelihood of detecting a true association between environmental exposure and lung function change, is reduced.

4) If lung function change is dichotomized to obtain exposure-response relationships, random measurement error causes an underestimation of the slope of the exposure-response relationship or relative risk.

With the exception of the effect on statistical power, all of these consequences of random error of measurement can be corrected at analysis stage. Several correction methods are suggested in the literature. The most attractive of them on theoretical and practical grounds is that using the reliability coefficient, defined as the ratio of the true (error-free) variance to the variance of the measured variable. This is easily estimated as the correlation between repeat measurements of the same true level. Correction for the effect of random measurement error requires multiplying the variance of lung function measurements by the reliability coefficient. The corrected variance is then used to calculate regression or correlation coefficients.

9.1.2 Implications for future research and methodological development

The implications for future research are explored specifically for lung function, but similar suggestions are applicable to epidemiological studies in other areas.

1) The reliability coefficient of lung function measurements should be determined in all studies of the consequences of environmental exposure. The reliability coefficient should then be used to correct analyses for the effect of random error of measurement. This is especially important for confounders, in this case the initial lung function measurement. The reliability coefficient is most easily obtained by repeat measurement on two occasions separated by a time interval chosen to include all sources of random error of measurement but minimizing the chance of true biological change.

2) Random error in the measurement of environmental exposures should also be determined. The regression of lung function (or change) on an environmental exposure variable is attenuated by error in the measurement of environmental exposure. However, statistical power is not affected. The regression coefficient can be deattenuated using the reliability coefficient of the environmental exposure variable. Methods need to be developed for determining the amount of error in environmental exposure assessment. If there are potential confounders, e.g. initial lung function or other exposure variables, the variances of all the predictor variables should be corrected in the variance-covariance matrix.[76] Experience needs to be accumulated on the effect of correction procedures for multiple predictors measured with error. The literature suggests that it may lead to unexpected results.[42,132]

3) Methods for examining interaction while correcting for measurement error need to be developed. The methods in this thesis deal only with a linear additive model. As pointed out in chapter 1,

a major concern in environmental health is the possible interaction between several environmental exposures, or between environmental exposure and initial lung function. Methods for taking account of measurement error in models including interaction terms need to be developed.

4) Studies of the effect of environmental exposure should be designed to minimize random error of measurement, e.g. by repeat measurements. Error is reduced if lung function is calculated as the mean of two (or more) independent measurements. If funds are available for 4 measurements of lung function, it seems preferable to do two at the beginning and two at the end of the study period rather than the more usual design of 4 measurements equally spaced over the exposure period.[99,133] The efficiency of alternative designs with repeat measurements not necessarily at regular intervals needs exploration.

5) Lung function screening programmes should take account of random measurement error. The implications of random measurement error for screening or detection of abnormality have not been considered in this thesis. In essence, random error increases the proportion detected beyond the threshold specified for screening or case detection. This is a reflection of misclassification across the threshold with the false positive rate exceeding the false negative rate. The implications of error for screening have received scant attention in the lung function literature[30] and should be explored, as they have been in other fields.[39,74,134,135]

6) Evaluation of new lung function tests should take account of measurement error. Current theories about the pathogenesis of airway obstruction suggest that the earliest manifestations of abnormality are found in small airways, i.e. those less than 2 mm in diameter.[19] A variety of new tests have been developed in an attempt to detect small airway abnormality. These include maximum expiratory flow volume curves with measurement of flow near the end of expiration, single breath nitrogen tests, frequency dependence tests, isovolumetric comparisons of maximum expiratory flow with gases of different density, and tests of airway reactivity.[8,19] If these tests detect earlier abnormality, they should be more sensitive to the effects of environmental exposure. However, they also exhibit considerable measurement error,[30] which raises several issues:

a) Measurement error is usually quantified as the coefficient of variation, i.e. the standard deviation of the difference between repeat measurements divided by the mean lung function value. For comparison of the reliability of tests, the reliability coefficient seems a much better measure as it indicates the proportion of measurement variance attributable to true variability of the lung function characteristic.

