PROJECT NUMBER

SIM 03-08-02

TITLE OF PROJECT

Development of sensitive tools for active case finding of tuberculosis (Phase 1)

PRIMARY OUTPUTS

Review of sensitivity and costs of existing and novel methods for active case finding

Analysis of any existing unpublished data and comparison of old and new tools

Systematic review or meta-analysis of data, if appropriate

Industry stakeholder workshop to discuss potential recommendations

Proposal for Phase II further evaluation of tool(s) if indicated

CONTRACTING ORGANISATION:

BUSINESS ENTERPRISES AT THE UNIVERSITY OF PRETORIA

PRINCIPLE INVESTIGATOR BV Girdler-Brown[1]
CO-INVESTIGATORS J Murray[2]
K Weyer[3]
S Thooe[1]
P Godfrey-Faussett[4]

[1] University of Pretoria
[2] National Institute for Occupational Health
[3] Medical Research Council of South Africa
[4] London School of Tropical Medicine & Hygiene

CONTRACT PERIOD   April 1 2003 to April 30 2004

ACKNOWLEDGEMENTS

Apart from the project leader, the following people attended the workshop and took an active and enthusiastic part in the deliberations:

Dr.	Dave	Barnes	Anglogold-Ashanti
Dr.	Lucille	Blumberg	National Institute for Communicable Diseases
Dr.	Salome	Charalambous	Aurum Health Research
Prof.	Gavin	Churchyard	Aurum Health Research
Prof.	Mary	Edginton	University of the Witwatersrand
Dr.	Ismael	Ganchi	Gold Fields
Dr.	Johan	Geyser	Harmony
Dr.	Peter	Godfrey-Faussett	London School of Hygiene and Tropical Medicine
Dr.	N	Govender	Angloplatinum
Dr.	Rob	Harrison	Angloplatinum
Dr.	Spo	Kgalamono	National Institute for Occupational Health
Dr.	Boitumelo	Mashitisho	National Institute for Occupational Health
Dr.	Thulani	Matsebula	Aurum Health Research
Dr.	Charles	Mbekeni	Chamber of Mines of South Africa
Dr.	Colin	Mupombwa	Billiton
Dr.	Jill	Murray	National Institute for Occupational Health
Dr.	Nesri	Padayatchi	Centers for Disease Control/University of Kwazulu-Natal
Dr.	Lindiwe	Pemba	Aurum Health Research
Mrs.	Amelia	Richards	Business Enterprises at the University of Pretoria
Prof.	Mary	Ross	Mines Health and Safety Council
Dr.	Stuart	Shearer	Gold Fields
Dr.	Jim	te Water Naude	University of Cape Town
Dr.	Selo	Thooe	University of Pretoria
Mr.	Paul	van der Heever	Mines Health and Safety Council
Dr.	Ingrid	Weber	University of Pretoria
Dr.	Michelle	Wong	Chris Hani Baragwanath Hospital

EXECUTIVE SUMMARY

This literature review and workshop were arranged in order to determine a possible research agenda for active case-finding for tuberculosis in South African mines. The review has focused on mass screening methods rather than other methods of active case finding. Over 400 articles were obtained as well as unpublished data from South African researchers. The articles were divided into categories according to themes and these were read by the research team members who then presented their findings at a 2-day workshop. The attendees at the workshop debated the relative merits of the available screening methods and came to a number of conclusions that have been incorporated into this review and discussion report. The literature review found that there was very little published research into methods of mass screening for tuberculosis (TB). There was a great deal of published research, however, into diagnostic tests that might be adapted for screening purposes, and this body of literature has been reviewed. Symptoms questionnaires have not been standardised nor formally evaluated. Chest X-rays were not, in general, recommended for mass screening purposes, although the possibility that an increased frequency of chest X-ray screening might lead to a reduced case fatality rate was believed to be worth further investigation. Skin tests were not useful for screening for active disease. Sputum smears might have potential as a screening method in spite of low sensitivity, but not as a mass screening method. Among the novel methods, those that were based on detection of three antigens specific to *Mycobacterium tuberculosis* showed most promise for the future. These tests are either in the form of an interferon-gamma production system or a patch test that is applied to the skin. The interferon-gamma systems are probably too impractical (and expensive) for mass screening purposes. The skin patch test, however, may prove to be useful. The patch test is currently undergoing rigorous evaluation in this country (as a possible definitive diagnostic test rather than as a screening test) and the results will be available at the end of 2005. It is recommended that the question of using the patch test as a screening tool be re-visited once those results are known. The main recommendation at this stage is that before screening methods are researched further, mines should work to improve the standard of delivery of the tried and proven methods for case detection. Once these are optimised it might be worth conducting research into new case-finding methods. Towards this end the recommendation is made that further research is carried out into the status of current practice in the mines and the nature of the decision-making and policy-making processes as far as TB is concerned. In addition, the methods whereby such decisions are implemented and the attitudes and practices of health care workers should be investigated. There are also two second level priorities. The first is that consideration be given to repeating the chest X-ray screening study, with randomisation of mines or shafts rather than randomisation of miners, in order to see whether the observed benefit of a lower case fatality rate can be duplicated. The other is that a study be conducted with a view to improving the quality of sputum bacteriology services in the mines.

INDEX

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This review of the literature has focused on active case finding methods for pulmonary tuberculosis with special reference to practice on South African mines. Although mass population screening for active tuberculosis is not recommended by the World Health Organization, the International Union Against Tuberculosis and Lung Disease, or the Chamber of Mines of South Africa, this review has been carried out from the perspective of mass screening rather than the screening of sub-groups that may be considered at increased risk, in order to investigate whether this position needs to be re-examined in the light of recent developments in TB diagnostic methods and the very high rates of tuberculosis on South African mines, especially the gold mines.

Screening of special groups such as room contact tracing has not been specifically addressed, although the application of screening methods to high risk groups will, logically, result in a higher yield of cases, and hence render the screening activities more economical. Lowe *et al*, for example, have shown that the application of a variety of screening methods (chest Xrays plus sputum microscopy and culture and perusal of serial weight records) to miners who were room contacts of smear positive patients is an efficient way of identifying new TB cases[1]. This review has had to be done in the context of the properties of diagnostic tests so that reports on the performance of the various methods when used as definitive diagnostic tests has also been incorporated into the review. The findings from the literature review were discussed at a 2-day workshop involving a number of role players and expert epidemiologists.

The results presented in this report incorporate the views and recommendations of workshop participants as well as evidence from published and unpublished scientific studies. The aim of the report is to tell a story, namely that while there may be several opportunities for useful further research into new tuberculosis diagnostic and screening methods on South African mines, these should not be the priority at present. The priority, instead, is to improve the management and cure of patients who are identified using the existing systems in place on South African mines.

The report begins with a brief background overview, followed by a description of the methods used. Thereafter, existing methods of diagnosis and screening are reviewed in turn. These are questionnaires about symptoms, skin testing, chest radiographs, and bacteriological examination of sputum used as a screening method. Possible novel methods of screening are then reviewed. Workshop commentaries and recommendations are integrated with the literature review throughout, and are clearly identified as such.

Finally, the findings of this review are discussed and recommendations are made concerning the prioritisation of future research into diagnosis and screening for tuberculosis on South African mines.

