Safety in Mines Research Advisory Committee

Final Report

Development of lung function reference tables suitable for use in the South African mining industry

Prof. Rodney Ehrlich
Prof. Neil White
Prof. Jonny Myers
Prof. Mary-Lou Thompson

Department of Public Health, University of Cape Town

Dr Gavin Churchyard

Aurum Health Research, Welkom

Dr David Barnes
Dr DB de Villiers

Anglogold Health Service, (Pty) Ltd, Carletonville

SIMHEALTH 610A

MAY 7, 2000

Executive summary

Because of the respiratory hazards of mining, lung function testing [specifically spirometry - the measurement of Forced Vital Capacity (FVC) and Forced Expiratory Volume in One Second (FEV1)] is used in the mining industry for baseline, medical surveillance and compensation purposes.

Reference equations are needed to enable practitioners to assess the normality or otherwise of measured values. The purpose of this desktop study was to review available reference equations in order to recommend those that would be suitable for use in the mining industry, both for cross-sectional and longitudinal measurement. An important question was whether specific equations were needed for black and white mineworkers, in view of consistently observed differences in average FVC and FEV1. Recommendations were also developed on implementation of spirometry, on quality control and on future research.

Variation of lung function by test technique, age, sex, body size, race/ethnicity/geographic origin, altitude, smoking, and disease states is well established. The international "best practice" recommendation regarding choice of lung function reference equations is that such equations should be derived from recently conducted studies in unexposed, non-smoking populations of similar origin to the participants in the lung function testing programme.

A number of South African studies were reviewed including a large recent but as yet unpublished study of goldminers by *Hnizdo et al.* Further analysis was required in two of the studies. Although the Hnizdo et al. study was not suitable as a reference equation because it contained a large proportion of dust exposed as well as smoking employees, it provided a database of lung function values of in-service mineworkers against which candidate reference equations could be tested for their validity in defining the statistically appropriate proportion of "abnormals".

For black in-service mineworkers, a study of non-smoking, healthy bank workers in Johannesburg by *Louw et al.* provided the best statistical fit as well as meeting the recommended criteria. For white in-service mineworkers the commonly used reference equations published by the European Community for Coal and Steel (ECCS) based on European populations best satisfied these criteria. Application of the ECCS equation to black mineworkers was found to produce a large proportion of "false positive" abnormal values, with potentially adverse effects in the form of unnecessary investigations and reduced job security.

It was considered useful to recommend reference values for women as well, although sensitivity analysis using current female employees was not possible. The only available study of black South African women, by *Mokoetle et al.* of university workers in Johannesburg, appeared a reasonable choice, while retaining the ECCS equation for white women.

Recommendation of the above equations for programming into spirometers does not prevent use of a single (sex-specific) equation for all mineworkers if a practitioner so chooses, as long as the differential performance characteristics of the equation selected are taken into account in making administrative or clinical decisions. Also, because of secular change, review of all of the equations every 15 years or so is advisable.

Some of the uncertainty about choice of equations for once off evaluation could be avoided by longitudinal monitoring, in which each employee is used as his own control in evaluating change in lung function and judging the impact of dust or other factors. With some qualification, cross-sectional equations can be used for longitudinal purposes. With current knowledge, the most effective method would be to generate percentile curves on which the individual's lung function is charted, so as to pick up early departures from the "expected path". The alternative is to use "expected annual loss" estimates as norms. Both of these are conceptually unsatisfactory because of failure to use all the information provided by previous lung function values for the individual and other members of the exposure cohort.

Recommendations for the development of "adaptive" (truly longitudinal) reference norms are made, which would require, as a start, studies of dose-response effects of dust on lung function, both cross-sectionally and prospectively, in a cohort of miners.

Important implementation recommendations include (1) the preferential use of the statistical definition of abnormal based on a given percentile or lower confidence limit (calculated using the standard deviation) rather than the typical fixed "percent of predicted", (2) the need for a Guidance Note on quality assurance in spirometry, and for quality assurance in the training of spirometrists, and (3) the need to collect adequate collateral information as part of lung function surveillance.

Acknowledgements

The authors are grateful to Dr Eva Hnizdo, National Institute for Occupational Safety and Health, Morgantown, U.S.A. for providing updated versions of the papers based on the Anglogold study, for being prepared to carry out additional analyses required for this report and for commenting on the final report.

Thanks are also due to Ms Magda de Beer of the City of Johannesburg, for providing the file of analysis print-outs from the Mokoetle et al study, and Dr Carl Lombard of the South African Medical Research Council for tracking down and providing the cleaned data files from the Louw et al. study. The experience of the present study, which required scrutiny of analyses completed a decade or more ago, confirms the value of archiving such research material for possible later use.

Valuable comments and opinions were received from a number of people, both those who attended the workshop on this subject in Johannesburg on December 1, 1999, and others who contributed individually. They do not of course share responsibility for the findings and recommendations of this report.

Table of contents

List of tables

List of figures (by place in text - actual figures appended at end for report)

Glossary

Lung function testing

The term lung function testing covers a range of tests designed to assess the function and integrity of different parts of the respiratory system. In routine occupational medicine practice, almost all testing is *spirometry*, the measurement of the mechanical ventilatory capacity of the lung.

Spirometry

The primary measurements made in spirometry are of Forced Vital Capacity (FVC) , the total volume the subject can exhale in a forced expiratory manoeuvre, and Forced Expiratory Volume in one Second (FEV1), the volume the subject can exhale in the first second of a forced expiratory manoeuvre.

The ratio of FEV1 to FVC expressed as a percentage is an important measure of the pattern of ventilatory function. Some studies include a separate reference equation for the FEV1/FVC ratio, while others do not. In the latter case, the predicted ratio is calculated from the predicted FEV1 and predicted FVC.

This project is concerned with the reference or population norms of FVC and FEV1 only, as these are found in almost all studies.

Most spirometers also measure Peak Expiratory Flow Rate and some variation on Mid-Expiratory Flow Rate. These measures are not dealt with in this report, other than briefly in section 6.37.

Lung function reference values, equations, tables

Reference values are derived from measuring lung function in a sample of people believed to represent a suitable population. The criteria for suitability are discussed in section 3.2.

The sample data are used to fit a statistical model where typically a linear relationship is assumed between mean FVC or FEV1 and age and height, the most important predictors at the individual level. Separate models are estimated for men and women.

The model is represented by an equation. These can then be used to generate a "reference" value given any individual's sex, height and age. This reference value is the mean ("expected" or "predicted") value for FVC or FEV1 of people of that sex, age and height in the reference population.

Equations are now programmed as software in spirometric equipment, allowing individual reference values to be printed for a given age, height, sex and in some cases "race", as well as the *percentage predicted* value (observed divided by reference value)

An array of reference values for different ages and heights could be derived as a *table*, but this is unnecessary given the availability of programmable spirometers.

In comparing *different* equations, it is useful to choose a standard person, e.g. a man 40 years of age and 1.7 m tall, to determine whether one equation generates reference values that are on average greater or less than another.

Variability about the prediction equation (see also Appendix 1).

The standard deviation (variability) of observations about the mean (which is estimated by the prediction equation) can also be estimated from the reference sample. It is the "spread" of values around the average or mean lung function at a given age and height. Used together with the prediction equation, this statistic enables one to calculate a lower limit of normal (for a specific age and height), i.e. lung function values below which a given percentage of the population lies.

One is often interested in the lung function value that cuts off the lowest 5% of the reference population (the 5th percentile), because this is typically taken as the (arbitrary) threshold between "normal" and "abnormal" (*American Thoracic Society 1991*). Assuming the data are normally distributed, this 5th percentile can be estimated from the lower one sided 95% confidence limit, i.e. 1.645 times the standard deviation (see Appendix 1). Given the estimated prediction equation and the associated standard deviation, it is a simple matter to programme this lower limit of normal.

The interpretation of individual lung function values falling below this lower confidence limit or 5th percentile is not a simple matter, however. Such interpretation may vary with the choice of population from which the reference equation is derived. For instance, the reference population may be pre-selected to exclude respiratory disease and smokers, or it may be mixed general population. In the occupational setting, the predictive value of falling below this threshold for job disability or susceptibility to future disease (e.g. as a result of dust exposure) may not be clearly known. By definition, however, whatever reference population is chosen, 5 percent of this population will be considered "statistically abnormal".

Cross-sectional versus longitudinal data

Lung function data that are collected in a survey with one measurement of each subject (or the average of a number of measurements over a short period) are called cross-sectional. Data that are collected by repeating the lung function test on the same individual over time, e.g. every year or every 5 years, are termed longitudinal.

1. Introduction

1.1. Evolution of the brief

The original brief stated that:

1. Lung function testing is used in the mining industry for baseline, medical surveillance and compensation purposes.

2. Reference tables are, however, derived from "Caucasian" populations, e.g. those on which the currently recommended European Community for Coal and Steel (ECCS) equations are based. Application of such tables to "non-Caucasians" may prejudice some employees.

   The brief requested:

   - A prospective study of "representative" and "appropriate" populations.

   - A standardised set of reference tables

After due consideration, a proposal was submitted which did not include a new field study, for the following reasons:

1. There exists no occupational population in South Africa which is the "same" as underground mineworkers except with no exposure to respiratory hazards. There are two reasons for this.

   - Mineworkers are a fitter subpopulation of the general population - one would therefore expect mineworkers, particularly, in the younger years, to have somewhat better lung function than a general population sample. This is commonly referred to as a *healthy worker effect.*

   - Blue collar worker populations doing strenuous work and which include a large proportion of workers from rural areas, such as in agriculture , construction or possibly manufacturing, are likely to be exposed to a variety of respiratory hazards. They therefore cannot be regarded as an unexposed control population.

2. There already exist a number of studies of Southern African workers, including studies of black workers, which may be suitable for use (*White et al. 1994*). To this list must be added a recent large study of white and black goldminers (*Hnizdo et al. 2000b).*

3. In addition, because of the potential implications of choice of reference values for access to employment, security of employment and access to compensation benefits, a critical review process involving all stakeholders was needed to contribute to the thinking on this subject.

In the course of the study, it has been possible to refine the brief further in the light of new information, and to add more specific primary outputs to the original one.

In refining the brief, the following questions emerged:

1. Should there be different tables for different groups of mineworkers or a single table for all?

2. Are cross-sectional tables suitable for longitudinal use ?

1.2 Primary outputs

1. A report on the suitability of available lung function reference tables for the various uses required in the South African mining industry.

2. Recommendations regarding lung function equations for once-off (cross-sectional) use.

3. Recommendations regarding lung function reference equations for longitudinal use.

4. Recommendations regarding implementation, with particular emphasis on quality assurance.

5. Recommendations regarding future research, particularly on the collection of longitudinal data. and on the need or otherwise of further field study to generate new lung function reference data.

1.3 Steps to accomplish enabling outputs (methodology)

1. Occupational health practitioners were consulted on current practice in the use of lung function testing in the South African mining industry, as well as on relevant written standards.

2. A literature review was conducted in order to:

    - Clarify the role of variables which may affect comparability of reference equations drawn from different populations, such as sex, age, “race”, socioeconomic status, smoking, dust exposure, the healthy worker effect, etc.

    - Identify requirements for “suitability”, in choosing a reference equation from one population to be applied to members of another.

    - Find relevant international but particularly South African cross-sectional studies, so as to identify potentially suitable reference equations.

    - Scrutinise international literature on longitudinal studies, including techniques for making longitudinal inferences from cross-sectional data.

3. A search was conducted for suitable unpublished databases to be reviewed for their suitability as reference equations.

4. A search was conducted for the original data analysis of two of the published South African studies because of missing information in the published reports. Analysis was re-run on one of the studies.

5. A “sensitivity analysis” of candidate equations against the unpublished database was carried out.

6. Other issues in the use of lung function testing in the mining industry which affect

interpretations of abnormality were addressed.

Two workshops were held as part of the process:

A stakeholder workshop, for the purpose of canvassing opinions and clarifying perceptions and disagreements, was held in Johannesburg on December 1, 1999, and the report incorporated in the Interim Report to SIMHEALTH in January 2000.

A technical workshop of the authors was held in Cape Town on March 1, 2000, in which the interim findings were critically examined, resulting in changes in some the key aspects of the report.

1.4 Ethical considerations

1. The questions entailed by the primary outputs are important, with potential impact in respect of access to employment in the industry, prevention of lung disease, and compensation access.

2. The questions are as yet unanswered .

3. No new data were collected. Where use was made of medical surveillance data, these were already in database form, with no new testing or identification of individuals required.

4. This report is to SIMRAC, a tripartite structure, after which it will be publicised to stakeholders.

   Stakeholders were identified as follows:

   - Safety in Mines Research Advisory Committee (SIMRAC) and its constituent committees.
   - Department of Minerals and Energy, including tripartite committees, particularly the Mine Occupational Health Advisory Committee (MOHAC).
   - The mining industry and mining unions.
   - Occupational medicine and specialist practitioners: Mine Medical Officers' Association, South African Society for Occupational Medicine. South African Pulmonology Society
   - Mining hygiene and safety professionals.
   - Compensation authorities: Medical Bureau for Occupational Diseases, Compensation Commissioner for Occupational Diseases (under Occupational Diseases in Mines and Works Act) and the Compensation Commissioner (under Compensation for Occupational Injuries and Diseases Act).
   - Academic and research organisations: National Centre for Occupational Health, Epidemiology Research Unit
   - Non Governmental Organisations: Industrial Health Research Group

5. Ethical approval was obtained from the Research Ethics Committee of the Health Sciences Faculty of the University of Cape Town, and from the Medical Research Ethics Committee of Anglogold Health Service.