b) Tests for airway disease may be regarded as detecting early reversible pathology on the basis of the reversibility of test results.[136,137] It is obligatory to ensure that reversibility of 'abnormal' results is not simply an artefact due to random error of measurement. New tests involving measurement of change[138] should take account of the possibility that change may be related to the

initial test result because of random error of measurement.

c) New tests need to be evaluated on the basis of whether they predict subsequent morbidity or mortality. If prediction is poor, the test is of little predictive validity. Analyses should be repeated with correction for random error of measurement to determine the prediction of subsequent disease by true variability in the new test. If the corrected predictive ability is good when the uncorrected prediction was poor, the test holds promise if its measurement error can be reduced by technical improvements.

9.2 RADIOGRAPHICALLY-DETECTED PNEUMOCONIOSIS AND ENVIRONMENTAL EXPOSURE

9.2.1 Summary: Effects of random error of measurement and correction procedures

Radiographically-detected pneumoconiosis is an example of a dichotomized variable, i.e. one that is usually read or analysed as a binary variable, but in fact has an underlying immeasurable continuum. The prevalence of abnormality and exposure-response relationships may differ between observers because of differing random error or differences in the threshold implicitly chosen for dichotomy by each observer. The biserial correlation coefficient can be used to measure the correlation between environmental exposure and the continuum (including random error) underlying the dichotomized readings of an observer. Normal distribution theory applied to the

biserial correlation can be used to estimate exposure-response relationships at any required threshold for each reader. Biserial correlation and exposure-response relationships can be corrected for random observational error using the reliability coefficient, calculated as the tetrachoric correlation between repeat observations by the same or different readers.

9.2.2 Implications for future research and methodological development

1) Can a correlation matrix of biserial and tetrachoric correlations be used to calculate multiple or partial effects in the same way as that derived from product-moment correlation coefficients? This question requires statistical research, for example using simulation studies. If biserial and tetrachoric correlations can be used in this way, exposure-response relationships could easily be estimated using partial correlations to take account of confounders.

2) More efficient studies of the effect of environmental exposure on the risk of pneumoconiosis can be designed. The efficiency of studies might be improved in several ways:

a) Investigators studying possible effects of environmental exposure could request observers to read comparatively rare abnormality with a high index of suspicion. This procedure increases statistical power and therefore reduces the sample size requirement. Exposure-response relationships at alternative (e.g. more stringent) thresholds can easily be constructed from the biserial correlation on these data if required.

b) Studies of pneumoconiosis commonly use several observers. The number of observers can be reduced if the objective is to quantify exposure-response relationships, rather than identify affected individuals. To choose the most extreme example, the exposure-response relationship using one reader could be deattenuated using a tetrachoric correlation obtained from his repeat reading of a sample of radiographs, assuming that his observational errors were independent on the two occasions.

3) <u>The recognition that many abnormalities are dichotomized rather than binary requires a reorientation of analytic techniques designed for binary data</u>. Many apparently binary variables, such as clinically detected abnormalities, must have underlying immeasurable continuous distributions and exhibit considerable observer disagreement.(139) Attempts to correct analyses for measurement error using methods designed for binary variables may lead to biased results. The methods I have described for dichotomized data therefore have applicability far beyond that covered by the examples in this thesis.

REFERENCES

1. Geisel T S, (Dr Seuss).
The Lorax
New York: Random House, 1971.

2. World Health Organization.
Health hazards from new environmental pollutants.
Technical Report Series No. 586
Geneva: WHO, 1976.

3. World Health Organization: Suess M J, Huismans J W, eds.
Management of hazardous waste - Policy guidelines
and code of practice.
World Health Organization Regional Publications
European Series No. 14,
Copenhagen: WHO Regional Office for Europe, 1983.