BACKGROUND TO THE REVIEW

The reported incidence rate of tuberculosis among South African miners has increased steadily over the last 20 years and this is illustrated, for gold miners, in the following graph (Professor Gavin Churchyard, various sources, personal communication):

Figure 1: The annual crude incidence rates for tuberculosis (all types) by year in South African gold miners, The South African population in general, and the United States of America.
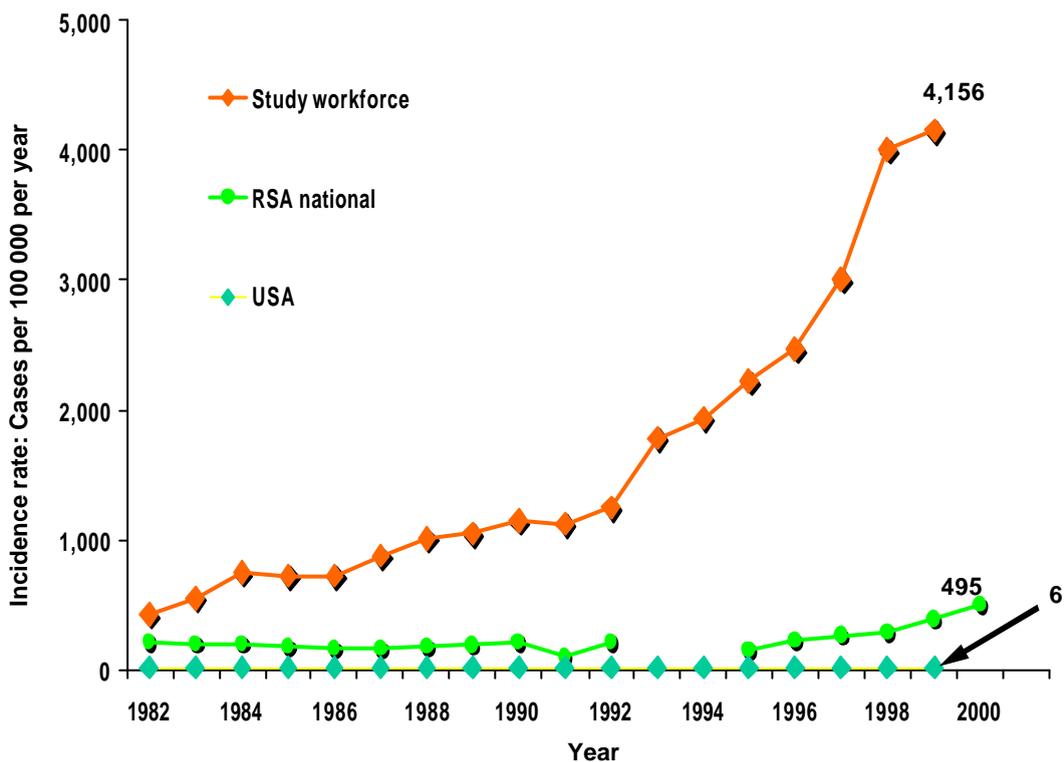
It is emphasised that the incidence rates in Figure 1 are crude incidence rates that are not age- or gender- specific. The South African national incidence rates would be expected to be lower since they use the entire population of South Africa as the denominator, including women and children. In South Africa the national age- and gender- specific reported rates are higher among men (an approximately 7-fold increased case-load among men versus women for the years 1990-2001) and among adults (approximately a 3-fold increase in the case numbers among adults versus non-adults for the period 1990-2001). Since the miners are adult males, direct comparisons with the national rates are not easy to interpret. Unfortunately the national TB reporting does not give the age- or gender- specific rates, and neither is the raw data available in the public domain that would make the calculation of such rates possible (the Department of Health merely presents the data as bar charts for the numbers of cases without any accompanying figures)[2].

Sonnenberg *et al* have reported on the change in the age-standardised incidence rate among male miners on 4 South African gold mines between the years 1991 to 1997, involving 53 296 person-years of follow-up. They found that the crude incidence rate increased by 165% among the HIV non-infected miners and 764% among the HIV-infected miners over this time period. The age-standardised rates increased by 108% and 129% respectively over the same time period[3]. These results indicate a compound rate of increase of the age-standardised incidence rates of approximately 10% per year over the study time period. The greatest increases in incidence rates for tuberculosis (TB) have been observed on those mines where the prevalences of the human immunodeficiency virus (HIV) are the highest and especially on gold mines where there is, in addition, a higher level of exposure to silica dust than in other types of mine. Lower incidence rates for TB have been reported from mines where the exposure to silica dust and the HIV prevalence rates are lower.

For example, in an audit carried out in a number of South African coal mines during 1999, the incidence rate for TB, as recorded in the TB registers, was found to be between 168 and 437 cases per 100 000 per year. These crude rates never exceeded the crude rates for the general surrounding populations, and in some case were 50% lower than in the surrounding populations. The reported cases were all cases of pulmonary TB (pTB) except for 2 cases[4;5].

Annual chest radiographs are routinely carried out on almost all of the mines in South Africa. In addition, there is a widespread practice of reminding miners of the symptoms of pTB at their annual re-induction medical examinations, and of encouraging them to present early for investigation should they develop any of these symptoms. Apart from a few studies that have been carried out examining the usefulness of alternative and/or additional screening practices, the annual chest radiograph represents the only widely used method of screening for active pTB on South African mines.

These prolonged and repeated increases in the annual incidence rates for pTB on South African mines have led to a re-examination of the strategies in place to bring the TB epidemic under control. One possible approach would be to shorten the time for which diseased miners are infectious, and thereby reduce transmission, through the adoption of additional or modified screening practices aimed at early detection of those with active disease. The rationale behind such a strategy is that cases of pTB that are detected by active case-finding, when compared to those detected by passive case-finding, are likely to have been infectious for a shorter period of time, are less likely to have symptoms of pTB, and hence are likely to have less advanced disease[6].

This literature review and workshop were commissioned by SIMRAC in order to review and discuss the evidence around screening for active tuberculosis.

METHODS

A Medline search was carried out for the period 1993 to 2002, using various combinations of the following key words: tuberculosis, diagnosis, screening, skin tests, chest x-rays, sputum examination, molecular methods, PCR. Well over 2000 articles were identified by this method and, after perusal of the titles, these were reduced to 455 articles of possible relevance. Abstracts were then requested for these articles. These abstracts were read and 390 were selected for probable relevance. Full text copies were obtained for these articles. The SIMRAC data-base was known to contain a number of articles in full text on CD-ROM that might be unpublished yet relevant. These CD-ROMs were obtained. The IUATLD (International Union Against Tuberculosis and Lung Disease) has published a number of reviews and guidelines over the last two decades.

These publications were used to cover review of earlier (pre-1992) work in the area. Appeals were published on the internet in two separate TB interest e-groups for any unpublished information. There were no responses to either of these invitations. The review was divided into several areas of common interest. The relevant articles were given to the individual researchers to summarise and critique, and they were asked to add any important and relevant sources that they may have, either published or not. The areas and researchers concerned were as follows:

Using questionnaires and weight loss:	Dr Jill Murray
The situation in South African mines:	Dr Jill Murray; Dr Michelle Wong
Skin testing and screening theory:	Prof Brendan Girdler-Brown
Economic evaluation:	Dr Seloacoe Thooe
Epidemiology:	Prof Gavin Churchyard
Sputum examination:	Dr Salome Charalambous
Chest X-rays:	Dr Peter Godfrey-Faussett
Novel methods:	Dr Karen Weyer

Dr Karen Weyer prepared the summary of the novel methods for screening. She also kindly provided a summary (presented in Appendix 4) of unpublished findings from her own evaluation of some proposed novel diagnostic tests. Based on these reviews, the researchers then prepared summaries that were presented at a two-day workshop. Participants in the workshop were identified in such a way as to include the known South African researchers in the field as well as operational experts from the mining industry and new researchers. Dr Peter Godfrey-Faussett from the London School of Hygiene and Tropical Medicine, was invited because of his recognised international stature as a TB epidemiologist with long experience in Africa as well as in the South African gold mines. A pre-workshop reading set was also prepared and sent to workshop participants in batches 1 month before the workshop so that the new researchers, especially, would come to the workshop with an understanding of the screening theory and terminology. Workshop participants were given a pre- and post-test to complete in which they were asked to indicate their views on the value of further

research into the various screening methods available. The returns were anonymous. Nick-names were used to link up the pre- and post- responses. The questionnaires are presented as Appendix 1. After each presentation, the workshop participants were invited to discuss the evidence and its relevance to the mining situation in South Africa. On the second day participants divided into groups where the evidence was discussed and possible priority research areas were identified. An adapted nominal group technique was then used, working with all the participants as a single group, to obtain a consensus about the research priorities. The principle investigator then collated the findings and wrote up the process. An explanation of the screening concepts and terminology used in this review is provided in Appendix 2, which contains the readings that were sent to workshop participants prior to the workshop.