1.5 Timetable

Event	Date
Contract start	October 15, 1999
Stakeholder workshop	December , 1999
Interim Report	January 15, 2000
Final Report	May 7, 2000
Launch	June 14, 2000

1.6 Structure of the report

Key concepts are defined in the glossary. The Introduction deals with the research brief and how it evolved in the light of initial consideration, emerging information and further analysis. The primary outputs are specified, together with the steps leading to the enabling outputs. As this was a desktop review study, with a limited amount of formal analysis, the research methodology is presented as part of these steps.

Section 2 reviews current practices regarding lung function testing in both medical surveillance and impairment determination in the mining industry. A critical commentary is included.

Section 3.1 reviews the sources of variation in lung function testing, as these all contribute to differences between populations. Section 3.2 summarises recommended criteria for choice of reference equations.

Section 4.1 reviews published South African studies potentially suitable for use in the mining industry, supplemented in section 4.2 by a summary of an as yet unpublished study of goldminers.

Section 5 discusses the value of longitudinal monitoring of lung function, and the methods and difficulties of deriving longitudinal reference values from cross-sectional data and longitudinal data respectively.

Sections 2 to 5 above constitute *primary output 1*.

Section 6 contains the recommendations forming *primary outputs 2 to 5,* viz. recommendations on cross-sectional reference equations, longitudinal reference values, implementation including quality control, and future research.

Appendices 1, 2 and 3 are technical notes on cross-sectional and longitudinal prediction equations, and appendices 4 and 5 are notes on quality control in lung function testing sites and in training of spirometrists respectively.

2. Lung function testing in the South African mining industry: current practice

Decisions regarding choice of lung function reference values need to be seen in the context of how such reference values will be used in the industry. It was thus considered necessary to review current practice regarding the use of lung function testing in medical surveillance and impairment assessment.

The new occupational health legislative framework has considerably expanded the potential use of lung function testing in the South African mining industry. For example, historically, lung function testing was not used as a selection criterion for fitness to work among black miners.

It was not the purpose of this report to review the full range of current practices with regard to lung function testing, which is likely to be wide, given that many of the regulations and guidelines regarding lung function testing are still in the development phase. This section thus summarises current *recommended* practice together with some elements of actual practice (Anglogold Health Service - personal communication).

2.1 Summary of current practice

Lung function testing is required where it is expected (and/or confirmed) that the dust exposure levels are > 50 percent of the Occupational Exposure Limit (OEL) (as described in the draft *Guideline for the Compilation of a Mandatory Code of Practice for an Occupational Health Programme. No 1. Personal Exposure to Airborne Pollutants, Department of Minerals and Energy, Draft Amendment, 3 April 2000*).

2.1.1 Occupational medical surveillance

The frequency of lung function testing is recommended in a Department of Minerals and Energy (DME) guideline (*RSA 2000a*):

- On initial examination.
- One year after initial examination as a confirmation of the initial baseline.
- During periodical examinations: every 3 years as per the guideline.
- At exit examination.

Although other tests in the forced expiratory loop may be available, the medical surveillance is a screening exercise and only the following are routinely used in deciding on lung function (RSA *2000b).*

- Forced Vital Capacity (FVC)
- Forced Expiratory Volume in One Second (FEV1)
- FEV1/FVC ratio expressed as a percentage.

A number of options are given regarding reference values:

For white miners, the European Community for Coal and Steel (ECCS) prediction equations (*Quanjer 1983*) are recommended:

For black miners, 3 options are given:

1. The ECCS equations as they stand.
2. The ECCS equations with a fixed percentage discount based on commonly observed population differences (discussed further below).
3. Louw et al. (*1996).*

The following are applied as "normality" thresholds to assist assessment of candidates for an initial certificate of fitness for work in dusty areas (*RSA 2000b*):

- FVC >80% of predicted.
- FEV1/FVC% >75%.
- FEV1 >80% of predicted (omitted from Guidance Note, presumably inadvertently).

A discount percentage [as recommended by the American Thoracic Society [*1991)*], is suggested for spirometric values of black employees or work applicants. A discount of minus 5 percent is applied (rather than the minus 12 percent recommended in the ATS guidelines) on the basis of a recent study of South African goldminers (*Hnizdo et al. 2000b*).

2.1.2 Compensation

Compensation for occupational lung disease in mineworkers is covered by the *Occupational Diseases in Mines and Works Act 78/1973) (ODMWA).*

The Medical Bureau for Occupational Diseases (MBOD) convenes a Certification Committee, whose decisions regarding compensability and degree of damage are conveyed to the compensation commissioner under ODMWA. Decisions are usually based on radiology, but for several years there has been a guideline (not published) referring to lung function impairment. This has recently been submitted to the Minister of Health as a draft Regulation and defines 1st and 2nd degree occupational lung disease as follows:

Table 2.1 Classification system of lung function impairment contributing to determination of compensation status under the Occupational Diseases and Mines and Works Act*

	FEV1 (percent predicted)	**FVC (percent predicted)**	**FEV1/FVC (%)**	**Percent impairment**	**Award****
Normal	>80%	>80%	>75%	<10%	Nil (no comp.)
Mildly Impaired	79-65%*	79-65%	74-65%	<10%	Nil (no comp.)
Moderately Impaired	65-52%	65-52%	65-55%	10-40%	First Degree (initial comp.)
Severely Impaired	<51%	<51%	<55%	>40%	Second Degree (maximum comp.)

**Source: Department of Health: Draft Regulation on Medical Examinations and Standards applicable in the Certification of Compensable Diseases (unpublished).*

** Other factors such as exposure history, clinical status and radiological findings are taken into account. Comp.: compensation

2.1.3 Action on basis of abnormal findings, including refusal of certificate of fitness

Lung function screening should be seen as a *part* of the medical surveillance programme. Decisions such as refusal of certificate of fitness should be made only after full clinical assessment. The inability of a certain percentage (approximately 2 to 5 percent) of employees to produce reproducible test results must be recognised. Inability to perform a reproducible test may in fact be a predictor of future accelerated lung function loss (*Becklake 1990).*

In cases of abnormal lung function findings among in-service workers, this should normally be referred to the occupational medical practitioner for assessment, or in some cases for specialist opinion and full diagnostic lung function testing. Where temporary reasons for abnormal test results are found, a test should be repeated after a suitable period. Consideration has to be given to the proposed specific occupation in terms of risk profile and inherent requirements of the job, in order to ensure that there will be no unfair discrimination against an employee / prospective employee or compromise with respect to health risk.

A statement of fitness or otherwise can be given to management. However, specific findings are part of the medical record and subject to confidentiality requirements.

2.1.4 Special considerations: Rescue Brigadesmen

At this stage, the minimum screening requirement for Rescue Brigadesman is a Peak Expiratory Flow Rate of > 600 litres/min. Otherwise the criteria for acceptability are as stated in section 2.1.1 above. After medical screening, these candidates are subjected to a heat tolerance test (one hour stepping in the climatic chamber as opposed to the normal heat tolerance screening of 30 minutes) and a rigorous work load test. Approximately 20 percent of those initially classified as medically fit to undergo training fail these physical tests.

2.2 Problems identified by occupational health practitioners in applying the above system.

2.2.1 Reference values

1. Which reference values should be used? This has practical implications as judgements regarding fitness will be influenced by which reference value is applied.

2. What, if any "ethnic" correction factor should be applied between 5 to 13 percent? The Guidance Note (*RSA 2000b*) suggests minus 5 percent. Is this appropriate?

3. What is the standard for a normal annual decline in FEV1 or FVC in longitudinal measurements? Currently (*RSA 2000b*):

    - "Normal" decline is accepted as 40 ml p.a. (i.e. 120 ml / 3 years).

    - The decline between the ages 20 to 40 years should not be more than 10 to15 percent.

- "Accelerated" decline is an annual average decline of $\geq$200 ml between tests.

4. What is the validity (sensitivity, specificity, and predictive values) of FEV1 and FVC for the various outcomes of interest, such as inability to do the job, increased susceptibility to the effects of underground dusts and gases, future clinically important lung disease, etc.

2.2.2 Quality problems in routine lung function testing

1. Large numbers of workers need to be screened, with considerable variation in the number of employees seen on any given day.

2. It may be difficult to motivate occupational health staff to maintain high technical standards in carrying out routine testing for long periods on a daily basis.

3. Quality control may be difficult. The standards attained in an occupational health centre cannot be expected to be those of a research laboratory.

2.3 Commentary on current practice

The use of routine lung function testing in the mining industry is a relatively new development. Spirometry serves a number of purposes in the occupational setting.

1. Current fitness: Lung functions that are within the range of normal for a given person contribute to a judgement of fitness to perform work requiring respiratory effort.

   It needs to be recognised that FEV1 and FVC are only crude indicators of respiratory fitness. Such fitness depends on other physiological capacities as well as the individual's subjective tolerance of the physiological changes associated with exertional stress.

   Administrative practicality is the main reason for reliance on spirometry rather than on more complex tests of exertional capacity or respiratory reserve such as maximum oxygen uptake during exercise.

2. Future fitness. Current lung function may be regarded to some extent as a predictor of the future trajectory of lung function loss.

   Studies of chronic obstructive pulmonary disease (COPD), particularly that due to smoking, have confirmed the predictive value of an abnormal FEV1 for future lung function loss and clinically significant lung disease (and in fact for other diseases such as stroke and heart disease) (*The National Lung Health Education Program Executive Committee 1998*). However, because of the lack of lung function testing in the industry in the past, there is uncertainty about whether mineworkers with a low FEV1 at baseline are at greater risk of lung function impairment due to dust exposure (or even of silicosis or tuberculosis). Such information is obviously central to the whole exercise of lung function surveillance.

3. Diagnosis. FVC and FEV1 are affected in a variety of disease states. The particular pattern of abnormality assists with diagnosis and treatment. A given pattern of functional disturbance is not necessarily specific for a single anatomical or pathologic diagnosis, however.

4. Impairment and disability. Defined departures from normality (i.e. from the person's expected level) is a measure of impairment or the degree of respiratory dysfunction. This is only crudely correlated with disability, which is the inability to perform a given occupation. However, spirometry is routinely used in the assessment of impairment for purposes of compensation and disability assessment.

These functions overlap and there is potential for confusion about the exact meaning of abnormal lung function values in the routine setting and about the objectives such of lung function testing.

The role of lung function testing is clearest where there is an abnormal chest film and/or other clinical correlation leading to a specific diagnosis, e.g. pneumoconiosis, or where one wants to rule out significant disease in the presence of symptoms. Here lung function is used to measure (or exclude) degree of pathology or impairment. Many true restrictive deficits (decline in FVC, rise in FEV1/FVC ratio) are likely to be found in this context.

The more problematic role of spirometry is where it is used to *screen* for abnormality or disease in populations with a relatively low prevalence (pre-test probability) of lung disease, as would be expected in a typical working population of mineworkers.

Although the term medical *surveillance* is used in occupational health, lung function testing in these contexts is closer to *screening* in the public health sense. The purpose of screening is to identify a disease at an earlier stage than would be the case without screening, and where early identification confers individual advantage - whether in the form of advice, redeployment, medical treatment, etc. Surveillance in the public health sense is the recording of disease events to identify trends, gain new knowledge, etc.

In this report the familiar term medical surveillance will be used to cover both functions.

At entry, a baseline against which to measure any future decline in lung function is essential. However, the meaning of an isolated abnormal lung function value (particularly if in the mild to moderate range) may create some difficulties of interpretation as to the extent to which it reflects current fitness for strenuous work and/or susceptibility to future accelerated lung function loss due to dust.

In periodic examinations, airflow limitation, in the sense of early COPD or asthma is one of the important targets of surveillance. In particular, it is the impact on lung function of dust (or of a combination of underground dust, gases, or diesel exhaust in some mines) that underlies the need for such surveillance. On current knowledge, decline in FEV1 in goldminers is a function of cumulative exposure to dust , smoking, or most likely, an interaction between the two (*Becklake et al. 1987*, *Hnizdo et al. 1992* , *Cowie et al. 1991)***.** Tuberculosis and silicosis are also independently associated with decline in FEV1 (*Cowie et al. 1991, Hnizdo et al. 2000a*)

[There is more general evidence that occupational exposure to dust (and in some context a combination of dust and gases) causes a decline in FEV1 (*Becklake 1989*). This has been shown in South Africa in the recent Demographic and Health survey, in which respondents who reported being exposed to dusts or gases at work reported almost three times the prevalence of symptoms of airflow limitation than respondents not exposed (*RSA 2000*)].

Finally it should be noted that the OELs on which the 50 percent criterion (see 2.1 above) which triggers lung function surveillance is based are for specific mineral dusts, e.g. silica or coal, rather than mixed dust, and are directed at limitation of pneumoconiosis risk rather than that of lung function loss or COPD. Although the rationale or otherwise of this criterion is an important subject for debate, further discussion is beyond the scope of this project.

3. CONSIDERATIONS IN COMPARING LUNG FUNCTION ACROSS POPULATIONS, AND CHOICE OF REFERENCE POPULATIONS

3.1 Variation in lung function

Choice of a suitable reference population requires consideration of why populations differ in their lung function. In order to understand this, it is necessary to examine why lung function varies in the same individual over time (*intra*-individual variation) and between individuals at the same time (*inter*-individual variation). In addition, there are additional sources of variation which arise only at the population level. These factors have been comprehensively reviewed by Becklake et al. (*1993*), and are summarised below.

In considering variation in lung function measurement, Becklake et al. make the useful distinction between the *signal,* which is the variation due to the factor one is interested in (in this context, dust impact) and *noise*, which is variation due to all other factors.

3.1.1. Intra-individual variation

The signal here is the true biological loss attributable to dust in any individual between repeat measurements.