4. Maugh T H.
Just how hazardous are dumps?
Science 1982; 215: 490-2.

5. Joyce C.
Counting the cost of chemical dumps.
New Scientist 1984; 102: 14.

6. Wyndham C H, Irwig L M.
A comparison of the mortality rates of various
population groups in the Republic of South Africa.
S Afr Med J 1979; 55: 796-802.

7. Office of Smoking and Health.
The health consequences of smoking.
Chronic obstructive lung disease.
A report of the Surgeon-General 1984.
Maryland: US Department of Health and
Human Services, 1984: 1-16.

8. Dosman J A, Cockcroft D W, Hoeppner V H.
Airways obstruction in occupational pulmonary disease.
Med Clin North Am 1981; 65: 691-706.

9. Speizer F F, Tager I B.
Epidemiology of chronic mucus hypersecretion
and obstructive airways disease.
Epidemiologic Reviews 1979; 1: 124-142.

10. Hunter R B.
Developments in tobacco products and the possibility of
'lower-risk' cigarettes.
Second Report of the Independent Scientific Committee of
Smoking and Health.
London: Department of Health and Social Security, 1979.

11. Kilburn K H.
Lung disease: lessons from the past - keys to the future.
Arch Environ Health 1984; 39: 139-143.

12. World Health Organization.
Health effects of combined exposures in the work environment.
Technical Report Series 662.
Geneva: WHO, 1981.

13. Acheson E D, Gardner M J.
Exposure limits - the scientific criteria.
In: McDonald J C, ed. Recent Advances in
Occupational Health I.
Edinburgh: Churchill Livingstone, 1981: 257-269.

14. McDonald J C, Harrington J M.
Early detection of occupational hazards.
J Soc Occup Med 1981; 31: 93-98.

15. Nicholson W J.
Research issues in occupational and environmental cancer.
Arch Environ Health 1984; 39: 190-202.

16. Smith M E, Newcombe H B.
Automated follow-up facilities in Canada for monitoring
delayed health effects.
Am J Public Health 1980; 70: 1261-1268.

17. Siemiatycki J.
An epidemiological approach to discovering occupational
carcinogens by obtaining better information on
occupational exposures.
In: Harrington, J M ed.
Recent advances in occupational health II.
Edinburgh: Churchill-Livingstone, 1984: 143-157.

18. Blot W J, Fraumeni J F Jr.
Studies of respiratory cancer in high risk communities.
J Occup Med 1979; 21: 276-278.

19. Ingram R H Jr, McFadden E R Jr.
Physiologic measurements providing enhanced sensitivity
in detecting early effects of inhalants.
In: Weill H, Turner-Warwick M, eds.
Occupational lung diseases: research approaches and methods.
New York: Marcel Dekker, 1981: 87-98.

20. Speizer F.
Environmental lung disease.
In: Petersdorf R G, Adams R D, Braunwald E,
Isselbacher K J, Martin J B, Wilson J D, eds.
Harrison's Principles of internal medicine. Tenth edition.
New York: McGraw-Hill Book Company, 1983: 1524-1532.

21. International Labour Office.
ILO U/C International classification of radiographs
of pneumoconioses 1971.
Geneva: International Labour Office, 1972.

22. American Thoracic Society (Committee of the Scientific Assembly
on Environmental and Occupational Health).
Guidelines as to what constitutes an adverse respiratory
health effect, with special reference to air pollution.
ATS News 1984; Winter: 6-8.

23. Liddell F D K, McDonald J C.
Radiological findings as predictors of mortality in
Quebec asbestos workers.
Br J Ind Med 1980; 37: 257-267.

24. Peto R, Speizer F E, Cochrane A L, et al.
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.

25. Ferris B G Jr, Speizer F E, Worcester J, Chen F Y.
Adult mortality in Berlin, N H, from 1961 to 1967.
Arch Environ Health 1971; 23: 434-439.