TUBERCULOSIS CONTROL ON SOUTH AFRICAN MINES

Almost all South African mines follow a policy whereby all mine workers undergo annual chest X-ray screening for lung disease (including TB). In addition, annual "re-induction" medicals are performed in most mines, but often affecting only the mine workers and excluding contract workers. Although contractors (whose staff may make up as much as 30-50% of the people working in a mine) are often required to screen their employees for occupational diseases, as part of their contract with a mine, the level of adherence has not been reported in the literature. The Chamber of Mines of South Africa last reported on incidence rates for TB among miners in 1997, aggregated by type of mine (personal communication, Dr. Charles Mbekeni). The practice has since been discontinued. The last figures released in an aggregated form by the South African Chamber of Mines are summarised in the next diagram, prepared by Professor G. J. Churchyard for the Mount Grace TB Screening workshop held as part of this review.
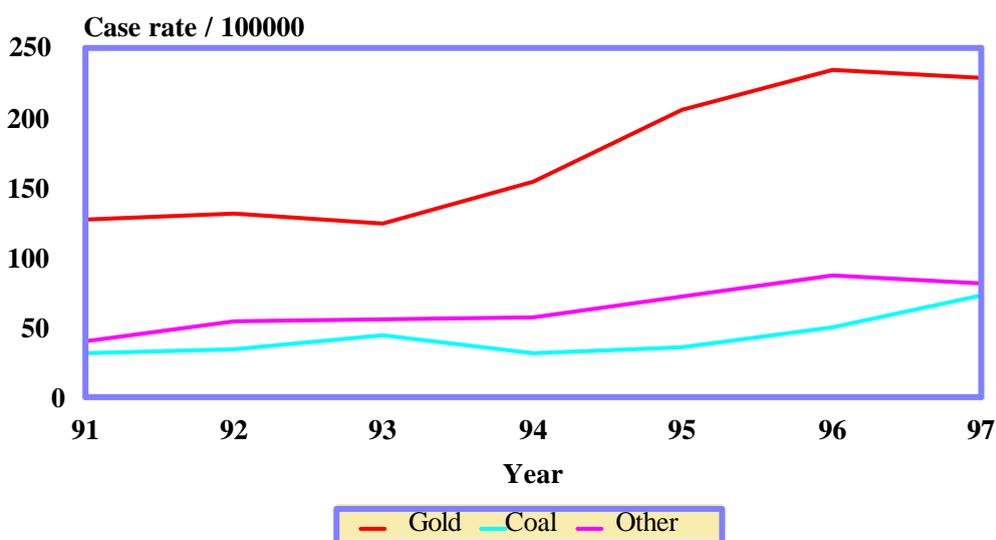

Figure 2: TB case rates on gold, coal and other (largely platinum) mines
(Chamber of Mines 1997)

In South Africa, the law requires that cardiorespiratory organs are examined, during legally mandated autopsy, on all miners and ex-miners who die, provided that the family agrees. The results of these autopsies are available in a unique database kept at the National Institute for Occupational Health (NIOH). Their report for the data from 2002, show that 20.7% of the autopsies carried out during that year revealed active pTB. This figure is probably an underestimate since, although the autopsies are mandated by law, 30% were carried out on white ex-miners who make up a much lower proportion than this of the workforce, and who are thought to have a lower risk for pTB. In fact, the autopsy rate among black gold miners was reported as 27.47%[7].

This implies that deaths of black ex-miners are under-represented. In fact, the prevalence of active pTB among black miners seen at autopsy was reported as 28.9%. Murray *et al* have in addition, reported on autopsies on a series of 1858 miners from all types of mine who died while in employment over a 1 year period[8]. They found that 250 of these miners had active TB, or were on treatment for definite TB, at the time of death, and 155 (62%) of these had not been diagnosed as having TB prior to their deaths. An additional 100 deceased miners were on TB treatment at the time of death although there was no convincing evidence, on examining their case notes and laboratory data, that they in fact had TB. On reviewing the case notes it was discovered that many of the missed cases of TB had not been adequately investigated for TB in spite of complaining of classical TB symptoms such as cough, weight loss and night sweats.

Furthermore, Churchyard *et al*[9] have reported that sputum-based measurements of incidence and prevalence rates among a group of gold miners indicate the mean duration of disease to be over 1.79 years in those HIV non-infected miners with pTB and 0.61 years in those who were HIV-infected. This suggests that there may be a considerable delay before TB is diagnosed and treated, especially in those patients who are not HIV-infected. This offers an opportunity for screening to be effective, and also an opportunity for improving the early detection of TB by improving the routinely available diagnostic services.

There is a national South African Tuberculosis Control Programme that is based on best practice recommendations of the IUATLD and the World Health Organization (WHO)[10]. In addition, the South African Department of Minerals and Energy (DME) has produced a guideline for practice on the mines that meets all the requirements of the national guidelines and, in some respects, exceeds the national guideline recommendations[11]. There has been no mining industry-wide study to determine how well mines are adhering to the DME's or the national guidelines. A postal survey carried out in 1998 managed to elicit only a 10% response, and the results of that survey were thus never published. However, the survey revealed that compliance with national policy might be extremely variable, with 1 respondent even stating that they "never" notified cases of TB (author, personal recollection). A number of smaller mines approached stated that they used part-time local general practitioners (GPs) to diagnose and treat TB patients, depending on which GPs were available to provide the

service as and when needed. As a result they were uncertain as to whom to ask to complete the questionnaire. None of these mines had a regular performance review with regard to their management of TB cases.

Girdler-Brown and Ross[4;5] conducted a practice audit for a South African coal mining corporation in 1998/1999 involving 7 different collieries. The results of that audit were very encouraging, and showed a good standard of management in the mines concerned. However, it is possible that mines that are unsure of their standards of practice might be reluctant to go through a similar process, so that those findings may be non-representative of the general situation. In the gold mining sector, where, due to the exposure to silica dust, the TB incidence rates are higher, there has been much published research on TB over the last decade. In addition, some large corporations have invited international experts to audit their practice and have fared well in these audits (Prof. Gavin Churchyard, Dr. Jill Murray, personal communications).

The fact remains, however, that it is not possible to make any scientifically valid general statement concerning the level of practice or performance in the industry as a whole. Although, in general, the overall increase in the incidence rates of TB over the last decade can be explained by the rising prevalence of HIV infection, it remains unexplained why some mines have reported much higher crude incidence rates than others with the same or similar HIV prevalence rates and the same or similar silica dust exposures. There is also some anecdotal as well as objective evidence of poor standards of practice in some mines and/or corporations.