Other intra-individual factors (noise) include everything affecting the way the test is done, including subject understanding, the tester, the equipment, and also a small degree of normal biological variation such as time of day and season.

Of relevance to this study is the observation that workers may show improved lung function in periodic examinations compared to the entry examination. This is a *learning effect*. It has relevance both for cross-sectional studies of subjects who have not undergone the learning process and for establishing a baseline for longitudinal monitoring - the impact would be to underestimate lung function at the first examination.

With good quality control, the total contribution of intra-individual factors to inter-individual variation should be relatively small.

Regarding subject posture during the test, current practices in spirometry include sitting and standing. Sitting is recommended. Standing gives a slightly higher value (*ATS 1991*).

3.1.2. Inter-individual variation

The signal here is the effect of dust on lung function in a comparison of a high dust with a low dust exposed group (or an exposed with an unexposed group). (See section 2.3 above).

All the intra-individual factors mentioned in section 3.1.1 could contribute to inter-individual variation in lung function. In addition, there are the following:

3.1.2.1 Sex

Men and women of the same height and age differ in their FVC and FEV1, with men consistently having higher values on average than women, implying a fixed biological effect. Sex accounts for up to 30% of inter-individual variation (*Becklake et al. 1993*). Reference equations are thus derived separately for men and women.

3.1.2.2 Body size

Body size and possibly shape is the other fundamental determinant of FVC and FEV1. Of the various ways of representing this dimension, standing height has been found to be the best predictor (explaining up to 20 percent of variation). Sitting height has also been used in reference equations, with the advantage that it accounts to some extent for the variation in leg to torso proportions across groups of different geographic ancestry.

The influence of weight, other than at the extremes, is small relative to that of height (*Becklake et al. 1993*). However, weight gain may contribute significantly to FVC variation intra-individually over time (*Wang et al. 1996).*

In calculating the reference value of an individual subject, standing height is the most commonly used size variable. However, where a reference equation is based on some other size variable, e.g. sitting height or the ratio of sitting to standing height, then this is the form in which the variable must be measured and entered during testing.

3.1.2.3 Age

Age is another important determinant of lung function, which is typically displayed graphically as a function of age. FVC and FEV1 show growth through childhood and adolescence, reaching a plateau during young adulthood, and declining slowly thereafter. Lung function reference equations always account for age. However, where equations are derived from a sample containing relatively few subjects of a certain age, typically the very young and subjects over 55 or 60 years, the reference values for those ages derived from that equation will be subject to greater uncertainty than for the ages for which there are many subjects.

3.1.2.4 Geographic ancestry ("race")[1]

Many studies have observed variation of FVC and FEV1 across populations defined racially or ethnically, both in Southern Africa, (*De Kock et al. 1988, Louw et al. 1996, Hnizdo 2000b)* and in many other settings (e.g. *Schoenberg et al. 1978*, *Yang et al. 1991 , Hankinson et al. 1999; also reviewed in White et al. 1994).*

The general finding of the literature is that, particularly when studied by the same methods, populations of European descent typically have higher height and age adjusted FVC and FEV1 than populations of Asian (*Yang et al. 1991*) or African descent (*White et al .1994*). Within Asia, differences have been observed between populations of Chinese descent and those from the subcontinent of India (*Yang et al. 1991*). In the United States, a difference of 10 to 15 percent has been observed between white and black Americans, leading to the recommendation of either racially/ethnically specific equations or a scaling factor applied to the sample drawn from a white population (*American Thoracic Society 1991, British Thoracic Society 1994*).

The source of these differences has been controversial (*Myers 1984*). Factors cited as explaining these differences vary. Biological explanations view chest size (at least as reflected in FVC) as a dimension of heritable body size or proportion which varies across the human species by geographic ancestry (*Molnar 1998).* Some of the difference may be explicable by variation in other dimensions of body size such as trunk/leg ratio, but differences persist after adjustment for dimensions of frame size (*Jacobs et al .1992).*

Environmental explanations favour the cumulative impact of poverty and racial discrimination, operating, for example, through undernutrition in early life (including *in utero*) or through recurrent respiratory infections (Myers 1984). For example, Barker has shown an association between birth weight and childhood respiratory infections and decreased FEV1 in adult life (*Barker et al. 1991*).

---

[1] "Race" is the generic term most commonly used in the four-category classification of South Africans entrenched by apartheid into whites, blacks (alternatively Africans), coloureds and Asians (alternatively Indians). Race in this sense, while ostensibly referring to skin colour and ancestry, is closely correlated with socioeconomic circumstances and a host of environmental differences. There is also considerable variability within each category. Despite its everyday "obvious" use, race has no precise or agreed scientific meaning (*Diamond 1994*). The notion of the human species as divided into a relatively small number of biologically distinct races has been scientifically discredited (*Hoffman 1994*). Most genetic variability is between individuals, with a relatively small amount (estimated at 7%) associated with the broad groupings commonly called "races" (*Hoffman 1994, National Institutes of Health 1998*). "Ethnicity" is commonly used as a substitute for "race", but is probably less suitable in this context as it includes a mix of "racial", geographic, cultural, religious and language characteristics that cuts across " race" in the South African sense (*Senior et al. 1994).*

The problem is how to categorise measures of biological variation (such as skin colour, body shape and possibly chest size) that may have arisen in populations across geographical gradients (both within continents and across continents). Such classifications always contain an arbitrary element strongly influenced by current historical usage. For the purposes of this report, the reference is to the two groups that make up the vast majority of mineworkers likely to come to lung function testing: migrant workers from various part of South Africa, Lesotho and Mozambique whose ancestry is traceable to Bantu speaking peoples who migrated from central to Southern Africa in past millenia; and those employees, generally of higher occupational and educational status, whose forebears migrated to Southern Africa from Europe after the start of the colonial era. These mineworkers are referred to as black and white respectively and "race" has a corresponding meaning. To quote Becklacke et al. (*1993) "*In the context of spirometric lung function, …race should be considered a naming device, to be used for studying human diversity*".* When reference is made to published studies, the terminology used by the authors of such studies has been retained.

Some environmental influence is given credence by limited evidence of apparent improvement in lung function in migrant populations (*Massey et al. 1986*), and over cross-sectional studies carried out across 40 years (with one study from the 19th century (*Glindmeyer et al. 1982, White et al. 1994*). However, the matter remains scientifically unresolved.

It should be noted that the differences observed for FEV1 and FVC do not necessarily apply to other indices of lung function measurement. Differences tend to be in the opposite direction or insignificant for the FEV1/FVC ratio (*Hankinson al. 1999*), with no difference in expiratory flow rates derived from the forced expiratory curve (*De Kock et al. 1988*).

3.1.2.5 Socioeconomic status

In contrast to the literature on lung function variation by race, t/here has been relatively little scientific investigation of variation by socioeconomic status. A review by Steinberg et al. (*1986*) concluded that lower socioeconomic status was associated with lower lung function in adults, independently of smoking.

The close correlation of socioeconomic status and “race” in South Africa makes it difficult to separate these two factors. For example, in the study by Goldin et al. (*1996)* of white collar bank and brewery workers in Johannesburg, the typical lung function differences found between black and white male employees could equally well be explained by a composite index of socieoeconomic factors without including “race” in the regression equation. These factors included income, occupation, number of dependants, home fuel type and home ownership.

In the recent national Demographic and Health survey in South Africa (*RSA 1998, RSA 2000*), prevalence of respiratory symptoms and of abnormal peak flow rates were very strongly negatively correlated with education level. Men with no education had over 6 times the prevalence of symptoms of airflow limitation and 5 times the prevalence of abnormal peak flow than men with greater than Std 10 education.

Further research is needed on the association of socioeconomic status with lung function and on the mechanisms for such associations. The results of such research may reduce some of the uncertainty in the interpretation of observed differences by “race”.

3.1.2.6 Exposures

Tobacco smoke

The impact of tobacco smoke in reducing lung function, particularly FEV1, is well known. Among smokers there may be a susceptible group who show accelerated lung function loss with age (*Fletcher 1977*).

Home and community air pollution

Indoor air pollution, particularly that from coal and wood burning as occurs in large parts of South Africa, has a significant impact on respiratory symptoms (*Terblanche et al. 1993*). The effects on lung function have not been measured but may be significant. In the Demographic and Health Survey (*RSA 2000*), exposure to indoor cooking fuels was only weakly associated with symptoms of airflow limitation or abnormal peak flow rates in adults. The association with symptoms of chronic bronchitis (productive cough) was somewhat greater.

3.1.2.7 Current and past lung disease

Different disease states lead to different profiles of lung function deficit. Of relevance to goldmining is that both simple silicosis (*Cowie et al. 1991*) and pulmonary tuberculosis (*Hnizdo et al. 2000b*) are strongly associated with lung function loss.

3.1.2.8 Altitude

Populations living at greater altitude have been shown to have higher lung function values than coastal populations. The effect on FVC in men has been estimated across populations of Sub-Saharan ancestry to be of the order of 250 ml per 1000 m of ascent (*White et al.1994*). Similar values have been shown elsewhere, including on the Indian subcontinent (*Cotes et al. 1975*).

3.1.3 Inter-population variation

Any or all of the above factors could cause differences between populations in lung function.

In addition, there are s*election fac*tors operating at the population level which may cause observed lung function differences.

For example, worker populations may differ from the general population by containing a higher proportion of "healthy" subjects. In particular, mineworkers can be expected to be a "fitter" population than both the general population and worker populations in less demanding industries, at least in the early years. This is because those with respiratory problems may avoid the mine work or be selected out at entry examinations. Further, those who experience health problems early on in service may leave. There is some evidence for this effect (*Hnizdo . 1992*).

In the longer term, the impact of work exposures on health as well as of any monitoring or medical removal policies will influence the lung function of remaining long-service workers (producing a "survivor" effect). The net effect on the remaining worker population compared to the general population is unpredictable.

Analogous to the healthy worker effect on lung function in goldmining, the association between smoking and lung function may be confounded by self-selection of less susceptible people into the smoking cohort. At younger ages, smokers may thus have better lung function than non-smokers, although their lung function may decline at faster rate with age as a result of smoking. The implication is that if one chooses a non-smoking population as a reference for a population which includes smokers, the former may understate “true predicted” lung function at all ages. Further, that a population with a mixture of smokers and nonsmokers may be a better reference population than one in which smokers have been removed.

Secular shifts in lung function across generations (cohort effects) can also be regarded as inter-population variation.

3.2 Selection of reference values for lung function testing

The subject of reference values for lung function has been comprehensively dealt with in the medical literature. Bodies such as the ATS (American Thoracic Society), ERS (European Respiratory Society) (*Quanjer 1993*) , and BTS (British Thoracic Society) have all published comprehensive technical documents relating to lung function testing, reference values and definitions of abnormality.

The ATS, in particular, has dealt with policy questions related to selecting reference values. It is useful to review the various considerations in selecting reference values that appear in the ATS documents (*ATS 1991, 1995*).

Selection of reference values requires careful consideration since no one set of reference values is likely to be applicable to all laboratories and all clientele. Unless there are specific policy recommendations, it is likely that by default the matter will be left to the judgement of manufacturers of automated equipment.

3.2.1 Epidemiological considerations

Reference values should be selected from cross-sectional studies of populations who are free of respiratory symptoms and disease. It is preferable to choose reference values for men and women from the same population source. Altitude is an important consideration.

The ATS uses the term ethnicity to refer to populations of different ancestries. The ATS view is that it is preferable to use equations based on populations with the same ethnic origins as the clientele.

An alternative strategy recommended is to use scaling factors (fixed proportions) to adjust for ethnic differences. However, in reality this proportion may vary by gender, age and height, and between FEV1 and FVC (*Hankinson et al. 1996*). The practice of using a single scaling factor regardless of age, height, gender or lung function index is thus clearly not optimal.

It is also preferable to select reference value studies based on populations in the country or region where they will be used. Furthermore, because of the variation between populations and the well documented healthy worker effect, the reference population should be employment based rather than general population based.

To assist in the choice of reference values, it may be useful to make an empirical assessment of how reference equations relate to subjects typical of the anticipated clientele. To be more rigorous it is desirable to conduct a formal *sensitivity analysis* of reference equations to evaluate whether they assign an appropriate proportion of statistically abnormal results when used in a population typical of the clientele (*Leibowitz et al. 1990*).

Finally, it is important that any reference values should be derived from recent studies, i.e. from within the last 10 years, given the possible secular trends in population lung function.

3.2.2 Statistical considerations

Reference values are typically based on linear prediction equations that include age and height as independent variables. Separate equations are used for men and women. The age range of the reference population studied must be known since, in general, reference values should not be extrapolated beyond this age range. Linear equations may predict values that are too high for young adults and too low for the elderly.

For lower limits of normal, it has become convention that normal ranges be based on calculated lower fifth percentiles. Estimates based on 1.64 x SD (see glossary and Appendix 1) are acceptable for FEV1 and FVC. These lower limits of normal should be considered as arbitrary limits that do not necessarily correctly classify all clients into normal and abnormal groups. Indeed, by definition 5% of the normal reference population are classified as “abnormal”. In particular, values for subjects being screened using this cutoff that lie close to the lower limit of normal should be interpreted with caution.

The use of fixed percentage thresholds such as “80 % of predicted” as a lower limit of normal is not recommended. This is a non-statistical convention and is unlikely to hold across different age groups. Its use at older ages may result in overestimation of reference values, i.e. excessive “false positives”.

It is also not recommended that a fixed value for the FEV1/FVC ratio be used as a lower limit of normal. This too is likely to decline with age (*Hankinson et al. 1999*).