26. Ashley F, Kannel W B, Sorlie P D, Masson R.
Pulmonary function: Relation to aging, cigarette habit, and mortality. The Framingham Study.
Ann Intern Med 1975; 82: 739-745.

27. Beaty T H, Cohen B H, Newill C A, Menkes H A, Diamond E L, Chen C J.
Impaired pulmonary function as a risk factor for mortality.
Am J Epidemiol 1982; 116: 102-113.

28. World Health Organization.
Principles and methods for evaluating the toxicity of chemicals - Part 1.
Environmental Health Criteria 6.
Geneva: WHO, 1978: 25-28.

29. Beveridge W I B.
The art of scientific investigation.
London: Heinemann, 1961: 106.

30. Becklake M R, Permutt S.
Evaluation of tests of lung function for 'screening' for early detection of chronic obstructive lung disease.
In: Macklem P T, Permutt S, eds.
The lung in the transition between health and disease.
New York: Marcel Dekker, 1979: 345-387.

31. Musch D C, Landis J R, Higgins I T T, Gilson J C, Jones R N.
An application of kappa-type analyses to interobserver variation in classifying chest radiographs for pneumoconiosis.
Statistics in Medicine 1984; 3: 73-83.

32. Amandus H E, Hanke W, Kullman G, Reger R B.
A re-evaluation of radiological evidence from a study of U S strip coal miners.
Arch Environ Health 1984; 39: 346-351.

33. Liddell F D K, Lindars D C.
An elaboration of the ILO classification of simple pneuomoconiosis.
Br J Ind Med 1969; 26: 89-100.

34. Eyssen G McK.
Development of radiographic abnormality in chrysotile miners and millers.
Chest 1980; 78 (Suppl): 411-414.

35. Musk A W, Peters J M, Wegman D H, Fine L J.
Pulmonary function in granite dust exposure: A four-year follow-up.
Am Rev Respir Dis 1977; 115: 769-776.

36. Graham W G B, O'Grady R V, Dubuc B.
Pulmonary function loss in Vermont granite workers. A long-term follow-up and critical reappraisal.
Am Rev Respir Dis 198[illegible]; 123: 25-28.

37. Douglas D B, Douglas R B, Oakes D, Scott G.
Pulmonary function of London firemen.
Br J Ind Med 1985; 42: 55-58.

38. Diem J E, Jones R N, Hendrick D J, et al.
Five-year longitudinal study of workers employed in a new toluene diisocyanate manufacturing plant.
Am Rev Respir Dis 1982; 126: 420-428.

39. Gardner M J, Heady J A.
Some effects of within-person variability in epidemiological studies.
J Chronic Dis 1973; 26: 781-795.

40. Davis C E.
The effect of regression to the mean in epidemiologic and clinical studies.
Am J Epidemiol 1976; 104: 493-498.

41. Healy M J R, Goldstein H.
Regression to the mean.
Ann Hum Biol 1978; 5: 277-280.

42. Kupper L L.
Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies.
Am J Epidemiol 1984; 120: 643-648.

43. Armitage P.
Statistical methods in medical research.
Oxford: Blackwell Scientific Publications, 1971: 275-278.

44. Anderson S, Auquier A, Hauck W W, Oakes D, Vandaele W, Weisberg H I.
Techniques for Bias Reduction.
Statistical methods for comparative studies.
New York: John Wiley & Sons, 1980: 7-17.

45. Goldstein H.
Some models for analysing longitudinal data on education attainment.
J R Statist Soc A 1979; 142: 407-442.

46. Buist A S.
Evaluation of lung function: concepts of normality.
In: Simmons D H, ed. Pulmonology, Volume 4.
New York: John Wiley, 1982: 141-165.