The crude incidence rates of TB have been rising in the South African mining industry over the last decade. This conclusion is based on actual studies[12], reports received by the Chamber of Mines up to 1997 and on serial autopsy reports[13]. The incidence rate data referred to also demonstrates that the rates are highest, and rising most steeply, among gold miners. In addition, the proportion of patients with recurrent pTB has been increasing, at least in some settings[12]. The autopsy reports show that the proportion of those examined with active TB rose from 0.9% to 3.9% between 1976 and 1991. This trend remains evident after correcting for increasing age and length of service, and is statistically significant. Kleinschmidt and Churchyard have, further, shown that within the gold mining industry, rates vary by occupational type and are higher in those groups where exposure to silica dust is greatest[14].

Exposure to silica dust[15;16] and the prevalence rate of HIV infection[9] are the two factors that are commonly cited in order to explain the high and increasing crude rates of TB among miners, especially gold miners, and the evidence of increased risks for tuberculosis following these two kinds of exposure is now very strong. The influence on the risk rates is multiplicative, so that a relative risk of 3 for developing TB if exposed to HIV and a relative risk of 2.5 for developing TB if exposed to silica, results in a relative risk of 7.5 (and not 5.5) if exposed to both factors. However, it should be remembered that these conclusions have been reached on the basis of crude incidence rates that have not always been adjusted for

age: the mining workforce is predominantly male, and although this is true for the entire period for which the dramatic increase in incidence rates has been observed, there is evidence that the average age of mine workers has increased in the same period[13].

The study by Churchyard *et al*[8] is thus far the only recently reported study in which the duration of infectiousness has been estimated in South African miners. In that study, as already mentioned, the average duration of disease was estimated at 7.32 months for those who are HIV-infected and 21.5 months for those who were HIV non-infected.

This difference is explicable on the basis that those who are HIV-infected are likely to become symptomatic, and to seek medical care, more quickly following the onset of disease. Furthermore, there is now a large body of evidence showing that TB is more likely to be either extra-pulmonary or smear negative if pulmonary, in those who are HIV-infected[17]. Thus, although the HIV epidemic has resulted in dramatically increased *incidence rates* of TB, it is likely that the contribution of the HIV epidemic to the *spread* of TB may be less than at first feared. Having said this, however, the recent paper by Sonnenberg *et al*[3] has shown that, even among those miners who are HIV non-infected, there has been a compounded growth rate in the age-adjusted incidence rate of approximately 10% per year between 1991 and 1997, and this cannot be explained by increasing exposure to silica dust, since age-standardisation will to some extent have adjusted for such exposure.

There have been a number of recent reports from developed countries[18-20] with relatively low TB incidence rates indicating that approximately 40% of cases of pTB are epidemiologically linked to other recent cases. Although the proportion is sometimes higher among those who are HIV-infected, the proportions in those who are HIV non-infected are higher than has been previously thought, and, in fact, Behr *et al*[21] have found that HIV infection status did not affect the proportion (39% in their study) of patients who were ill with recent infections rather than reactivation of long-standing infections. Other researchers have found somewhat lower rates of clustering (20%) in communities in developed countries[22].

Godfrey-Faussett *et al* have reported on a prospective cohort study of TB in South African gold miners using DNA fingerprints[23]. They showed that, in their cohort, over 50% of cases were likely due to recent transmission. They also showed that although 12% of this transmission was from index cases living in the same room, the majority of the transmission was between those who did not share the same room, and, especially, from index cases who were "chronic" cases. A further report involving 326 cured patients followed up for a mean of approximately 2 years, found that of 39 cases of repeat episodes of pTB in whom paired DNA fingerprints were available, 14 had the repeat episode with a new strain of *M. tuberculosis*[24].

However, the proportion of patients differed when they were grouped according to HIV infection status. In that study, 13/21 of the HIV-infected patients had a repeat episode with a new strain (implying that the infection that led to the disease was recently acquired) as opposed to only 1/18 of those who were HIV non-infected.

9

Ten Asbroek *et al*[25] have studied the serial interval (the time elapsing between the onset of illness in an index case and its secondary case) using DNA fingerprinting in the Netherlands. They studied 233 such primary-secondary pairs. The distribution of the serial intervals was markedly positively skewed (range 0-130 weeks) with a geometric mean of 27.5 weeks for pTB pairs. The serial intervals for those who were HIV positive were far shorter (8, 12 and 23 weeks).

These findings, that a fairly high proportion of active TB disease is caused by recent infections rather than reactivation of latent infections, were predicted by the results of the Prophit survey which was carried out on a cohort of 20000 young English adults between 1936 and 1944. In that study it was found that the rates of tuberculosis disease among those who were already skin test positive at entry to the study were double in those groups with high exposure compared to those with low exposure to cases of active pTB[26].

USING SYMPTOMS QUESTIONNAIRES AS SCREENING TOOLS

A large number of studies reported from around the world have examined the associations between symptoms and TB disease. Most of these studies report on the prevalences of symptoms among those detected by passive case finding. Ward *et al* have recently published data from aboriginal patients in Saskatchewan in which they compared symptoms rates among those detected by active and passive case finding.

Their findings are fairly typical of what has been reported elsewhere and are presented below (patients often had more than 1 symptom each, so the percentages add up to more than 100%)[27]:

SYMPTOM	% of all those with disease	
	Active	Passive
N	450	453
Daily cough for more than 1 month	59%	84%
Unexplained fever for more than 1 week	19%	47%
Weight loss	9%	30%
Night sweats	9%	17%
Blood in the sputum	1%	12%
Smear positive sputum	4%	37%

Table I: Symptoms percentages among TB cases detected by active and passive methods

Of note is that while 36% of those detected by active case finding had no symptoms, the percentage of those detected by passive case finding with no symptoms was less than 1%. In another survey of patients in Vietnam, it was shown that male patients were significantly more likely to admit to cough, sputum expectoration, fever, tiredness, anorexia, chest pain, haemoptysis and headache, but not weight loss, than female patients.

This suggests that, if the same is true in South Africa, symptoms screening may be more productive in the mines setting, where the vast majority of subjects are male[28].

A report from the Sudan[29] shows that the vast majority of symptomatic patients will be detected passively within 2-4 months of the onset of their symptoms, so that it is unlikely that symptom screening performed only once a year will detect a large proportion of patients who would not have self-presented.

Although the use of symptoms has been shown to be of little use to discriminate between patients with pTB and those with other respiratory disease[30], since pTB is likely to be the only common prevalent respiratory disease present in those being screened for pTB, this fact is not of concern in the screening context. Other studies have specifically looked at the prevalence of symptoms among those who are sputum smear negative. Since any screening programme will result in the identification of predominantly smear negative patients, their findings are of interest here.

In Guinea[31], 396 adult TB suspects were examined who had cough for more than 3 weeks and who had all had 3 negative sputum smears. 110 of these patients were subsequently found to have positive sputum cultures for *M. tuberculosis*. Those with smear negative pTB were more likely than those who did not have pTB to admit to symptoms of sputum production, a familial TB contact, and weight loss. In a different study[32], this one in Senegal, and involving 450 patients with pTB, in whom 85 were smear negative, cough and sputum production were significantly less common in the smear negative group than in the smear positive group.

Although 9% of the patients were HIV-infected, the symptoms differences were not reported by HIV infection status. In the pre-HIV era, up to 51% to 54% of smear negative pTB patients have been reported to have one or more symptoms[33;34].

The prevalence of symptoms may be different among pTB patients who are also HIV-infected. A study from Italy, for example, reports that 94% of HIV-infected TB patients had fever, 40% had a productive cough, and 55% had night sweats[35].

Marciniuk *et al*[36] have shown, in Saskatchewan, that screening people by means of symptoms questionnaires is becoming more efficient as a means of case finding in an era when, due to the increasing prevalence of HIV infection, the yield from chest X-rays is declining[36].