4. SOUTH AFRICAN REFERENCE EQUATIONS: PUBLISHED AND UNPUBLISHED DATA

4.1 Published studies

Reference values for lung function testing in the South African mining industry should be based on a consideration of the epidemiological, statistical and other relevant issues described in the previous sections.

Ideal reference values for the mining industry would be based on a recent, technically well conducted, cross-sectional study of non-smoking unexposed healthy workers who live at highveld altitudes and include a good proportion of older workers. It should include men and women, as there is no longer any statutory bar to women working in the industry. It should be possible to calculate any desired statistical lower limit of normal from the study by use of the standard deviation of the values around the prediction equation.

Although there are a number of candidate South African studies that satisfy some of the above considerations, there is no one study that satisfies all of them. However, based on consideration of original study data, sensitivity analysis, or a combination of these techniques, it should still be feasible to recommend existing reference equations for use in the South African mining industry.

Eight published reference equations for lung function were identified in a recent review of lung function reported in studies of healthy black South Africans (*White et al. 1994*).

Table 4.1 : Comparison of the Predicted Value for FEV1 and FVC (age 38 years, height 171 cm) from eight studies of spirometry in black workers conducted in South Africa (in ascending order of predicted FVC)

First Author (Year published)	**Sample size**	**Sample source**	**Predicted FEV1**	**Predicted FVC**
Yach (1985)	a) 582 b) 183	a)Grain mill b)Controls	2.49 2.63	3.74 3.74
Myers (1985)	203	Stevedores		4.00
Johannsen (1968)	120	Hospital workers	2.98	4.06
White (1989)	109	Textile workers	3.46	4.12
Goldin (1989)	128	Bank workers	3.44	4.21
Hessel (1989)	653	Vermiculite miners	3.29	4.23
Coetzee (1989)	518	Asbestos workers	3.55	4.30
Mokoetle (1990)	206	University workers	3.47	4.42

* Source: *White et al. 1994*

One further, as yet unpublished, study specifically of gold mining employees is detailed separately below (*Hnizdo 2000b*).

It is of interest to note that a number of the above studies, particularly the more recent ones, have produced generally similar results for standardised predicted values of FEV1 and FVC.

However, based on the considerations recommended by the ATS above, there are two published studies that emerge as serious contenders for reference values for lung function in the mining industry: those of *Louw et al. 1996* and *Mokoetle et al. 1994*.

Both of these studies are cross-sectional studies of South African workers in Johannesburg. Both cover the age range 20 to 60/65 years, although not beyond this.

The study by Mokoetle (*1994)* was of black university employees, the majority of whom were classified as being in semi-skilled or skilled occupations, with a small proportion of professionals (6.8% among men, 16.1% among women). Of note is that of the studies cited in Table 4.1, this study found the highest standardised predicted FVC (although not so for FEV1).

The Mokoetle et al. study has the advantage that it included both men and women, but the potential disadvantage that the full sample included smokers and symptomatic respondents. Given that it is the only study thus far of black South African women (with a sample size of 203), it is a candidate for providing a reference equation for women in the mining industry.

Although results for asymptomatic, non-smoking women were available, the small size of this stratified sample require that the whole female sample, i.e. including smokers (defined as “ever” versus “never) and symptomatic subjects, be used. However, there was no adverse smoking effect demonstrated by Mokoetle et al, probably because of the “healthy smoker” effect (never smokers in fact had a lower standardised mean FVC than the full sample) and the relatively low smoking rates only (14 percent of the women smoked).

The study by Louw et al*. (1996*) had the advantage of having sufficient numbers to exclude smokers and symptomatic persons and a population unexposed to respiratory hazards. The authors studied a large sample of both black and white employees drawn from the staff of a large banking group in Johannesburg, although the final sample size was modest due to strict exclusion criteria. Only men were studied. The proportions of black subjects below 25 years and above 55 years of age were also relatively small.

Neither Mokoetle et al. nor Louw et al. provided the standard deviation (SD) for their reference equations of lung function in relation to age and standing height in their published reports. Louw et al did provide a SD for an equation based on both sitting height and the sitting to standing height ratio. (Because of the possible administrative burden of having to measure both sitting and standing height in routine lung function testing, this was regarded by the present authors as an obstacle to routine use of that reference equation).

In the case of the *Mokoetle et al.* study the original summaries of the analysis were found, providing the necessary information on the SDs.

In the case of the *Louw et al*. study, the original data were obtained. These had already been cleaned and were in analysable format. The SDs were obtained by re-running the regression model for age and standing height to obtain a model equivalent to that reported in the published study. Their Vitalograph data were used rather than those derived from the

Autolink spirometer.[2]

4.2 Unpublished study (*Hnizdo et al . 2000b)*

Although not yet published, this database was considered important because of a number of factors.

1. It contains large numbers of subjects, making it one of the largest such studies ever.
2. The population is the same or similar population to the one to which the reference equations have to be applied, viz. goldminers in service.
3. It includes black and white miners from Southern Africa.
4. It provides data against which the suitability of other candidate reference equations can be judged, by comparing the proportion of these workers categorised as statistically abnormal (below a given lower limit of normal) by each of the candidate reference equations.

4.2.1 Methods

Data from a lung function screening programme conducted at a large South African gold mine from 1994 to 1998 were used to estimate the prediction equations.

Lung function tests from the most reliable period of testing (coefficient of reliability G=0.93) were used for this study (*Hnizdo et al.1999*). Values from periodic testing rather than from the baseline examination were used, to allow for the learning effect.

Miners with a history of pulmonary tuberculosis or with radiological abnormalities were excluded.

The prediction equations were estimated cross-sectionally on 15 772 black and 2 752 white miners and compared with published reference equations.

4.2.2 Results

The estimated prediction equations are provided in Table 4.2.2.

---

[2] Louw and Goldin observed a systematic difference between the lung function values measured on the Autolink (sleeve sealed piston) and Vitalograph (wedge-bellow) spirometers. It is known that different instruments used to test the same people may produce systematic differences despite the production of precise results on each. It is thus impossible to know which of the two instruments is the more accurate. The equations derived from the Vitalograph were used in the report because the Vitalograph was used in four of the eight studies listed in Table 4.1 (*Myers et al., White et al., Coetzee et a.l and Louw et al.*) with comparable results.

Table 4.2.2 Prediction equations derived by Hnizdo et al. (2000b)

	Height (m)	**Age (yrs)**	**Intercept**	**SD**
White miners				
FEV1	4.314	-0.038	-2.341	0.484
FVC	5.940	-0.036	-4.407	0.547
Black miners				
FEV1	3.665	-0.30	-1.654	0.477
FVC	4.655	-0.025	-2.901	0.528

There were significant differences between the white and black miners (higher values in whites) in the height and age adjusted mean values (standardised to 40 year of age, 1.7 m in height):

- FVC: 220 ml (5.2 percent mean difference).
- FEV1: 110 ml (3.2 percent mean difference).

These mean differences are smaller than the “ethnic” differences typically cited in the literature (*ATS 1991*). However, these were dust exposed and smoking populations. The heavier smoking histories among white than black mineworkers, in particular, might have reduced any baseline difference.

It should be noted that the linear equations did not provide a good fit for the 20-29 years and 55 years or older age categories.

Also, a strong cohort effect on height was observed. Men 50 years of age were 3.6 cm shorter than those 20 year of age. This difference is greater than would be expected from a simple aging effect, suggesting that succeeding generations of mineworkers are getting taller. As a result, the linear regression statistics for age obtained from these data may provide a biased estimate for loss of lung function with age.

4.3 Sensitivity analysis of other reference equations against the *Hnizdo et al.* data

Sensitivity analysis in this context is defined as the derivation of a lower limit of normal (statistical abnormality threshold) from the reference equation of interest and applying it to a distribution of lung function tests *actually observed* in the population in which the equation would be used. In this case, the observed values are derived from in-service goldminers without TB or pneumoconiosis.

This enables one to evaluate what proportion of this workforce would be classified as statistically abnormal using the prediction equation of interest. In this way the practical implications of the use of a specific reference equation in different contexts, i.e. in pre-assignment/ pre-employment, periodic medical examination or in impairment/ compensation evaluations, can be assessed.

“Best fit” was assessed visually by Hnizdo et al. by comparing the height standardised regression lines of lung function against age. The equations reported by Knudson et al. (*1976*) and those of the ECCS report (*Quanjer et al. 1983)* provided the closest fit for the white miners.

This was confirmed by observing that the ECCS equation provided lower limits of normal (i.e., one-sided 95% lower confidence limits) that were closest to the *observed* one-sided lower 95% confidence limit, for the age stratum 30 to 34 years (Table 4.3b). This was not the case for the age stratum 50 to 54 years, however, for which the ECCS equation cuts off larger percentages than in the younger stratum. This is most likely due to the cumulative impact of high smoking levels and their interaction with dust among white miners.

For black miners, the reference equations derived for black Johannesburg bank employees by Louw et al. (*1996*) provided the best fit. This is confirmed in Table 4.3a which indicates that the lower limit of normal derived from Louw et al. cuts off proportions that are reasonably close to 5 percent for both FEV1 and FVC for both age strata.

The Mokoetle et al. equation works less well in this respect, with a large discrepancy between the proportion cut off for FEV1 (low) and that for FVC (above 10 percent).

The ECCS equation cuts off high proportions (greater than 10 percent) for both FEV1 and FVC among black miners. Hnizdo et al. noted that when the ECCS reference equations were scaled by a conversion factor of 0.92 for the predicted FVC and 0.95 for the predicted FEV1, the resulting line fitted well.

To round off the comparison, the equations derived from Louw et al. and Mokoetle et al. for black university and bank employees respectively were applied to the observed values for white miners (Table 4.3b). As expected, in the younger age group, these equations cut off relatively low proportions of white miners. However, in the 50 to 54 year stratum the proportions rose well above 5 percent (13.6 percent for FVC predicted by Mokoetle et al.), consistent with more rapid decline in lung function among white miners than in the reference populations.

TABLE 4.3a Proportion of population of in-service black gold miners classified as "abnormal" using lower limits of normal from different equations: (2 age strata; height 1,7m.)*

	30-34 years		**50 –54 years**	
	FEV1	**FVC**	**FEV1**	**FVC**
Mean	3,60	4,21	3.06	3.74
SD	0,46	0,51	0,46	0,51
HNIZDO (black miners)				
LLL (litres)	2,81	3,34	2,24	2,86
% abnormal	*4.3%*	*4,5%*	*3,9%*	*4,2%*
LOUW (black workers)				
LLL	*2,78*	*3,44*	*2,26*	*2,98*
% abnormal	*3,8%*	*6,6%*	*4,3%*	*6,8%*
MOKOETLE				
LLL (litres)	2,78	3,58	2,10	3,19
% abnormal	*3,8%*	*11,1%*	*2,0%*	*12,1%*
ECCS (no scaling factor)				
LLL (litres)	3,07	3,68	2,47	3,07
% abnormal	*12,5%*	*15,2%*	*10,4%*	*9,3%*

*Percentage classified abnormal calculated from:
(Observed mean - reference lower limit of normal) / observed SD

*TABLE 4.3b Proportion of population of in-service white gold miners classified as abnormal using lower limits of normal from different equations: (2 age strata; height 1,7 m).**

	30-34 years		**50-54 years**	
	FEV1	**FVC**	**FEV1**	**FVC**
Mean	3,74	4,54	3,04	3,83
SD	0.45	0.53	0,56	0,58
HNIZDO (white miners)				
LLL (litres)	2.96	3.62	2,23	2,93
% abnormal	*4.2%*	*4.1%*	*7,6%*	*6,2%*
LOUW (black workers)				
LLL	2,78	3,44	2,26	2,98
% abnormal	*1,6%*	*1,9%*	*8,2%*	*7,4%*
MOKOETLE				
LLL (litres)	2,78	3,58	2,10	3,19
% abnormal	*1,6%*	*3,6%*	*4,8%*	*13,4%*
ECCS				
LLL (litres)	3,06	3,68	2,47	3,07
% abnormal	*6,4%*	*5,4%*	*15,6%*	*9,5%*

*Percentage classified abnormal calculated from:
(Observed mean - reference lower limit of normal) / observed SD

The Hnizdo et al. prediction equations although derived from the source population of interest has a number of disadvantages regarding their use as reference equations

The study did not exclude or measure the effect of smoking nor of occupational exposures, e.g. specific occupation, concentrations of respirable silica, or duration of employment/exposure.

Study lung functions were collected under a routine surveillance system, and not specifically as part of a quality controlled study. However, the reliability coefficient was 93 percent, suggesting a high degree of quality control.

There were also selection processes affecting the workforce studied. Selection into the industry at pre-employment examination would elevate FVC and FEV1 relative to the general population. Additionally at measurement those with pneumoconiosis and TB were excluded, so that the remaining sample represents a survivor population who are also likely to have higher values than the general population. Over the long-term mining exposures and smoking would reduce FEV1, although not necessarily the FVC. Hence relative to an unexposed population FVC may be higher, while FEV1 could be either higher or lower in longer service miners than in the general population.

Consistent with the above argument is the observation that the Hnizdo et al. manuscript (*2000b: Figure 2*) in fact shows that the median FEV1 compared to that from Louw et al. is greater at younger ages, with the difference diminishing with age.

5. CONSIDERATIONS IN LONGITUDINAL MEASUREMENT OF LUNG FUNCTION

5.1 Potential uses and advantages of longitudinal measurements

There are 3 main categories of use in medical evaluation of mineworkers: at intake to employment, at periodic examinations, and for compensation purposes for mining-related occupational diseases.

1. There is a general question as to whether pre-employment examinations have any preventive value, and this has not yet been settled for any industry, particularly not South African industry and especially not for mining. However, if measurement of lung function at entry to employment is viewed as part of a longitudinal monitoring programme, it will increase the effort put into ensuring the quality of the baseline measurement, which has to be accurate for meaningful monitoring over time.