47. Hofman A,
Change viewed on the level.
Int J Epidemiol 1983; 12: 391-392.

48. Weisberg H I.
Statistical adjustments and uncontrolled studies.
Psychol Bull 1979; 86: 1149-1164.

49. Fletcher C, Peto R, Tinker C, Speizer F E.
The natural history of chronic bronchitis and emphysema -
An eight-year study of early chronic obstructive lung disease
in working men in London.
Oxford: Oxford University Press, 1976: 195-208.

50. Oldham P D.
A note on the analysis of repeated measurements of
the same subjects.
J Chronic Dis 1984; 15: 969-977.

51. Svärdsudd K, Blomqvist N.
A new method for investigating the relation between change
and initial value in longitudinal blood pressure data.
Scand J Soc Med 1978; 6: 85-95.

52. Brinkman G L, Block D L, Cress C.
Effects of bronchitis and occupation on pulmonary
ventilation over an 11-year period.
J Occup Med 1972; 14: 615-620.

53. Ferris B G Jr, Chen H, Puleo S, Murphy R L H Jr.
Chronic nonspecific respiratory disease in Berlin,
New Hampshire, 1967-1973. A further follow-up study.
Am Rev Respir Dis 1976; 113: 475-485.

54. Higgins I T T, Gilson J C, Ferris B G Jr. Waters M E,
Campbell H, Higgins M W.
Chronic respiratory disease in an industrial town:
a nine-year follow-up study. Preliminary report.
Am J Public Health 1968; 58: 1667-1676.

55. Pham Q T, Mastrangelo G, Chau N, Haluszka J.
Five year longitudinal comparison of respiratory symptoms and function in steelworkers and unexposed workers.
Bull Europ Physiopath Resp 1979; 15: 469-480.

56. Poukkula A, Huhti E, Mäkaräinen M.
Chronic respiratory disease among workers in a pulp mill. A ten-year follow-up study.
Chest 1982; 81: 285-289.

57. Musk A W, Peters J M, Bernstein L, Rubin C, Monroe C B.
Pulmonary function in firefighters: a six-year follow-up in the Boston Fire Department.
Am J Ind Med 1982; 3: 3-9.

58. Love R G , Miller B G.
Longitudinal study of lung function in coal-miners.
Thorax 1982; 37: 193-197.

59. Siracusa A, Cicioni C, Volpi R, et al.
Lung function among asbestos cement factory workers: cross-sectional and longitudinal study.
Am J Ind Med 1984; 5: 315-325.

60. Beaty T H, Menkes H A, Cohen B H, Newill C A.
Risk factors associated with longitudinal change in pulmonary function.
Am Rev Respir Dis 1984; 129: 660-667.

61. Dontas A S, Jacobs D R Jr, Corcondilas A, Keys A, Hannan P.
Longitudinal versus cross-sectional vital capacity changes and affecting factors.
J Gerontol 1984; 39: 430-438.

62. Sparrow D, Bossé R, Rosner B, Weiss S T.
The effect of occupational exposure on pulmonary function:
A longitudinal evaluation of fire fighters and nonfire fighters.
Am Rev Respir Dis 1982; 125: 319-322.

63. Kauffmann F, Drouet D, Lellouch J, Brille D.
Occupational exposure and 12-year spirometric changes
among Paris area workers.
Br J Ind Med 1982; 39: 221-232.

64. 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.

65. Woolf C R, Zamel N.
The respiratory effects of regular cigarette smoking in women.
A five-year prospective study.
Chest 1980; 78: 707-713.

66. Jedrychowski W.
A consideration of risk factors and development
of chronic bronchitis in a five-year follow-up study of
an industrial population.
J Epidemiol Community Health 1979; 33: 210-214.

67. Krzyzanowski M.
Changes of ventilatory capacity in an adult population
during a five year period.
Bull Europ Physiopath Resp 1980; 16: 155-170.