The study by Churchyard *et al*[9] has reported on the prevalence of different types of symptom among South African gold miners who were screened primarily through microbiological

examination of sputum. The situation, particularly regarding cough as a symptom, may be expected to vary between types of mine and even between mines of the same type, depending on the dust exposure (quantity and composition). Churchyard *et al*'s findings on one gold mine were as follows:

SYMPTOM	% of all those screened (n = 1960)
Cough	2.50
Night sweats	6.20
Weight loss	5.40
Sputum production	1.30
Blood in the sputum	0.70
Cough more than 3 weeks	1.20
Fever more than 3 weeks	0.80
Symptom combination*	10.10

*Any one or more of cough, weight loss and night sweats

Table II: Symptoms prevalences for gold miners screened for pTB

Forty eight cases of pTB were eventually identified, mainly on the basis of bacteriological diagnosis (microscopy and/or culture). Sensitivities were all low, with the best sensitivity being 29% for the symptom combination. Specificities, on the other hand were all in the order of 90%. This is a potentially useful result since the sensitivity for the symptom combination was slightly higher than for either smear microscopy or routine chest x-ray alone. The high specificity means that there are not too many false positives that would then have to be subjected to definitive diagnostic testing with negative results.

Of note is that the combination of the results of routine chest x-ray and the symptom combination resulted in a sensitivity of 60%, which is double the sensitivity for x-ray examination, symptoms combination or smear microscopy alone.

Other studies, from other parts of the world, have reported much higher percentages of symptoms among *patients* with pTB, presumably because these are mainly self-presenting patients and more likely to have symptoms. Toman has pointed out that over 90% of smear positive pTB patients will have symptoms in carefully designed studies, and there was no difference in the rate of symptom reporting in developed and underdeveloped countries. Approximately 70%+ of patients will admit to a cough and a further 20% will admit to fever or flu-like symptoms[34]. In this same vein, other authors have reported relatively high reporting levels for symptoms among *patients* with disease, whether smear positive or smear negative, and whether or not they are co-infected with HIV[35;37].

Some reasons why studies may result in different percentages of symptoms reporting include poor history taking, lack of understanding by the patient, and that they may be using different definitions of the symptoms.

For example, in the study by Churchyard *et al*, the definition of cough was "a new or worsening cough in the last month". Other definitions might differentiate between dry or productive coughs, coughs of different durations, cough on waking in the morning, etc… Furthermore, one needs to be sure of the way in which the question is understood, especially if it is translated, since in some cultural settings hawking or spitting may be understood to be "coughing" as well. It is of interest that only 2.5% of those questioned in Churchyard *et al*'s screening study admitted to a cough, and that the specificity of this symptom was over 95%, since the subjects are underground miners exposed to dust pollution, and many would also be smokers. Asking about symptoms in an empathetic way may be a cheap and feasible method of screening miners for further investigations.

Other symptoms such as coughing up of sputum containing blood are suggestive, but not particularly common among those with active TB, so that many patients will deny such a symptom. In the study by Churchyard *et al* the sensitivity of this symptom was less than 5%.

Since it may be argued that the longer a patient has untreated pTB the more likely that the patient will suffer more severe permanent damage to the lungs, and the greater the spread of the infection to others, there is a need to detect patients with disease as early as possible, and to place them on effective anti-tuberculosis treatment. This applies particularly to those HIV non-infected patients who are smear positive since they are more likely to contribute to *M. tuberculosis* transmission. There have been several reported studies in which researchers have attempted to estimate the average duration of symptoms prior to starting treatment, and to identify the causes of delays in the diagnosis and treatment so that these can be targeted for reduction[30;38-46]. These published studies commonly give values of between 2 and 5 months for the duration of symptoms prior to starting treatment. Delays are sometimes ascribed to late presentation (which may be due to poor access) and sometimes up to 2 months in diagnostic delays within the health system. The service delay components in these studies were frequently due to prolonged time taken in obtaining sputum results and other factors, such as consulting of private doctors that may not be relevant for South African miners.

In the single published study of this nature that was found that presented South African findings, it was determined that, for rural patients from a community in the Limpopo province the median delay to presentation was approximately 10 weeks which included a median service provider delay of 1 week[47]. These delays are considerably less than the 21.5 months reported by Churchyard *et al* for gold miners, partly because in these studies delay was defined as the interval between the onset of symptoms and the initiation of treatment. In the gold miners study the duration of infectiousness was estimated and this would include an additional time during which many patients would be asymptomatic.

In the study reported by Churchyard *et al* [9] the duration of active disease (defined as the presence of *M. tuberculosis* in the sputum, detected by microscopy and/or culture) was estimated rather than the duration of symptoms. This approach is interesting for several reasons: Firstly, it has been shown that smear negative patients are also infectious to others, and secondly it is possible that severely ill patients with positive sputum smears would be more likely to seek medical attention in the mining situation where there is easy access to diagnostic and treatment services. That study showed durations of 0.61 years for HIV-infected patients and 1.79 years for those who were HIV non-infected. From an epidemic control point of view the authors of that report concluded that HIV non-infected patients were more likely to be playing the dominant role in the transmission of *M. tuberculosis* in the population that they studied. This is fortunate from a screening perspective, since the longer that a diseased person remains undiagnosed, the greater the proportion of cases that will be detected by the screening measure. It can also be argued that the long duration of active disease observed in the HIV non-infected group, occurring as it did in a well-equipped mine, indicates a failure of the service to detect disease early.

One needs to be cautious, however. Even if the specificity of smear microscopy examination is 99%, if the entire workforce is screened (in this case subjects were asked to give saliva if they could not produce sputum), 1% of the non-tuberculous workforce might be expected to provide false positive specimens. If the actual prevalence of active disease is 2% then this means that 1/3 of those who test positive will not actually have active tuberculosis. By including all those with positive smears as cases of active tuberculosis, it is possible that the duration of infectious disease will be over-estimated.

In addition, there are no bench marks with which to compare durations estimated in this way. The other published durations refer to the durations of symptoms, and it is possible that many of the cases of active disease identified in the study of Churchyard *et al* might have been asymptomatic. Furthermore, that study did not attempt to identify the causes of the delays, so that the delay cannot be assumed to be due to failures within the service. Nevertheless, the high proportion of TB patients who have one or more of the symptoms of TB suggests that screening by means of asking patients about the presence of symptoms might offer a cheap and effective way of screening people for TB. The literature search did not produce any formal evaluations of using questionnaires regarding symptoms as a screening tool for TB. However, there is some anecdotal evidence available from coal mines in South Africa. In their TB audit of 7 coal mines Girdler-Brown and Ross[4;5] specifically asked the health staff at the 7 mines that were visited about the use of symptoms questionnaires. They elicited two reports of such practices.

In one mine a nurse with experience of working with TB patients was trained in interviewing techniques and was employed full-time to go about the mine interviewing miners at their workplace as well as in the recreational facilities and refectories. He carried a clip-board for entering details of miners who he thought should be investigated further. The mine had

detected 50% of their registered TB patients *via* this referral route. The impact on the duration of symptoms was not assessed, and it is possible that all those identified would have come to the clinic (as they were symptomatic) of their own volition even without being encouraged to do so by the interviewer. The interviews were informal, the findings were not recorded, and there were no controls, so that the intervention could not be evaluated. In the second mine, a simple self-administered questionnaire was used at the annual re-induction medical since the average standard of education is relatively high in the highly mechanised South African coal mining industry. Miners were asked to tick any of the 3 classic symptoms if they had them, and anyone with one or more was called for a medical examination and possible further investigations. The auditors were told that this method had been abandoned after 1 year because there had been virtually no new cases identified as a result. Once again there were no records available and there were no controls so the intervention could not be evaluated.