   Baseline testing should be followed by a surveillance process which is able to successfully flag and follow up an abnormal initial value. One or more (cross-sectional) reference ranges could be used to interpret the initial value as part of the decision as to whether to recommend the person as fit for the relevant category of work.

2. In periodic examinations, longitudinal measurements measuring a change from baseline potentially allow removal of known constant confounders or effect modifiers such as "race", and allow adjustment or control for known determinants of lung function (e.g. age, smoking, past and present tuberculosis or other respiratory disease) in assessing the impact of dust.

3. For compensation assessment, longitudinal measurements maximise information about the true determinants of loss of function over time, and potentially enable the separation of work-related and other determinants.

The value of longitudinal monitoring is illustrated in the following example.

See as attachment *Figure 5.1 Examples of longitudinal changes in FEV1 for three different persons (subjects A, B and C)*

Figure 5.1 from Hankinson and Wagner (*1993*) plots percent predicted FEV1 against age and shows different tracks across time for three subjects. This illustrates three possible scenarios based on longitudinal measurements.

Subject A is normal for age at all ages compared with reference values.

Subject B is shown to be a problem case. Despite beginning at 100 percent function there is a rapid decline, which might be preventable, to very low levels.

Subject C begins below 100 percent function but maintains this level throughout life. Subject C could just as well have begun at 75 percent of normal function and have been flagged as abnormal (thereby affecting employment chances) despite being "healthy".

The use of longitudinal measurements to distinguish such cases is thus potentially of great value.

5.2 How might reference values for longitudinal lung function be derived?

Reference ranges for lung function by age (for a given height) based on longitudinal follow-up data may differ from those based on cross-sectional data owing to a number of factors:

- Selection effects (as in the healthy worker /survivor effect)
- Cohort effects (e.g. secular changes in lung function due to improved socioeconomic conditions).
- Changes in height with age causing cross-sectional reference equations to be biased upwards at older ages relative to those based on longitudinal measurements.
- Changing risk profiles (e.g. improvement , or deterioration, in occupational risk factors and occupational hygiene).

Dockery et al. (*1993*) used both cross-sectional and 6-year longitudinal data in developing longitudinal reference values. Figure 5.2a derived from Dockery et al. shows percentile curves based on cross-sectional values for both FVC and FEV1 for a sample of black American men from six cities. These are percentile curves for pulmonary function (PF) values adjusted by height (PF/height^2).

See as attachment: ***Figure 5.2a. Percentiles of lung function (liters) divided by height (meters) squared (PF/ht^2)for asymptomatic, never smoking black males aged 25 to 74 years from six U.S. cities***

Studies based on longitudinal data show similar rates of decline to those based on cross-sectional data in the middle of the adult age range, but significant differences mainly for the elderly. Dockery et al. also followed a sample of white Americans in the six cities for 6 years at two time points separated by 3 years, plotted the mean change for the different age groups over this period and superimposed the points on the cross-sectional percentiles as shown in Figure 5.2b from Dockery et al. (1993).

See as attachment: ***Figure 5.2b. Comparison of estimated population mean track for FEV1/ht^2 determined from longitudinal, repeated measurements for asymptomatic never-smokers (solid bullets) and a smoker of one pack of cigarettes per day starting at age 25 years (open triangles)***

This figure shows that for FEV1 the percentile values are much the same (cross-sectional versus longitudinal) for the 50th percentile, except at ages over 60 or 65 years, when the percentiles based on longitudinal data seem to decline faster than those based on cross-sectional data. This discrepancy can be removed by using the height at the time of the *latest* measurement of lung function as opposed to the height measured initially. As people lose height when they get older, use of the lower (“older”) height value will result in the solid curve rising to meet the dotted curve in the figure.

A survivor effect might also have been expected to give rise to the difference observed between longitudinal and cross-sectional curves, but adjustment for height was sufficient to remove most of this difference.

An example of the smoking effect for an individual is also shown in Figure 5.2b to be quite dramatic and easily detectable, as percentiles are crossed over time.

The conclusion drawn by Dockery et al. was that differences between cross-sectional and longitudinal data in expected lung function up to the age of 65 years are very small for asymptomatic never smokers, and it is possible to rely on cross-sectional data to develop percentile charts.

However, given that they have provided comparison data for the 50th percentile only, the curves based on cross-sectional or longitudinally derived data still need to be examined at all percentiles – especially those at the extremes (5 percent or lower).

There are currently no longitudinal data in non-exposed South African populations. However, existing cross-sectional South African data could be used as if they were longitudinally derived to generate percentiles. The values at different ages obtained cross-sectionally could be used as if they were equivalent to longitudinal measurements on the same person across time.

Among South African studies, the closest cross-sectional equivalent to Dockery's data are those of Louw et al. (*1996*), although they did not study women. Hnizdo's data (*2000b*) could also be considered, although, unlike Louw et al., they included smokers and dust exposed workers.

5.3 HOW ARE LUNG FUNCTION SURVEILLANCE DATA TYPICALLY USED IN PRACTICE?

It is not usually recognised that lung function measurements over time, whether data are collected longitudinally or cross-sectionally, are typically *interpreted* cross-sectionally. This is usually done by examining the spirometer printout measured value against the predicted value. If the former is less than 80% predicted, the subject is flagged for further action, otherwise the test is considered normal. In this case all benefit from existing past measurements are lost. Worse still, as is shown in Appendix 3, the specificity (or the percentage of those testing normal who are truly normal) decreases with each successive cross-sectional test. This means that with each successive periodic lung function test the percentage of false positives (those testing abnormal who are in fact normal) increases leading to more work for the occupational health service staff.

Typically any comparison with past measurements is “informal” or qualitative, involving looking at differences from test to test (usually only the last test and the current test) and using some rule of thumb about what is an excessive loss.

An alternative, although yet not in routine use, is to plot observed values over time using percentile charts (see Figures 5.2a and 5.2b). However, this also involves what are essentially qualitative judgements about how many percentiles can be crossed before triggering action. The ideal comparison would enable the precise quantification of changes in an individual over time against reference ranges for expected or normal changeover time. These are genuine “adaptive” longitudinal reference ranges (Thompson and Fatti, 1997a; Thompson and Fatti, 1997b).

5.4 Adaptive reference ranges

A longitudinal sample of unexposed individuals could be used as the basis for developing estimates of percentiles for “normal” change over time (where “normal” change refers to age rather than exposure-related change).

One approach developed by Thompson et al. (*1997a, 1997b*) involves the development of personalised adaptive centiles for an individual, conditional on his/her previous path.

See as attachment: ***Figure 5.4 Illustration of what is being compared at two successive timepoints using cross-sectional or longitudinal reference data from an unexposed population***

The graph in the figure refers to an unexposed population measured at ages 25 and 35 years.

The adaptive unexposed curve **[_ _ _]** is the lung function distribution at age 35 years of unexposed individuals who have had a previous measurement taken at age 25 years (represented by the black star).

The factors that determine this distribution are the means and variances for lung function for the entire unexposed population at both ages (reflected in the two cross-sectional unexposed curves **[__ . . __ . .]** at ages 25 and 35 years) and the *correlation between these two successive measurements*. The greater the correlation, the narrower the adaptive unexposed curve. If there were no correlation between successive measurements, the cross-sectional and adaptive unexposed curves at age 35 years would be the same.

Take an exposed miner tested for the first time when beginning work at say age 25 years, with lung function at the value of the black star. At the second point at which lung function is tested (age 35 years) it is possible to compare his lung function test with the “normal” range for unexposed individuals who had the same lung function value at age 25 years as he did. This is done by observing where the measured value at age 35 years for the exposed miner falls within the expected adaptive unexposed distribution curve.

The adaptive unexposed curve is personalised for each exposed miner and has to be calculated by a computer programme based on that miner’s previous value(s). This can easily be programmed into a spirometer or attached PC. For personalised adaptive reference ranges to improve upon cross-sectional comparisons however, the adaptive exposed curve **[. . . . .]** **(**exposed miners at age 35 years who had similar lung function when they were 25 years) would need to be adequately separated from, and lower than, the adaptive *unexposed* curve (unexposed individuals at age 35 years who had similar lung functions as the exposed miners at age 25 years).

5.5 What are appropriate criteria for abnormality when using longitudinal data?

Abnormality criteria are important in relation to any set of reference values, whether cross-sectional or longitudinal.

Mean changes reported in the literature typically involve annual losses in FEV1 and FVC of the order of 20 - 30 ml for non-smokers, and about double that (40 - 50 ml) for smokers. In the *Louw et al .* study, average annual losses were 20 and 24 ml for FVC and FEV1 respectively for non smokers. Doubling these could be used a crude abnormality criterion for smokers. Mean changes in lung function are thus quite well characterised. The variability around these mean values, however, is more difficult to estimate and is not well reported in the literature. It leads to 95 percent confidence intervals of more than 0.5 litres around the mean change of 20 to 30 ml, or massive year-on-year change of > 20 percent of baseline lung function at the first measurement.

The American Thoracic Society (*1991*) uses a >= 15 percent decline in FEV1 year on year as threshold for normal/abnormal change. This is quite an extreme value corresponding to an expected loss over an entire working lifetime in the unexposed. It must be stressed that these large changes are based on the situation where one is comparing two values – a current lung function value with the immediately past value for the preceding year. There is no way in which such crude criteria can be used to follow a person across time with many measurements of lung function, as they lack sensitivity to early or gradual loss of function. These abnormality criteria are useful mainly for flagging processes involving chronic respiratory disease with accumulated loss based on a comparison of two measurements of function over a long time (e.g. COAD, pneumoconiosis), or alternatively rapid loss over a shorter annual period due to the advent of a significant respiratory illness (e.g. TB).

Tracking change in lung function measurements over time could be observed by plotting lung function values on a percentile chart, and would be a more sensitive indicator of early abnormality. This remains a qualitative or “informal” assessment, however, based on deciding the degree of abnormality involved in an individual’s crossing percentiles over time. For example, would the criterion require crossing one or two percentiles (5% or 10%)?

Recommendations regarding normal *change* in lung function cannot be based on cross-sectional data. Any rigorous assessment of what constitutes normal change will require longitudinal data which allow estimation of the correlation between successive measurements on an individual. Such adaptive reference ranges would allow the most precise quantitative assessment of an individual’s progress over time. This would provide a method for identifying as abnormal those individuals falling below the 5th percentile of the personalised adaptive reference range with a high degree of specificity.

5.6 Conclusions regarding longitudinal reference equations

1. It is possible to adopt a crude criterion of abnormal change such as that recommended by the ATS as a year on year change of >=15 percent. It should be recognised that this is an extreme degree of change.

2. Currently available cross-sectional reference ranges (such as *Louw et al*.) could be used to produce a percentile chart which could be used either in hard copy form, or as a programmed spirometer output, to track the progress of an individual across time. Such a chart would be for height adjusted lung functions with age. In practice, charts of lung function by age for each particular height could be generated in advance. When an employee is tested for the first time, their height is measured and they are assigned the particular chart corresponding to their height. This is a good starting point to get health professionals used to tracking individuals over time and, at least in an informal way, taking account of their prior lung function path.

3. The ideal would be to have reference ranges for changes in lung function (or adaptive reference ranges) over time, so that the observed change in an individual over time could be explicitly assessed for abnormality . Longitudinal data for unexposed individuals, however, do not currently exist, and would have to be generated by further research.

6. RECOMMENDATIONS

6.1 Reference equations for cross-sectional use

6.1.1 A single reference equation or different ones for different groups of mineworkers?

The following are the points of departure, following the workshop and further discussions with stakeholders:

1. The use of longitudinal monitoring, with the individual's baseline as his own control would go some way to reducing dependence on a single cross-sectional reference equation, in both the medical surveillance and compensation contexts (See section 5). It is therefore important that wherever possible, practice should move towards longitudinal monitoring. It is recognised that this will take some time.

2. An abnormal spirometric result, if to be used for decision making, needs to be interpreted wherever possible in a clinical context, while recognising that its predictive value as part of *screening* within the South African mining industry is largely unknown.

3. For entry examinations, and for most medical surveillance and compensation assessment for the foreseeable future, cross-sectional reference equations are still going to be needed to aid interpretation of spirometric test results.

4. The building of distinctions based on race into medical surveillance and compensation procedures is regarded as problematic, given the legacy of past racial discrimination, and the desirability of deracialising the practice of occupational health.

5. However, the practical consequences of choice of a particular reference equation should be clearly spelled out with regard to access to employment and security of job grade once employed, and access to compensation and to specific degrees of award. In particular it is felt that no significant group of employees should be materially disadvantaged by the way in which uncertainties in the scientific evidence are dealt with.

6. The use of scaling factors is not regarded as a desirable option.

6.1.1.1 Conclusions

1. Accumulated evidence internationally is that when the same methods are used for comparison (ideally in the same study), FVC is on average about 5 to 15% higher in people of European ancestry compared to people of African ancestry, at all ages, and controlling for height. In the Southern African context this has been observed in the studies of De Kock et al. (*1988*) of uranium miners, Louw et al. of bank employees, and in the very large study of South African goldminers by Hnizdo et al*. (2000b).* The same holds for FEV1 although the differences are somewhat smaller.

2. Whatever the source of these differences (see sections 3.1.2.4 and 3.1.2.5), there are practical implications in using a reference standard drawn from one population to judge the current status of an individual belonging to another. These practical implications are demonstrated in Tables 4.3a and 4.3b. above.