68. Taylor R G, Joyce H, Gross E, Holland F, Pride N B.
Bronchial reactivity to inhaled histamine and annual rate
of decline in FEV_1 in male smokers and ex-smokers.
Thorax 1985; 40: 9-16.

69. McCarthy D S, Craig D B, Cherniack R M.
Effect of modification of the smoking habit on lung function.
Am Rev Respir Dis 1976; 114: 103-113.

70. Bosse R, Sparrow D, Rose C L, Weiss S T.
Longitudinal effect of age and smoking cessation
on pulmonary function.
Am Rev Respir Dis 1981; 123: 378-381.

71. Rosner B.
The analysis of longitudinal data in epidemiologic studies.
J Chronic Dis, 1979; 32: 163-173.

72. Tashkin D P, Clark V A, Coulson A H, et al.
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.

73. Bronowski J.
The ascent of man.
London: British Broadcasting Corporation, 1973: 358.

74 Shepard D S.
Reliability of blood pressure measurements: implications for
designing and evaluating programs to control hypertension.
J Chronic Dis 1981; 34: 191-209.

75. Richardson D H, Wu D.
Least squares and grouping method estimators in
the errors in variables model.
J Am Statist Assoc 1970; 65: 724-748.

76. Warren R D, White J K, Fuller W A.
An errors-in-variables analysis of managerial role performance.
J Am Statist Assoc 1974; 69: 886-93.

77. Fuller W A, Hidiroglou M A.
Regression estimation after correcting for attenuation.
J Am Statist Assoc 1978; 73: 99-104.

78. Werts C W, Linn R L.
A general linear model for studying growth.
Psychol Bull 1970; 73: 17-22.

79. Wu M, Ware J H, Feinleib M.
On the relation between blood pressure change and initial value.
J Chronic Dis 1980; 33: 637-644.

80. Blomqvist N.
On the relation between change and initial value.
J Am Statist Assoc 1977; 72: 746-749.

81. Blomqvist N, Svärdsudd K.
A new method for investigating the relation between change and initial value in longitudinal blood pressure data.
II. Comparison with other methods.
Scand J Soc Med 1978; 6: 125-129.

82. Capocaccia R, Mariotti S, Menotti A, Verdecchia A.
Evaluation of risk factor variations in relation to their baseline values in a controlled preventive trial.
Application to the Rome project of CHD prevention.
J Chronic Dis 1982; 35: 509-520.

83. Snedecor G W, Cochran W G.
Stastical methods. 6th ed.
Iowa: The Iowa State University Press, 1967: 164-166.

84. Lord F M, Novick M R.
Statistical theories of mental test scores.
Reading, Massachusetts:
Addison Wesley, 1968: 198-223.

85. Greenland S, Neutra R.
Control of confounding in the assessment of medical technology.
Int J Epidemiol 1980; 9: 361-367.

86. Maddala G S.
Econometrics.
New York: McGraw-Hill, 1977; 296-300.

87. Madansky A.
The fitting of straight lines when both variables are
subject to error.
J Am Statist Assoc 1959; 53: 173-205.

88. Republic of South Africa.
Report of the Department of Mineral and Energy Affairs, 1982.
Pretoria: Government Printer, 1983.

89. Republic of South Africa.
Report of the Miners' Medical Bureau, 1967.
Pretoria: Government Printer, 1967.

90. American Thoracic Society
ATS Statement - Snowbird workshop on standardization
of spirometry.
Am Rev Respir Dis 1979; 119: 831-838.

91. Peslin R, Bohadana A, Hannhart B, Jardin P.
Comparison of various methods for reading maximal expiratory
flow-volume curves.
Am Rev Respir Dis 1979; 119: 271-277.

92. Wright B M, McKerrow C B.
Maximum forced expiratory flow rates as a measure of ventilatory capacity.
Br Med J 1959; 2: 1041-1047.

93. Page-Shipp R J, Harris E.
A study of the dust exposure of South African white gold miners.
Journal of the South African Institute of Mining and Metallurgy 1972; 73: 10-24.