There are a few comments that need to be made at this stage. Firstly the mines were coal mines with relatively low incidence rates for TB, so that the yield of any screening programme would be expected to be relatively low. Secondly, the interview method was continuous whereas the questionnaire method was intermittent, and annual. Depending on the duration of symptoms, the questionnaire method might therefore be expected to identify only a small proportion of cases (those who happen to be symptomatic at the time of the annual questionnaire). Thirdly, the self-administered questionnaire might be less sensitive *ceteris paribus*, than the interview method, since it would be easier for miners to underplay their symptoms if they were not being observed. They might do this if they believed that their symptoms were unimportant and they wished to avoid unnecessary trouble, or if they thought that the presence of symptoms might prejudice their chances of securing a re-appointment.

Although the literature search also looked for articles that reported on the use of serial weights as a method of screening for TB, only one such article was found and it involved short-term repeat weighing in HIV-infected individuals. As a result it was not possible to evaluate such a screening strategy. However, in the era of high HIV prevalence it is likely that such a strategy might lack specificity.

No references were found for the costs of using symptoms questionnaires for screening for pTB. In addition, there was no opportunity for meta-analysis because the settings reported on were too dissimilar. Workshop participants felt that the use of symptoms as a screening tool might be worth exploring in the future, but that the first priority must be to improve case detection methods among those who self-report with symptoms.

SKIN TESTS AS SCREENING TOOLS

Skin testing has been used to identify people who have been *infected* with *M. tuberculosis*. It has also sometimes been used to help in the diagnosis of TB, especially in children, in very low prevalence situations where infection with *M. tuberculosis* is uncommon.

There are three broad groups of skin tests that are either in use or being developed. The first is the Mantoux test using tuberculins. Tuberculins are extracts of tubercle bacteria, and usually contain multiple antigens.

Where an HIV non-infected patient has active tuberculosis, the commonly used Mantoux skin tests using tuberculins have a high sensitivity (few false negatives). However, where there is a high prevalence of prior exposure to *M. tuberculosis* or related organisms in the environment the specificity for diagnosing active disease is low (a large proportion of false positives). Among South African adults, there is both a high prevalence of HIV infection and a high prevalence of prior exposure to *M. tuberculosis*, so that the test has both a low sensitivity and a low specificity for diagnosing active disease. As a result the workshop participants felt that the method should not be given priority attention as a screening tool for identification of those likely to have active disease.

Nevertheless the literature contains a great number of recent and past publications on the subject of skin testing as a screening tool. The reason is that, in very low prevalence countries, where prophylactic treatment with isoniazid is practised on all people who are infected, even if not diseased, skin testing (as well as chest x-ray screening, and sometimes both) is used extensively to screen high risk groups in order to identify people who are infected in order to provide prophylactic treatment. There is some evidence of a reduction in the annual risk of infection (from 7% to 4%) over the period 1972 to 1977 in a rural area of the Transkei[48]. If this trend has continued it is possible that the rate of positive Mantoux skin reactions among adults may now be considerably lower than in the past. A more recent survey, carried out in Lesotho and Botswana, two traditional recruiting areas for the mines in South Africa, showed that, in 1986, the Mantoux positive rate for skin test reactions of 10mm and more in diameter was 51.5% in Botswana and 57.9% in Lesotho (% of all those tested over the age of 14)[49].

A recent study carried out among adult health care workers at a South African TB referral hospital showed that the prevalence of a positive Mantoux tuberculin skin test (>10mm diameter) was 64.6%. These workers had an average age of 43.43 years and an average length of service of 13.06 years. The rate of positive skin tests among low exposure staff such as stores and switchboard personnel was 52.19%[50]. Furthermore, there is anecdotal evidence (participants in the Mount Grace workshop, personal communication) that, within a year or so of starting to work in the mines, over 85% of miners convert to skin test positive (it appears there have been some sporadic unpublished surveys in some of the mines). However, the epidemic of HIV that has intervened would make the test less useful, even if the background rate of skin test positivity has fallen in the interim. It has been shown that even if one were to include the "density" of the skin reaction with the diameter as diagnostic criteria there would be little or no improvement in the diagnostic properties of the test[51].

One strategy that has been used to try and improve the sensitivity is two-step testing where subjects are tested twice, relying on a booster effect to detect additional cases in the second

round: this, however, has only resulted in a less than 10% improvement of sensitivity, and the published studies have not involved high HIV-prevalence settings[52-55]. Another strategy is to combine the skin test with anergy skin testing tailored to ensure that the anergy test includes locally prevalent antigens (such as pertussis). This is sometimes referred to as "Combi-testing"[56;57]. This might improve sensitivity, but would not address the local problem of low specificity.

The second group of skin tests, the Mantoux test using sensitins, or mixed antigens derived from non-tuberculous mycobacteria, is used in epidemiological work only, where its purpose is to determine what level of tuberculin Mantoux reactions are probably caused by exposure to these non-tuberculous mycobacteria.

Recently, there has been increasing interest in developing new skin tests (the third of the three groups) that are reactive only in cases of exposure to *M. tuberculosis* (existing tests might react following exposure to other Mycobacterial species found in the environment)[58]. This would be of use, say, when screening immigrants to the United States of America who come from areas of the world where such environmental bacteria are ubiquitous, for *M. tuberculosis* infection, and would prevent unnecessarily placing some migrants on 6 months of prophylactic therapy. However, these tests would still not differentiate between past exposure and current disease, and neither would they overcome the problem of anergy in populations with high HIV infection rates.

There has been a recent report of a study in which BCG (Bacille Calmette Guérin) vaccine was given to TB suspects in an attempt to diagnose active disease, referred to as the "BCG test": if there was a marked reaction this was said to reflect active disease[59]. However, there were marked side effects in the form of tuberculous adenitis in up to 10% of patients, the diagnosis of TB used as a gold standard was not based on smear examination or culture of sputum, and the control group was inappropriate, so that the high sensitivity and specificity that was claimed cannot be relied upon, quite apart from ethical considerations bearing in mind the high rate of untoward side effects.

There were no reports on costs of skin testing programmes, and meta-analysis was not considered appropriate since there are so few studies with similar settings (as pointed out, the sensitivity and specificity will vary from community to community).

Workshop participants felt that there were no useful further studies indicated on the use of skin testing in South African mines.

CHEST X-RAYS AS SCREENING TOOLS

Chest X-rays have been investigated both from the perspective of use as a definitive diagnostic tool and from the perspective of screening large populations to identify those

requiring further investigation. Since as long ago as 1974, the use of X-rays for mass population screening has been discouraged by the World Health Organization[60].

Chest X-rays have a poor sensitivity and specificity for the diagnosis of pTB[34;61-63]. In a South African study the sensitivity on a gold mine was found to be only 27%[9]. The reading of chest X-rays for pTB also has poor reproducibility[34;64;65], although this improves to fair reproducibility as judged by kappa scores, for lymphadenopathy in children[66]. The use of computerised tomography does not appear to offer any useful improvement[67]. The experience of the Tuberculosis Research Institute in South Africa, which carried out 6 prevalence surveys, showed that the positive predictive value of chest X-rays for identifying those with culture positive pTB was only 13.8%[68]. These results are similar to results reported from Bangalore[69] in which, in spite of chest X-rays having a sensitivity of 87.7% and a specificity of 95.9%, when all those with abnormal chest X-rays were subjected to sputum smear examination the yield was only 10%. This low positive predictive value was found in a group that might be considered to be at high risk, since they were all symptomatic, but in whom the prevalence was only 13% to 14%.