3. For example, applying the Louw et al. equation to the FEV1 of in-service black miners aged 30 to 35 years in service would result in 3,8 percent of individuals falling into the “statistically abnormal” group. Applying the ECCS equation to the same group, 12,5 percent would fall into this category. In other words, there is an increase of 8,7 percent in the proportion of individuals judged to have possibly abnormal lung function when ECCS rather than Louw et al. is used. Assuming that 220 000 employees have spirometry done in a three year period, 27 500 would be flagged as possibly abnormal using ECCS versus 8 360 using Louw et al. (It should be remembered that the observed population in this example excluded employees with pneumoconiosis or tuberculosis).

4. This implies that the specificity of the test falls significantly in moving from Louw et al. to ECCS, and would lead to larger numbers of individuals with normal function being flagged as abnormal - i.e. false positives. The consequences might include inaccurate information or advice given to the employee, unnecessary medical follow-up and even loss of grade for in-service workers or denial of employment to work-seekers (although the dependence of the latter action on lung function alone is unclear).

5. Conversely, applying (say) the Louw et al reference equation (i.e. the one based on black bank employees) to white mineworkers could result in inappropriately small proportions of such workers being flagged, resulting in false negatives. This might result in real early disease being missed.

6. Illustrating similar examples in the compensation context is difficult since the thresholds for compensation are somewhat lower than the 5th percentile (“below 65 percent of predicted” for first degree and below “51 percent of predicted” for second degree - see Table 2.2). Further, mineworkers typically come to compensation assessment at older ages, when the equations based on working populations are subject to greater uncertainty. It is possible that the practical implications of using different equations may be less marked in this context than in the surveillance context, but in the same directions.

7. In summary, there are practical implications, involving potential disadvantages for some groups in using one reference equation for everybody. The size of the effect is difficult to predict – much depends on what else other than lung function testing is used to make the relevant decision. The more that other clinical considerations are brought into contention, the smaller are likely to be the implications of using one reference equation rather than another.

6.1.1.2 Recommendation

Of the three options suggested in the recently published Guidance Note (*RSA 2000b*), the one most supported by available evidence is the use of separate equations appropriate for white and black miners. Scaling factors are not recommended (see section 3.2.1).

Applying the criterion of *primum non nocere* to medical examinations, i.e. that one should minimise harm, and given the limitations of current scientific knowledge, the use of two separate reference equations, one for black mineworkers and one for white mineworkers is justified.

This recommendation is qualified in the following way:

1. Because there is some evidence that lung function may improve in successive cohorts in a population where there is socioeconomic advance, reference values should be reviewed every 15 years or so to ensure that they are still applicable.

2. Should any practitioner believe that the practical consequences of using one rather than two distinct reference equations in the medical surveillance program are small (e.g. because mild departures from normality are ignored in the absence of other clinical indicators), such a user can always exercise the option of using one equation only, as both will be programmed into the spirometer. However, in such a situation, it is recommended that the user track the practical implications of such a choice on the relative proportions of black and white employees classified in a particular way, as a form of validity check.

6.1.2 White Southern African mineworkers

The use of the ECCS equations is recommended (Table 6.1.2a and 6.1.2b).

1. The populations used in that ECCS study represent the ancestral source populations of most white miners.

2. The Hnizdo et al. study has confirmed the close approximation of the ECCS equation to that derived from the study of white in-service mineworkers.

3. A further advantage is that the equation is already in wide use.

4. It should be noted that the ECCS equations are derived from a meta-analysis which includes some small studies, dust exposed populations and smoking populations. There are advantages of such a “mixed” population, however, which have been mentioned in section 3.

5. By parallel reasoning, the relevant ECCS equation could be used for white women in the mining industry requiring lung function testing.

Table 6.1.2 Prediction equations derived from ECCS (Quanjer 1983)

	Height (m)	**Age (yrs)***	**Intercept**	**SD**
Men				
FEV1	4.301	-0.029	-2.492	0.51
FVC	5.757	-0.025	-4.345	0.61
Women				
FEV1	3.953	-0.025	-2.604	0.38
FVC	4.426	-0.026	-2.887	0.43

*Values at age 25 years applied to subjects 18 to 24 years.

6.1.3 Black Southern African male mineworkers

The equation derived by Louw et al. is the preferred population, on the following grounds:

1. The subjects were in “white collar” i.e. not involving exposure to respiratory hazards.

2. The subjects were non-smoking.

3. Strict inclusion criteria were used to ensure healthy subjects.

4. The equation provided a good statistical fit when applied to the lung function distributions of in-service black mineworkers.

Table 6.1.3 Prediction equations derived from Louw et al.: black men

	Height (cm)	**Age**	**Intercept**	**SD***
FEV1	0.029	-0.027	-0.535	0.46
FVC	0.048	-0.024	-3.08	0.54

* Calculated by current authors from original Vitalograph data

The equations of Hnizdo et al. were considered less suitable for use as reference equations because of the combination of the pre-existing healthy mineworker effect compounded by the selection into the study (“double healthy worker effect”) and the cumulative dust effect at older ages)

6.1.4 Black Southern African female mine employees

Mokoetle et al. provide the only local reference equation for black women, derived from university employees in Johannesburg. This was therefore an unexposed population, although the exclusion criteria for the full sample were not as strict as those in the study by Louw et al.

The equation is included in Table 6.1.4

Table 6.1.4 Prediction equations derived from Mokoetle et al. for black women

	Height (cm)	**Age**	**Intercept**	**SD**
FEV1	0.034	-0.028	-1.87	0.39
FVC	0.045	-0.023	-3.04	0.41

6.2 Longitudinal tracking of lung function

At present there is very little in the way of longitudinal data, especially for unexposed, but also for exposed workers in the mining industry. There are no standard instruments in use for longitudinal tracking of workers either in South Africa or elsewhere, irrespective of the industry in which they work.

The recommendations in this section refer to the use that may be made of currently existing data in the form of either cross-sectional references for lung functions at various ages, or in the form of more than one measurement over time for individual workers. Recommendations are also made with respect to future research needs.

6.2.1 With current data

1. It is possible to adopt a crude criterion of abnormal change such as the ATS year on year change of >=15 percent for two measurements viz. current and last year’s values. It should be acknowledged that this is an extreme degree of change and would significantly lack sensitivity in detecting early decline. This would produce many false negatives and miss people with early or gradually developing abnormalities

2. Preferably, the prediction equations recommended above should be used in the interim to produce a percentile chart for lung function decay based on cross -sectional data. Miners crossing percentiles at periodic examinations should be flagged for preventive action. A decision would need to be taken about how many percentiles must be crossed before taking action. Candidate suggestions are changes in percentiles of 5 percent or more or 10% or more. Those with low values at initial examination, but otherwise healthy, should be flagged for special surveillance.

6.2.2 Outstanding methodological and operational issues

The possibility of developing more precise quantitative estimates (personalised adaptive reference ranges) of expected or normal decay of lung function over time requires further research to be done. In order for any future research to be worthwhile several methodological and operational issues need to be addressed.

Operational issues include:

1. The possibility of identifying a cohort of unexposed subjects similar in other ways to miners and of following their lung functions over time.

2. The possibility of obtaining quality cumulative exposure information for miners at all levels of exposure.

3. The possibility of obtaining quality lung function measurements for exposed miners over time.

Methodological issues are addressed in more detail in Appendix 2 and include:

1. The ability of lung function tests to distinguish between the dust exposed and the unexposed groups, and then between groups at different levels of dust exposure.

2. Questions of lung function test validity (sensitivity and specificity):

   What is the concurrent validity of test results in respect of their correlation with other markers of abnormality such as symptoms, effort tolerance, and chest x-ray findings?

   What is the predictive validity of lung function tests, i.e. are the lowest 5 percent at greater risk of future disease or acceleration of lung function loss due to dust, than those above that threshold?

   Put differently, what is the criterion validity against current or future respiratory diseases of miners being classified in the lowest 5 percent of FEV1 or FVC as a definition of screening abnormal? Further, what are the clinical diagnostic entities that would serve as a gold standard against which to evaluate low lung function at the screening test?

   These questions reflect critically the clarity or otherwise of the objectives in performing lung function testing in medical surveillance programmes in industry. They are discussed in more detail in Appendix 3.

6.2.3 Future Research Needs

The authors recommend that truly adaptive reference ranges should be generated by further research in a longitudinal study in the mining industry for the most sensitive and earliest detection of abnormal loss of lung function. Such research addresses the questions raised in section 6.2.2.

Specifically, the following research is recommended in order to progress from our current situation where we are wholly dependent on cross-sectional reference values external to the industry or country for evaluating miners' lung functions. Three phases of research are proposed:

1. A cross-sectional study of exposed miners to determine whether given the existence of cumulative exposure information for silica and general dust, it is possible to distinguish between the lung function distributions for low, medium and highly exposed miners (equivalent to the unexposed, A and B population lung function distributions in Figure A3). Such a study could begin by comparing miners in the age range 60 to 65 years across three exposure groups (low, medium and high) to obtain maximal contrast. If appropriate, the study could be extended to include all age groups across the employable range.

2. A longitudinal study of exposed miners based on follow-up measurements of a cohort to see whether it is possible to distinguish between the adaptive distributions (distributions conditional on previous lung function measurements) of low and high exposure groups. This is equivalent to being able to distinguish between the adaptive unexposed curve and the adaptive exposed curve in Figure 5.4. Lung function measurement quality assurance protocols recommended elsewhere in this report should result in the generation of quality longitudinal data on exposed miners. It is to be hoped that cumulative exposure data will also increasingly be available on miners in future.

3. If the results of phases 1 and 2 indicate that such separation is possible and therefore that (I) there is a dose-response relationship of lung function with exposure, and (2) adaptive reference ranges have utility in providing improved sensitivity compared with the use of purely cross-sectional reference values, it is recommended that the mining industry and SIMRAC invest in the establishment of a cohort of *unexposed* subjects similar in other ways to miners for the purposes of longitudinal follow-up of their lung functions. The unexposed should be selected from the source population of miners. An exposed cohort of miners should be selected from mines with quality routine medical surveillance systems and studied in parallel to the unexposed cohort.

A more detailed research proposal would need to be developed for this longitudinal study of unexposed workers and exposed mineworkers to collect the data necessary for the development of an effective longitudinal tracking instrument.

6.3 Recommendations regarding implementation, including quality control

6.3.1 Statistical lower limit of normal or percent predicted?

The American Thoracic Society recommends the use of a statistical threshold,, viz. spirometric values below the fifth percentile (approximately 1.64 times the SD for any individual tested) should be regarded as being below the expected range. This is clinically the most appropriate and avoids the overestimate of the appropriate threshold at older ages which occurs when a fixed percentage of predicted (e.g. 80 percent) is used as the threshold.

However, there are a number of factors to take into account in using the statistical definition.

- Since users have got used to using the “percentage of predicted” concept, training will be needed to switch users to the percentile concept implied by use of standard deviations.
- Impairment classification systems like that of the ATS use percentage of predicted.
- The statistical definition cannot be computed manually in the same way that percentage of predicted can, and will rely on programming of the software in the equipment used. This is, however, a relatively simple matter.

It is therefore recommended that the statistical definition be programmed into spirometers to define the lower limit of normal, when a categorical classification (normal vs abnormal) is required.

The percentile on which the individual lies is also capable of providing a quantitative scale which takes into account the actual distribution of lung functions of people of that age and height.

It should be emphasised that whatever the reference equation or criterion of abnormality used a single abnormal test value needs to be interpreted with caution.

6.3.2 Quality assurance

The point has been repeatedly made among practitioners that quality assurance is as important as choice of the appropriate reference equation.

Recommendations

1. A Guidance Note on Quality Assurance should be developed by MOHAC. A sample version of such a Guidance note prepared by one of the authors is appended as Appendix 4.
2. The critical shortage of skilled personnel to carry out lung function testing needs to be brought to the attention of the authorities responsible for funding clinical technologist posts and training posts in public sector institutions. The MBOD may have a role in subsidising such posts in state facilities where there is a significant load of benefit examinations being done.

3. Given that most spirometry in industry will be done by people who are not pulmonary function technologists, quality assurance is needed in the training of such testing personnel in spirometry. Liaison with training institutions and professional bodies is needed. A note on what such training might entail is included as Appendix 5.

4. Methods are needed for doing in-house quality checks. Reliability testing is recommended.
   This can be done in two ways. One method is more demanding in that it demands repeat measurements of individuals within a short period of time to calculate a reliability coefficient (see *Hnizdo et al. 1999*).

   A simpler method which is operationally easier is to take a mean of all the lung function tests (*Hankinson et al. 1993*) done in a month and to compare this with the mean for the previous month. The difference should not be more than 2 SDs. A big difference calls for a review of the lung function test measuring process.

5. Quality assurance checks could also be built into the software of equipment This should allow automatic calculation of rolling averages of the last batch of spirograms for all the lung function indices using the criterion set out in point 4 above.

6.3.3 Additional information to be captured at time of lung function testing

Additional information that is of relevance to the interpretation of lung function tests should be recorded at the time of lung function testing. The purpose of occupational lung function screening is to detect an accelerated loss of lung function due to an occupational exposure. A number of non-occupational exposures could result in loss of lung function. It would be important to try control for these other factors when interpreting lung function values.