94. Beadle D G.
An epidemiological study of the relationship between the amount of dust breathed and the incidence of silicosis in South African gold miners.
In: Davies C N, ed. Inhaled particles and vapours II.
Oxford: Pergamon Press, 1965: 479-491.

95. Burrows B, Cline M G, Knudson R J, Taussig L M, Lebowitz M D.
A descriptive analysis of the growth and decline of the FVC and FEV_1.
Chest 1983; 83: 717-724.

96. Krige D G.
Lognormal-de Wijsian geostatistics for ore evaluation.
Johannesburg: South African Institute of Mining and Metallurgy 1981: 1-39.

97. Stebbings J H Jr.
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.

98. Tashkin D P, Detels R, Coulson A H, Rokaw S N, Sayre J W.
The UCLA population studies of chronic obstructive respiratory disease. II. Determination of reliability and estimation of sensitivity and specificity.
Environ Res 1979; 20: 403-424.

99. Berry G.
Longitudinal observations. Their usefulness and limitations with special reference to the forced expiratory volume.
Bull Europ Physiopath Resp 1974; 10: 643-655.

100. Koutsoyiannis A.
Theory of Econometrics. An introductory exposition of econometric methods. 2nd ed.
London: MacMillan Press 1977: 185-186.

101. Chatfield C.
The analysis of time series: an introduction. 2nd ed.
London: Chapman and Hall 1980: 44-50.

102. Snedecor G W, Cochran W G.
Statistical methods. 6th ed.
Iowa: The Iowa State University Press 1967: 447-448.

103. Fletcher C, Peto R, Tinker C, Speizer F E.
The natural history of chronic bronchitis and emphysema.
An eight-year study of early chronic obstructive lung disease in working men in London.
Oxford: Oxford University Press, 1976: 183-186.

104. Glindmeyer H W III.
Predictable confusion.
J Occup Med 1981; 23: 845-849.

105 Knudson R J, Lebowitz M D, Holberg C J, Burrows B.
Changes in the normal maximal expiratory flow-volume curve with growth and aging.
Am Rev Respir Dis 1983; 127: 725-734.

106. Gregg I, Nunn A J.
Peak expiratory flow in normal subjects.
Br Med J 1973; 3: 282-284.

107. Brooks A G F, Waller R E.
Peak flow measurements among visitors to a public health exhibition.
Thorax 1972; 27: 557-562.

108. Hankinson J L, Reger R B, Fairman R P, Lapp N L, Morgan W K C.
Factors influencing expiratory flow rates in coal miners.
In: Walton W H, ed. Inhaled particles IV, Part 2.
Oxford: Pergamon Press, 1977: 737-752.

109. Beck G J, Maunder L R, Schachter E N.
Cotton dust and smoking effects on lung function in cotton textile workers.
Am J Epidemiol 1984; 119: 33-43.

110. Wiles F J, Faure M H.
Chronic obstructive lung disease in gold miners.
In: Walton W H, ed. Inhaled particles IV, Part 2.
Oxford: Pergamon Press, 1977: 727-735.

111. Theriault G P, Peters J M, Johnson W M.
Pulmonary function in granite shed workers of Vermont.
Arch Environ Health 1974; 28: 18-22.

112. Johnson N L, Kotz S.
Distributions in Statistics: continuous multivariate distributions.
New York: John Wiley & Sons 1972: 117-122.

113. Goldberg J D.
The effect of misclassification on the bias in the difference between two proportions in the four-fold table.
J Am Statist Assoc 1975; 70: 561-567.

114. Copeland K T, Checkoway H, McMichael A J, Holbrook R H.
Bias due to misclassification in the estimation of relative risk.
Am J Epidemiol 1977; 105: 488-495.