The proportion of pTB patients with chest X-ray abnormalities has also been reported to be lower in patients who are infected with HIV[35;70], although Kanaya et al[37] found that 96% of their pTB patients had some or other abnormality on chest X-ray in a series in which 46% of the patients were HIV-infected in California. Finally, chest X-ray abnormalities have been found to be poorly correlated with the duration of disease: pTB patients are just as likely to develop X-ray cavitation early in the course of disease, so it is not a good method for detecting infectious patients earlier[34]. In fact, as Toman points out, even intensive screening programmes with high levels of participation from the public, have failed to result, over the years, in a reduction of the proportion of newly detected cases that are smear positive.

On South African mines, the proportion of pTB patients that are first detected by the routine annual chest X-ray surveillance programme has been as high as 77% in the past but has fallen to less than 50% in recent years[12]. This may be ascribed to the rising proportion of pTB that is HIV related: Churchyard et al have shown that HIV-infected cases are more likely to present earlier (after an average of 0.61 years), and are thus less likely to be detected by the annual chest X-rays[9]. The high proportion of cases detected by screening does not mean that these patients would not have been detected in the absence of the screening programme; it could be that cases are being detected through passive self-presentation relatively late in their disease.

Furthermore, it has been shown that with serial mass population screening by means of chest X-ray there was no difference in the interval since having a normal chest X-ray and the onset of disease between those who were smear positive and those who were smear negative, suggesting that serial chest X-rays may not shift the pattern of disease towards a greater proportion being smear negative and hence less infectious[33]. In that study, however, carried

out in Czechoslovakia, the interval between the serial X-rays was variable and often as long as 3 years.

If the serial X-rays are carried out at intervals that are less than the average duration of infectiousness then the result might be different. Churchyard *et al* have reported on such a study, in which annual chest X-rays were compared with chest X-rays repeated every 6 months in a randomised controlled trial involving 22000 miners on a gold mine with a high incidence of pTB[71]. They found that there was no difference in the annual incidence rate after 1 year. However, they randomised individuals in the same mining population rather than randomising communities, so any effect in the intervention arm would have been diluted by exposure to infectious individuals from the non-intervention arm of the study. As a result the findings are inconclusive. They did, however, report a lower death rate while on treatment among those in the 6 monthly arm of the study, suggesting that disease might have been detected at an earlier, and more curable stage.

In another study reported by Churchyard *et al*[9] the combination of an abnormal chest X-ray finding with one or more abnormal symptoms was found to have a sensitivity of 60%.

There were no reports found on the costs of chest X-ray screening programmes. In addition, meta-analysis of papers that presented findings for sensitivity and specificity were not attempted because these parameters are setting-specific and there were too few studies from the South African mining setting.

Workshop participants felt that the findings (especially of a lower mortality rate among those who were screened twice a year as opposed to those who were screened once a year) were potentially of interest and warranted some further research to see if they are repeatable.

SPUTUM EXAMINATION AS A SCREENING TOOL

Sputum examination forms the cornerstone of modern recommended definitive diagnostic methods. In the first instance, sputum smear examination of 2 or 3 consecutive sputum specimens is recommended for the diagnosis of (by definition) "smear positive" pTB[10;72;73]. This is because patients with smear positive pTB are considered to be more infectious and hence the principal sources of transmission in a community.

In addition, smear positive patients are believed to be the ones who are at higher risk of progressive lung damage if not treated. It has been shown that the grade of the sputum smear result correlates with the severity and number of symptoms[29]. Finally, the diagnostic test is affordable and relatively easy to perform in resource-poor countries[74], and the incidence of pTB is higher in these countries than in those where more expensive, more sensitive tests might be considered appropriate.

In the setting of South African mines, however, lack of resources is not an important issue and the Department of Minerals and Energy recommends, in its guidance note, that sputum

cultures, in addition to smears and chest X-rays, should be performed for diagnostic purposes for all suspected cases[11].

The sensitivity of the sputum smear examination for cases of pTB is poor[75] and variously estimated at between 20% and 70% [69], depending on the prevalence of HIV infection, how early the patients are investigated (i.e. on the average duration of symptoms or the average duration of infectiousness), how many smears are performed, how many microscopic fields are examined per smear, on the staining technique used, and on the quality of the laboratory services. However there is little additional gain in sensitivity if 3 or more smears are examined per patient[34;76;77]. Furthermore, examination of only 100 microscopic fields (versus 300) reduces the workload by two thirds with very little sacrifice of yield[78].

The prevalence of HIV may also affect the results of microbiological examination of sputum in pTB patients. A Ugandan study, for example, has shown that HIV-infected pTB patients are less likely to be smear positive, and less likely to be culture positive. They also showed that where smears were positive the numbers of bacilli seen per high-power field were lower than for positive smears of patients who were HIV non-infected[70]. A South African study has shown similar results[79].

Quality issues in the taking of the sputum specimens, their storage and transport, and on the preparation of smears and their reading, have been extensively investigated and reported on. Whether or not sensitivity can be improved by induction of sputum appears to be somewhat controversial. A study from Malawi reports that, where induction was used on those patients who were unable to produce a specimen without the procedure, 18% of smear positive cases would have been missed if the intervention had not been used[80]. In a study reported from China almost 1/3 of TB suspects with poor or inadequate sputum production were found, after induction, to be smear positive[81]. The estimated direct cost of the procedure (in China) was found to be $0.37 (1999 prices).

However, in a study carried out in the United States of America, with a 36% HIV prevalence among the study group, induction of sputum yielded only one additional smear positive case (out of 11 smear positive patients with 2 sputum specimens obtained by expectoration and 1 by induction). The total direct cost per induction was estimated at $28.65. Roughly half of this cost was for a physiotherapist's time and the other half was for supplies[82]. It is likely that the poor yield is due to the fact that induction was performed only on patients who could produce a spontaneously expectorated specimen for comparison: one would then not expect to detect many additional cases from such an induction strategy. The high prevalence of HIV infection may also have affected the yield, since two patients who were smear positive with spontaneously expectorated specimens were then smear negative on their induced specimens. In addition the sample size (11) was very small.

The American costs are higher than those reported from China, and this may be partly due to higher salaries in the United States. It is also unclear from the American report whether the cost of "supplies" is an incremental cost for patients who would not normally be able to

produce an adequate specimen. If inductions are reserved for those without a sputum specimen available from spontaneous expectoration, then there is a "saving" of the supplies that would have been used for the expectoration specimen, and these need to be deducted from the cost of supplies used in the induction procedure to give the incremental cost of the procedure.

The traditional staining technique used in smear examination is the Ziehl-Neelsen technique. The use of fluorescence microscopy, although more costly, has been shown to be somewhat more sensitive, especially in paucibacillary smears, and also less labour-intensive, since this method is more rapidly performed[83].

There have been numerous published reports on quality of sputum examination from developing countries. First of all, compliance of health workers with sputum collection policy is sometimes poor[84;85]. Secondly, there may be long delays before sputum results are available at the clinics[85].

In the laboratories themselves, there may be few quality assurance protocols, or they may not be adhered to. Liquefaction of the sputum prior to smear preparation may be sub-optimal[86], smears may fade with time[87], especially if stored under conditions of high temperature and/or humidity[88], tubercle bacilli may be dislodged from the specimen during staining[89], and technicians may not examine a sufficient number of fields. Problems may be especially serious in private laboratories[90]. The quality assurance step that is often used whereby laboratory staff are asked to report on "standard" smears sent to the laboratory periodically merely certifies competence and may not correlate well with the routine practice standards[91].

It is reported that false negative results outnumber false positive results[92]. It is recommended that laboratory staff undergo regular training and support from the national TB control programme even if the laboratory is privately owned. It has also been recommended that re-checking of smears, especially if reported as negative, should be routine[93]. The routine use of quality check lists has been advocated[94]. Finally, because the number of smears to be examined in high prevalence settings may be high, the management systems in place may need to be quite sophisticated in order to efficiently manage the workload, reduce costs, and support the maintenance of a high quality service[95].