Ideally, a detailed history would be obtained with respect to:

- Smoking, If the individual has ever smoked the following history should be noted, date started and stopped, the brand and the average amount smoked in pack years. The form in which the tobacco is smoked should also be recorded, i.e., commercial cigarette, hand rolled, pipe or a combination.
- Occupation. The job description and duration of each occupation should be obtained from the period of first exposure. It would be ideal if the occupational history could be linked to the hygiene exposure data for each employee. This would enable a more precise determination of a dose response relationship.
- Pulmonary tuberculosis (TB). This is associated with a significant loss of lung function which resolves partially with time (*Hnizdo et al. 2000a*). The degree of loss of lung function increases with each episode of pulmonary TB. If an individual has had an episode of pulmonary TB then the date and treatment category i.e., new or retreatment TB and treatment outcome, should be recorded.
- Radiological silicosis. This is associated with loss of lung function and its presence and grade of severity, according to the International Labour Organisation system, should be recorded.
- Other acute or chronic pulmonary diseases, which may also impair lung function tests should be noted. This includes clinical conditions such pneumonia, asthma, chronic bronchitis and bronchiectasis.
- Weight gain may also be an important determinant of longitudinal lung function loss (*Wang et al. 1996*). The individual’s weight and body mass index should be recorded at the time of lung function testing.

6.3.4 Incorporation of recommended reference equations into existing equipment

Recommendation:

1. It is recommended that bodies responsible for medical surveillance, e.g. the Mine Medical Officers' Association liaise with the suppliers of lung function equipment trade regarding the programming of the requisite software into new spirometers bought for use in occupational health centres.

2. The implications of reprogramming existing spirometers need to examined.

6.3.5 Publicisation and adoption

Recommendations:

1. Future revisions of the Guidance Note for Lung Function testing issued by MOHAC should take the findings of this report in account.

2. Liaison with The South African Society for Occupational Medicine, The Mine Medical Officers' Association and the South African Pulmonology Society is needed.

6.3.6 Spirometry in Benefit examinations for ex-mineworkers

Under the Occupational Diseases in Mines and Works Act, all ex-mineworkers are entitled to Benefit Examinations at specified minimum intervals to determine whether they are suffering from an occupational disease. As apparent from section 2.1.2 above, spirometry is relevant to the determination of degree of impairment in all occupational lung disease.

The current policy of the Medical Bureau for Occupational Diseases is that ex-mineworkers seeking Benefit Examinations should be referred to approved medical facilities, including provincial facilities.

Concern has been expressed about the availability of spirometry and its quality in state facilities, particularly in remote areas.

The current situation is that lung function tests are not regarded as essential in the performance of Benefit examinations under the Occupational Diseases in Mines and Works Act (Dr A. Banyini, Director, Medical Bureau for Occupational Diseases – personal communication). This is because many such examinations are carried out at provincial facilities or by practitioners who do not have lung function testing capacity.

The Certification Committee of the MBOD may ask for the claimant to be referred for lung function testing if the clinical information so indicates. However, even in the absence of lung function results, the Committee may award 2nd Degree benefits on clinical grounds alone. The establishment or standardisation of lung function testing capacity at remote provincial sites, a difficult task given severe resource constraints within the Department of Health, is thus not regarded as a priority by the Medical Bureau for Occupational Diseases.

Recommendation:

A guidance note on lung function testing for provincial departments of health and private providers doing Benefit Examinations should be prepared by the Medical Bureau for Occupational Diseases.

6.3.7 Peak Expiratory Flow Rates (PEFR)

PEFR measures have been advocated and used as screening tests for lung function testing in the mining industry.

Recommendation:

The use of this test is no longer regarded as acceptable for the following reasons:

1. PEFR will not detect restrictive lung function abnormalities caused by obstructive lung disease.

2. PEFR is not a sensitive test for chronic obstructive pulmonary disease (COPD). Much of the early abnormality in COPD occurs in mid-expiratory flow which is not detected by PEFR.

3. As a test, PEFR is known to have a high coefficient of variation for intra-individual measurements. This property diminishes it’s usefulness for repeat measures, as used in occupational medical surveillance.

References

American Thoracic Society. **1986.** Quality assurance in pulmonary function laboratories. *American Review of Respiratory Disease*, 134:625-627.

American Thoracic Society. **1991.** Lung function testing: selection of reference values and interpretative strategies. *American Review of Respiratory Disease*, 144:1202-1218.

American Thoracic Society. 1995. Standardization of spirometry. 1994 Update. *American Journal of Respiratory and Critical Care Medicine,* 152:1107-1136.

Barker, D.J.P., Godfrey, K.M., Fall, C., Osmond. C., Winter. P.D. & Shaheen, S.O. 1991. Relation of birth weight and childhood respiratory infections to adult lung function and death from chronic obstructive lung diseases. *British Medical Journal*, 303:671-675.

Becklake, M.R. 1989. Occupational exposures: evidence for a causal association with chronic obstructive pulmonary disease. *American Review of Respiratory Disease*, 140:S85-S91.

Becklake, M.R. 1990. The epidemiology of spirometric test failure. *British Journal of Industrial Medicine.* 47: 73-74.

Becklake, M.R. & White N.W. 1993. Sources of variation in spirometric measurements. Identifying the signal and dealing with noise. (*In: Eisen, E,A, Occupational Medicine. State* of *the Art Reviews, Vol. 8, No. 2 : Spirometry.* Philadelphia: Hanley and Belfus Inc.)

British Thoracic Society and Association of Respiratory Technicians and Physiologists. **1994**. Guidelines for the measurement of respiratory functions. *Respiratory Medicine*, 88: 165-194.

Burrows, B., Lebowitz, M.D., Camilli, A.E. & Knudson, R.J. 1986. Longitudinal changes in forced expiratory volume in one second in adults: methodological considerations and findings in healthy non-smokers*. American Review of Respiratory Disease*, 133: 960-974.

Coetzee, A.M., & Becker, P.J. 1989. Lung function screening in industry. *South African Medical Journal* 67:550-553*.*

Cotes, J.E., Dobbs, J.M., Hall, A.M., Lakhena S.C., Saunders, M.J. & Malhotra, M.S. 1975. Lung function in healthy men in India: contributory roles of genetic and environmental factors. *Proceedings of the Royal Society of London (Biology).* 191:413-425*.*

Cowie, R.L, & Mabena, S.K. 1991. Silicosis, chronic airflow limitation, and chronic bronchitis in South African Gold Miners. *American Review of Respiratory Disease*, 143:80-84.

De Kock, M.A., Swiegers, W.R.S., Kotze, T.J. vW. & Joubert, G. 1988. Cross-sectional study of uranium mine workers to develop predictive equations for lung function with reference to chronic obstructive pulmonary disease. *South African Medical Journal* (suppl.), 3:1-20.

Diamond, J. 1994. Race without colour. *Discover, 15 (Nov.*) 82-91.

Dockery, D.W. 1993. Percentile curves for evaluation of repeated measures of lung function. (*In: Eisen, E,A, Occupational Medicine. State* of *the Art Reviews, Vol. 8, No. 2 : Spirometry.* Philadelphia: Hanley and Belfus Inc.)

Fletcher, C. & Peto, R. 1977. The natural history of chronic airflow limitation. *British Medical Journal,* 1:1645-1648*.*

Goldin, J.G. 1989. Spirometric and gas transfer measurements among normal South African men – an investigation into anthropometric, socioeconomic, racial and environmental factors influencing lung function. *Ph.D. Dissertation.* Cape Town: University of Cape Town.

Goldin, J.G., Louw, S.J. & Joubert ,G. 1996. Spirometry of healthy adult South African men. Part II. Interrelationship between socio-environmental factors and 'race' as determinants of spirometry.. *South African Medical Journal*, 86:820-826.

Glindmeyer, H.W., Diem, J.E., Jones, R.N. & Weill, H. **1982.** Noncomparability of longitudinally and cross-sectionally determined annual lung function change in spirometry. *American Review of Respiratory Disease*, 125: 544-548.

Hankinson, J.L. & Wagner, G.R. 1993. Medical screening using periodic spirometry for detection of chronic lung disease. (*In: Eisen, E,A, Occupational Medicine. State* of *the Art Reviews, Vol. 8, No. 2 : Spirometry.* Philadelphia: Hanley and Belfus Inc.)

Hankinson JL, Krisley KB, Wagner GR. **1996**. Comparison of spirometric reference values for Caucasian and African American blue collar workers. *Journal of Occupational and Environmental Medicine,* 38: 137.

Hankinson, J.L., Odencrantz, J.R., Fedan, K.B. 1999. Spirometric reference values from a sample of the general U.S. population. *American Journal of Respiratory and Critical Care Medicine*, 159:179-187*.*

Hessel, P.A. & Sluis-Cremer, G.K. Prediction equations for lung function in black industrial workers at Phalabora mining company*. South African Medical Journal,* 76:548-549*.*

Hnizdo, E. 1992. Loss of lung function associated with exposure to silica dust and with smoking and its relation to disability and mortality in South African gold miners. *British Journal of Industrial Medicine,* 49:472-479*.*

Hnizdo, E., Churchyard, G.J., Barnes, D., & Dowdeswell, R. 1999. Assessment of reliability of lung function screening programs or longitudinal studies. *American Journal of Respiratory and Critical Care Medicine,* 160:2006-2011*.*

Hnizdo, E., Singh, T. & Churchyard, G.J. 2000a. Chronic pulmonary function impairment caused by initial and recurrent pulmonary tuberculosis following treatment. *Thorax,* 55:32-38*.*

Hnizdo, E., Churchyard, G. & Dowdeswell, R. 2000b Lung function prediction equations derived from healthy South African gold miners. *Occup Environ Med* (in press).

Hoffman, P. 1994. The science of race. *Discover,* 15 (Nov): 4.

Johannsen, Z. & Erasmus, L.D. 1968. Clinical spirometry in normal Bantu*. American Review of Respiratory Disease*. 97:585-597*.*

Jacobs, D.R., Nelson, E.T. , Dontas, A.S., Keller, J. & Slattery, M.L., Higgins M. 1992. Are race and sex differences in lung function explained by frame size? The Cardia study. *American Review of Respiratory Disease, 146:644-649.*

Knudson, R.J, Slatin, R.C., Lebowitz, M.D. &, Burrows. B. 1976. The maximal expiratory flow-volume curve. Normal standards, variability and effects of age. *American Review of Respiratory Disease*, 113:587-600.

Lebowitz, M.D. &, Holberg, C.J. 1989. Comparisons of spirometric reference values and the proportions of abnormal subjects among male smokers and those symptomatic in a community population. *American Review of Respiratory Disease*. 141:1491-1495.

Louw, S.J., Goldin, J.G. & Joubert, G. 1996. Spirometry of healthy adult South African men. Part I. Normative values. *South African Medical Journal*, 86:814-819.

Massey, D.G., & Fournier-Massey. G. 1986. Japanese -American pulmonary reference values – influence of environment on anthropology and physiology. *Environmental Research,* 39:418-433.

Mokoetle, K., de Beer, M. & Becklake, M.R. 1994. A respiratory health survey of a black Johannesburg workforce. *Thorax*, 49: 340-346.

Molnar, S. 1998. Human Variation. Races, types and ethnic groups. New Jersey: Prentice Hall.

Myers, J.E. 1984, Differential ethnic standards for lung functions, or one standard for all? *South African Medical Journal,* 65:768-772.

Myers, J.E., Garrisch, D., Myers, H.S, & Cornell, H.S. 1985. A respiratory epidemiologic survey of stevedores intermittently exposed to asbestos in a South African port. *American Journal of Industrial Medicine,* 7:273-283*.*

Quanjer, Ph. H. 1983. (Ed). Report of the Working Party on Standardisation of Lung Function Tests: European Community for Coal and Steel. *Bulletin Europeen de Physiopathologie Respiratoire*; 19 (suppl. 5): 7-95.

Quanjer, P.H., Tammeling, G.J., Cotes, J.E., Pedersen, O.F., Peslin, R. & Yernault, J.C. 1993. Lung volumes and forced ventilatory flows. Respiratory Working Party Standardization of Lung Function Tests. European Community For Steel and Coal. Official Statement of the European Respiratory Society. *European Respiratory Journal* (suppl.), 16:5-40.

Richards, G. COPD and occupation. *South African Respiratory Journal* (in press).

Republic of South Africa. Department of Health. 1998. South Africa Demographic and Health Survey 1998: Preliminary Report. Pretoria: Government Printer

Republic of South Africa. Department of Minerals and Energy. Mine Health and Safety Inspectorate. 2000a. *Guideline for the Compilation of a Mandatory Code of Practice for the Establishment of a System of Medical Surveillance*. Pretoria: DME, March 2000

Republic of South Africa. Department of Minerals and Energy. Mine Health and Safety Inspectorate. 2000b. Guidance Note for Occupational Medical Practitioners: Lung Function Testing. Pretoria: Government Printer. (*30 March 2000*)

Republic of South Africa. Department of Health. 2000c. South Africa Demographic and Health Survey : Final Report. Pretoria: Government Printer (in preparation).

Schoenberg, J.B., Beck, G.J. & Bouhuys, A. 1978. Growth and decline of respiratory function in healthy blacks and whites. *Respiratory Physiology* 33:367-393.

Senior, P.A., & Bhopal, R. 1994. Ethnicity as a variable in epidemiologic research***.*** *British Medical Journal,* 309: 327-330*.*

Steinberg, M. & Becklake, M.R. 1986. Socio-environmental factors and lung function. *South African Medical Journal,* 70:270-274.

Terblanche, P, Nel, R, & Danford, I. 1993. Health and safety aspects of household fuels. Phase II. Report to the Department of Mineral and Energy Affairs. Pretoria: Council for Scientific and Industrial Research.

The National Lung Health Education Program Executive Committee. 1998. Strategies in preserving lung health and preventing COPD and associated diseases. *Chest*,113 (suppl.): 123S-163S.

Thompson, M.L. & Fatti, L.P. 1997a. The use (and abuse) of reference centiles with an application to weight gain in pregnancy. *South African Medical Journal*, 87:206-213.

Thompson, M.L. & Fatti, L.P. 1997b. Construction of multivariate centile charts for longitudinal data. *Statistics in Medicine,* 16:333-345.