115. Hui S L, Walter S D.
Estimating the error rates of diagnostic tests.
Biometrics 1980; 36: 167-171.

116. Marshall J R, Graham S.
Use of dual responses to increase validity of case-control studies.
J Chronic Dis 1984; 37: 125-136.

117. Walter S D.
Commentary on 'use of dual responses to increase validity of case-control studies'.
J Chronic Dis 1984; 37: 137-139.

118. Jacobsen M.
Quantifying radiological changes in simple pneumoconiosis.
Appl Statist 1975; 24: 229-249.

119. Irwig L M, du Toit R S J, Sluis-Cremer G K, et al.
Risk of asbestosis in crocidolite and amosite mines in South Africa.
Annals of the New York Academy of Sciences 1979; 330: 35-52.

120. Greene R, Boggis C, Jantsch H.
Asbestos-related pleural thickening: effect of threshold criteria on interpretation.
Radiology 1984; 152: 569-573.

121. Harber P, Schnur R, Emery J, Brooks S, Ploy-song-sang Y.
Statistical 'biases' in respiratory disability determinations.
Am Rev Respir Dis 1983; 128: 413-418.

122. Cary J, Huseby J, Culver B, Kosanke C Jr.
Variability in interpretation of pulmonary function tests.
Chest 1979; 76: 389-390.

123. Newell R R, Garneau R.
The threshold visibility of pulmonary shadows.
Radiology 1951; 56: 409-414.

124. Lusted L B.
Decision-making studies in patient management.
N Engl J Med 1971; 284: 416-424.

125. Christensen-Szalanski J J J, Diehr P H, Bushyhead J B, Wood R W.
Two studies of good clinical judgment.
Med Decis Making 1982; 2: 275-283.

126. Wilson M E, Williams N B, Baskett P J F, Bennett J A, Skene A M.
Assessment of fitness for surgical procedures and the variability of anaesthetists' judgments.
Br Med J 1980; 280: 509-512.

127. Lord F M, Novick M R.
Statistical theories of mental test scores.
Reading, Massachusetts: Addison Wesley, 1968: 335-349.

128. Tate R F.
The theory of correlation between two continuous variables when one is dichotomized.
Biometrika 1955; 42: 205-216.

129. Diaconis P, Efron B.
Computer-intensive methods in statistics.
Scientific American 1983; 248: 96-108.

130. Efron B, Gong G.
A leisurely look at the bootstrap, the jackknife and cross-validation.
Am Statisticiar 1983; 37: 36-48.

131. Irwig L M, Groeneveld H T, Pretorius J P G, Hnizdo E.
Relative observer accuracy for dichotomized variables.
J Chronic Dis 1985; In press.

132. Barnow B S.
The use of proxy variables when one or two independent variables are measured with error.
Am Statistician 1976; 30: 119-121.

133. Schlesselman J J.
Planning a longitudinal study.
J Chronic Dis 1973; 26: 553-570.

134. Rosner B, Polk B F.
The implications of blood pressure variability for clinical and screening purposes.
J Chronic Dis 1979; 32: 451-461.

135. Donner A, Bull S.
The mean versus the minimum as a criterion for hypertension screening.
J Chronic Dis 1981; 34: 527-531.

136. Sharp J T, Oglesby P, McKean H, Best W R.
A longitudinal study of bronchitic symptoms and spirometry in a middle-aged, male, industrial population.
Am Rev Respir Dis 1975; 108: 1066-1077.

137. Bode F R, Dosman J, Martin R R, Macklem P T.
Reversibility of pulmonary function abnormalities in smokers. A prospective study of early diagnostic tests of small airways disease.
Am J Med 1975; 59: 43-52.

138. Fairshter R D, Wilson A F.
Relationship between sites of airflow limitation and severity of chronic airflow obstruction.
Am Rev Respir Dis 1981; 123: 3-7.

139. Koran L M.
The reliability of clinical methods, data and judgments.
N Engl J Med 1975; 293: 642-6 and 695-701.