United State of America. National Institutes of Health. National Cancer Institute. 1998. Report of the President's Cancer Panel. The Meaning of Race in Science - Considerations for Cancer Research. Bethesda.

Vollmer, W.M, Johnson, L.R., McCamant, L.E. & Buist, S.A. 1988. Longitudinal versus cross-sectional estimation of lung function decline – further insights. *Statistics in Medicine,* 7: 685-696.

Vollmer, W.M. 1993 Reconciling cross-sectional with longitudinal observations on annual decline. (*In: Eisen, E,A, Occupational Medicine. State* of *the Art Reviews, Vol. 8, No. 2 : Spirometry.* Philadelphia: Hanley and Belfus Inc.)

Wang, M-L., McCabe, L., Hankinson, J.L., Shamssain, M.H., Gunel, E., Lapp, N.L. & Banks, D.E. 1996. Longitudinal and cross-sectional analyses of lung function in steel workers. *American Review of Respiratory and Critical Care Medicine*, 153: 1907-1913.

White, N., Hanley, J.H., Lalloo, V.G., & Becklake, M.R. 1994. Review and analysis of variation between spirometric values reported in 29 studies of healthy African adults. *American Journal of Respiratory and Critical Care Medicine*, 150: 348-55.

White, N.W. 1989. Byssinosis in South Africa: a survey of 2411 textile workers. *South African Medical Journal;* 75:435-442

Yach, D., Myers, J., Bradshaw, D., & Benatar S.R. 1985. A respiratory epidemiologic survey of grain mill workers in Cape Town, South Africa. *American Review of Respiratory Disease*, 131:505-510.

Yang, T-S., Peat, J., Keena, V., Donnelly, P., Unger, W. & Woolcock, A. 1991. A review of the racial differences in the lung function of normal Caucasian, Chinese and Indian subjects. *European Respiratory Journal,* 4:872-880.

APPENDIX 1

Statistical note on the use of the terms standard deviation and standard error applied to prediction equations.

The term “standard error of the estimate” is typically used to indicate the square root of the estimated variance of an estimate.

In the simplest case, assume that one is wanting to estimate the mean μ of a variable Y which has a N(μ, σ^2) distribution. The sample mean estimates μ and the standard error of this estimate is $\sigma / \sqrt{n}$. What this means is that, in repeated samples of the same size, the variability of the sample mean (i.e. the variability in the estimate of μ) is reflected by $\sigma / \sqrt{n}$. This does not represent the variability in Y which is reflected by σ.

The regression context is simply a generalisation of the above example, where the mean of Y varies with the level of one or more covariates, X. The estimate of the mean of Y is the regression equation (e.g. the slope β) and the term *standard error of the estimate* should refer to the variability in this estimate (e.g. the SE(β), and not to the variability in Y itself at that level of X. Standard regression analysis texts will distinguish between a *confidence interval* for the regression line (based on the standard error of the estimated regression line) and a *prediction interval* which represents the “typical” range of values for Y at each X. It is this latter interval which is of interest in establishing reference ranges for Y as a function of covariates.

See as attachment: *Figure A1 Prediction limits for individual subjects tested and confidence limits for the population mean*

APPENDIX 2

A2.1 Methodological considerations in developing reference ranges

The authors wish to highlight several important methodological considerations in the development of reference ranges.

1. All existing South African cross-sectional equations are based on *assumptions* that need to be tested**.**

They all assume (1) that the variance in lung function does not change with age and height, (2) that the mean lung function is associated in a linear fashion with age and height, and (3) that lung function at given age and height is normally distributed.

While it may be that linearity and normality are acknowledged approximations to the true relationship and are felt by some investigators to be reasonable, it is not clear that homogeneity of variance has been assessed. The consequence of this assumption is that all the percentiles are artificially parallel to the mean line.

One way of validating any particular set of equations in relation to some of the concerns expressed above is to compare observed empirical percentiles of the data on which the reference equations were based with the predicted percentiles based on the estimated prediction equation.

2. The prediction equations are for *mean lung function* and not the height adjusted lung function values used in some of the literature (e.g. where FEV/ht^2 is modelled on age). Equations based on different models are not directly comparable.

3. The utility of the lower percentiles in being able to detect exposure-related “abnormality” needs to be assessed by exploring whether the percentage of individuals falling below a certain lower limit of normal increases with increasing level of exposure, i.e. is there a *dose-response relationship* with dust, for example?

4. There are additional validity considerations relating to the objectives of lung function surveillance: such as the relationship between what the tests are being used to detect, on the one hand, and the inherent sensitivity and specificity of these tests, on the other (i.e. do they discriminate between exposed or unexposed workers, or between those with or without a condition requiring flagging or attention).

5. Lastly, *choice of the 5th percentile as the definition of “abnormal” is arbitrary*. It would be of interest to evaluate the Receiver Operating Curve characteristics relating to different choices of abnormality threshold as a screen for some defined outcome (e.g. using a clinical diagnosis of Chronic Obstructive Pulmonary Disease as a gold standard).

APPENDIX 3

Note on the potential performance of lung function screening tests in meeting their objectives: validity (sensitivity and specificity) considerations

A3.1 Once-off cross sectional comparisons

It is important to point out that even for one-off measurements at a single time point, say 25 years, the utility of lung function testing for surveillance depends on there being separation between the lung function distributions of exposed and unexposed to the risk factor (e.g. dust) thought to be associated with the outcome of interest (e.g. respiratory impairment).

See attached *Figure A3: Sensitivity and specificity of a test against a gold standard*

Specificity is determined by the screening threshold chosen on the unexposed distribution e.g. all individuals below the 5th percentile are screened as abnormal. This latter definition implies a specificity of 95 percent as 95 percent of those screening negative are by definition normal.

The corresponding sensitivity (or the percentage of those falling below the 5 percent level of the unexposed distribution) depends on the degree of separation of the exposed distribution from that of the unexposed. In the case of exposed population A (e.g. a low exposed population) in Figure A3, this will be greater than 5 percent (by definition in the unexposed) and around 40 percent. For population B (e.g. a high exposed population), whose distribution is even more separated from the unexposed distribution the sensitivity will be around 90 percent.

However this evaluation of screening effectiveness is somewhat circular as the same instrument is being used both to define and evaluate the screening procedure. An external gold standard based on a definition of health status (e.g. clinically defined chronic obstructive pulmonary disease) or impairment which is independent of lung function evaluation would be preferable. In lung function surveillance, abnormality is being defined somewhat arbitrarily in terms of lung function values. One is then at best assessing sensitivity and specificity of a lung function classification system in its response to exposure rather than the accuracy of lung function values relative to an external gold standard. This raises important questions about the objectives of lung function testing, as outlined in section 2.3. Those conducting lung function surveillance need to be clear about these objectives.

A3.2 More than one comparison (longitudinal comparisons) with the reference values

Any assessment of the characteristics of a screening tool involves a consideration of both sensitivity and specificity and the trade-off between the two. For instance, referring back to

Figure A3 , if "abnormality" were defined as occurring below the 10th percentile instead of the 5th, then sensitivity would increase (more exposed individuals would be identified) but specificity would decrease (more unexposed individuals would be incorrectly identified).

With repeated screenings, the specificity of the overall surveillance scheme will decrease (there is an increased chance of false positives), but it would be hoped that there is also an increase in sensitivity (increased true positives).

The first time an individual is screened, the probability that an unexposed individual will not be screened positive is by definition 0.95 if the 5th percentile of the unexposed is used as a threshold. If an individual is screened on two occasions and the measurements are uncorrelated, and cross-sectional centiles (cross sectional unexposed curves in Figure 5.4) are used, an unexposed individual will have a 0.95^2 probability of not screening positive on either of the two tests. And so on. Hence by the time a fifth measurement is taken in this setting, specificity is no longer 95 percent, but rather 0.95^5 or about 77 percent.

In other words the chance of an unexposed individual being (correctly) identified as "normal" at each of 5 successive assessments (when the results of each assessment are assumed to be independent) is only 77 percent, or, with 5 successive evaluations, 23 percent of unexposed individuals will be screened "abnormal" at some point thereby generating more work and incurring greater costs for the occupational health service

If there is correlation between measurements at different times, the specificity will be still be reduced (relative to the nominal 0.95) with successive evaluations, but the extent of the reduction will be less than in the case of independence and will depend on the (unknown) degree of correlation.

A motivating idea behind adaptive reference ranges is that they will have improved sensitivity relative to cross-sectional assessments that ignore the prior path of the individual. The specificity of adaptive longitudinal screening (equivalents of the adaptive unexposed curve) is well-defined. Here the specificity after 5 measurements (this time based on the adaptive unexposed curve) will be 77 percent. However, one would anticipate that the associated sensitivity will be superior to using cross-sectional reference values, and that one would be detecting someone who falls below the 5th adaptive centile (on the adaptive unexposed curve) at an earlier time point than would have occurred if they were to fall below the 5th cross-sectional centile (cross-sectional unexposed curve).

APPENDIX 4

Draft guidance note for quality assurance in lung function testing for medical surveillance

Quality control programmes are necessary to ensure a well functioning lung function testing unit and to secure clinically useful results. The principles of a lung function quality assurance programme are identical for screening, diagnostic and compensation lung function units.

The American Thoracic Society (ATS) has produced comprehensive documents regarding the standardisation of spirometry (ATS 1994) and quality assurance in lung function laboratories (ATS 1986). The ATS guidelines should be modified to provide minimum standards for spirometry and quality control.

In summary, the basic requirements of a lung function quality assurance programme are:

1. Responsible practitioner. Appointment of an individual responsible for overseeing the lung function testing and quality assurance programme is critical for ensuring a well functioning programme.
2. Procedure manual. A procedure manual should be present which should contain the quality control programme, guidelines regarding performance of spirometry and reporting of results, and appropriate reference values.
3. Testers. Minimum standards of training should be agreed to. The most important component in successful spirometry is a well-motivated enthusiastic technician. The quality control programme should ensure continuous monitoring of technician performance and regular feedback regarding their performance.
4. Hygiene and infection control. Recommendations would be determined by the setting under which lung function testing is being done. For mass lung function screening only an expiratory effort should be recorded and therefore the use of a disposable mouthpiece is all that is required. However, when closed circuit techniques are used, exacting requirements should be met for infection control as outlined in the ATS guidelines.
5. Equipment quality control. Attention to equipment maintenance and calibration are an important part of good practice. A calibration and maintenance log should be kept.
6. External audit. National or industry approved laboratories should be responsible for accrediting pulmonary function units on a regular basis.
7. Internal audit. Internal audit should also be carried out on a more frequent basis (every 3 months). Ongoing surveillance of the lung function programme can be achieved through monitoring a rolling mean for both the equipment and technicians. Similarly, a reliability (G) co-efficient for equipment and technicians can be determined by re-testing a proportion of individuals within a period of less than one year (See *Hnizdo et al. 1999*).

APPENDIX 5

Note on training and certification to perform spirometry

A5.1 Current situation and required action

Current arrangements for training and certification to perform spirometry for occupational medical purposes are not optimal. Professional bodies, such as South African Society of Occupational Medicine and The South African Pulmonology Society, have been aware of this situation for quite some time, but for a variety of reasons have been unable to bring about a satisfactory solution which enables a system of training to be put in place that will ensure that there is an acceptable standard of performance of spirometry for medical surveillance/primary care purposes.

At the present time performance of spirometry is one of the competencies of medical technologists who choose to specialise in pulmonary technology. Trainees is this discipline register and receive part of their basic training through certain of the technikons. Their practical training is completed in a tertiary centre. In recent years there has been an alarming decline in the number of new students in the field of pulmonary technology, largely because of freezing of training posts the public sector tertiary hospitals.

Clearly only certain of the skills of a pulmonary technologist are required to perform spirometry for occupational medical purposes. Training of other health care professionals in these skills does take place but tends to be ad-hoc, unstandardised and does not lead to a recognised certification.

Recognised training and certification is already in place for performance of screening audiometry and similar mechanisms need to be put in place for training and certification in performance of screening spirometry.

The steps that are required in such a process are:

1. Consensus among professional bodies and interest groups on the syllabus and core skills that need to be acquired for certification and re-certification in performing spirometry. This process has been going on for some time but needs to be brought to finality.

2. Certifying authorities will also need to meet certain criteria and consensus is also required on this.

3. With consensus on curriculum/skills to be acquired the matter should be considered by MOHAC for referral to the Mine Qualification Authority (MQA) for consideration. What requires to be formalised is the definition, in terms of the Mine Health and Safety Act, of what constitutes a competent person for performance of spirometry.

4. With recognised competencies, accepted by the MQA formal certification can take place. Once adopted for the mining sector these standards are sure to become more widely used.

A5.2 Proposed curriculum for training in spirometry

1. Entry criterion: any person already registered with Health Professions Council.

2. Twelve hours of theoretical and practical training to enable the candidate to:

    2.1 Assess and select an appropriate spirometer and understand the principle of it's working mechanism.
    2.2 Calibration.
    2.3 Instruct correct subject technique.
    2.4 Evaluate test data for acceptability and reproducibility.
    2.5 Select the best test.
    2.6 Select appropriate reference values and understand their applications and limitations.
    2.7 Recognize common patterns of functional disorders.
    2.8 Carry out sterilisation.

3 Proof of acquisition of skills through:

    3.1 Completing Multiple Choice Questions on aspects of theory
    3.2 Returning 20 acceptable spirograms performed without supervision.

A5.3 Certification centres

Any centre wishing to certify candidates in performance of spirometry should have or include access to:

1. A faculty that includes a registered pulmonary trained medical technologist and a medical specialist in internal medicine, preferably a pulmonologist.
2. A pulmonary function laboratory with an adequate daily average number of tests performed and a range of equipment.

***