In one reported study from a South African gold mine with an HIV prevalence of 29%[9], the sensitivity of smear examination was found to be only 25%. This is lower than the sensitivities that are reported from the pre-HIV era, and this could well be explained by the high prevalence of HIV infection, in keeping with the findings from the Uganda study. However, in that study there were stringent quality control measures in place and a number of discrepancies were reported between the study results and the results obtained in the reference laboratory used for quality control. The sensitivity for infectious cases is much higher. For example, if only those cases that are smear positive are considered infectious, then, by definition, the sensitivity for detecting infectious cases will be 100%.

Specificity of smear examination is generally high, and in excess of 98%[69], since the number of false positives is usually expected to be very low. Nevertheless, with a sensitivity of only 25% and a specificity of 98%, if the prevalence of disease among those being tested is only 2%, the positive predictive value of this test, when used as a mass screening test, is only 16.8%! Of course, it is unlikely that sputum specimens would be obtainable from 100% of the miners, and it is more usual to find that smear examination is reported as having a sensitivity of over 50%, so this scenario over-exaggerates the situation. As expected, the HIV epidemic may reduce the sensitivity, but does not reduce the specificity[96;97].

False positives arise when the specimen contains mycobacteria other than *M. tuberculosis* complex organisms or in the rare cases where these may contaminate the water supply used in the staining procedure. In HIV non-infected patients these mycobacteria are not usually pathogenic.

However, in the mines, they have been found to be especially common, with up to 30%[9] of positive cultures (for Mycobacterium species) being due to these organisms, so that the specificity of sputum smear examination may also be relatively low. For this reason, among others, the Department of Minerals and Energy recommends that sputum cultures should be performed routinely and species identification be performed on all positive cultures[11].

In a mass screening programme, sputum examination is only possible if the subject being screened is able to produce a sputum specimen. The screening study performed by Churchyard *et al* showed that only 1.3% of miners admitted to the symptom of sputum production[9]. This symptom, on its own, had a sensitivity of less than 10%, and, since a majority of patients diagnosed with pTB are reported elsewhere to have sputum production[34], this finding suggests that miners with pTB are under-reporting the symptom. In their study, Churchyard *et al* took specimens from all miners whether or not they were able to produce a sputum specimen: saliva was collected from those without sputum. No published reports were found that define the costs of such an intervention in the South African mines setting.

The study reported by Churchyard *et al*[9] used a case definition that allowed culture negative patients to be classified as tuberculosis patients if the smear was positive or if there were new radiological abnormalities and the patient's symptoms had not responded to 5 days of antibiotic treatment. As a result, the sputum culture had a sensitivity of less than 100% (it was not the "gold standard"). In fact, only 70% of their patients had one or more positive cultures.

Approximately 30% of the smear positive patients in this study were subsequently culture negative. These findings are at odds with findings reported by Nagpaul *et al* from Bangalore (in Toman[34]) which showed that only 10% of smear positive patients were culture negative. This difference can be only partially explained by the high proportion of Mycobacteria other than tuberculosis found in the mining population.

In practice, in countries with a high HIV prevalence rate, it is usual to have to examine 9-10 suspects (27-30 smears) in order to detect 1 smear positive case[74]. This is enough work to

keep one technician busy for a day for each case detected. If one were performing mass screening the number of smears examined per case identified would be considerably higher.

A report from Italy, in which the presentation and outcome of TB was studied in HIV-infected people, those with negative chest Xray findings and negative bacteriological findings were found to be less likely to start on anti-tuberculosis treatment and to have diminished survival times, perhaps as a result of the delayed diagnosis and treatment initiation[35].

There were 3 other reports found in this review that describe the use of sputum examination as a mass screening tool. In a study reported from Thailand[98], villages were visited and villagers had the symptoms of pTB described: they were then invited to present for a sputum smear examination if they had any of the symptoms. The method was followed, 2 weeks later, by mass screening of the entire village population using smear examination. The yield was similar (17 cases versus 18) in both cases, although the time and cost of the first, more selective intervention were much lower.

In a study in the remote mountain areas of Nepal, mobile microscopy "camps" were held periodically over a three year period[99]. These camps were preceded by intense publicity in the villages in which people were encouraged to present for sputum examination if they had a cough of more than 3 weeks duration. The results were disappointing, with no appreciable impact on the case detection rate. However, it was noted that the camps increased the number of contacts with the health services for women in these remote areas.

In an American study among the homeless who made use of a large city shelter, spot sputum screenings (microscopy and culture were performed on all specimens) 10 months apart were used[100]. Of 120 people who were assessed, 4 cases of pTB were detected by this intervention, although only 1 of the 4 was detected by microscopy alone. Unfortunately, it was not possible to determine the incremental benefit of the intervention since there was no control group.

Since there were only 4 studies found in which sputum examination was used as a screening tool, and all 4 were in different settings, meta-analysis was not possible.

The workshop participants did not feel that there should be further investigation of the use of mass sputum screening in South African mines at this stage, mainly on the grounds of poor sensitivity of smear examination for detecting asymptomatic and smear negative cases, and the high cost of cultures. However, they also felt quite strongly that laboratory services should be quality controlled and that automated methods of examination should be considered.

NEWER METHODS OF DIAGNOSIS AS POTENTIAL SCREENING TOOLS

NEWER TECHNOLOGIES FOR SPUTUM SMEAR EXAMINATION

Fluorescent stains allow for more rapid smear examination (and hence a larger number of smears examined per given staff complement) as well as a marginally higher sensitivity as compared to Ziehl-Neelsen staining[74;101]. The sensitivity can be increased by concentration methods[102;103] and centrifugation[104;105] at forces higher than can be achieved using standard equipment[106]. A number of different sedimentation methods have been used and it is possible that some are better than others in terms of the ultimate sensitivity of the test[75] although not a "panacea"[107] for poor technique.

In addition, a number of automated procedures have been evaluated and found to result in greater reliability of the test. These include automated stainers[108] and computer aided automated microscopy[109;110], and have resulted in marginal improvements in sensitivity when compared to manual methods. The promise is that such methods may outperform the routine manual laboratory examination to a considerable extent, especially where very large numbers of negative smears are to be examined (as with a mass screening operation). The few positive smears can then be checked more carefully using the manual method.

The Foundation for Innovative New Diagnostics (FIND), based in Geneva and set up under the auspices of the World Health Organization and Tropical Diseases Research, and currently partly funded by The Bill & Melinda Gates Foundation, has prioritised the development of late-stage TB tests[111]. They have begun their work with a programme of research aimed at optimising sputum microscopy.

They have funded a number of projects that are examining new ways of processing the sputum specimen prior to smear preparation, and the results of these projects should be available during 2005. One of their projects is examining the improvement of programmatic aspects in laboratory diagnosis. These initiatives are aimed at improving the sensitivity and reliability of sputum smear examination as a diagnostic tool for TB suspects, but the findings may also be relevant to active case finding.

NEWER TECHNOLOGIES FOR SPUTUM CULTURE

The sensitivity of sputum culture for detecting *M. tuberculosis* can be increased by about 5% if the radiometric BACTEC system is used for culture rather than the traditional Lowenstein-Jensen culture medium alone, and the time taken for the cultures to appear positive is greatly reduced as well[112]. However, the equipment required is somewhat costly and there may be issues concerning the disposal of the radioactive material used. Alternative rapid culture methods have been investigated, all with very high specificities, and the use of the Mycobacteria Growth Indicator Tube (MGIT™) using non-radioactive materials is proving

promising with a number of reports showing equivalent rapidity and sensitivity when compared with the radiometric BACTEC system[103;112-114].

Another method that has proved to be equally as efficacious as the BACTEC method, both in terms of time taken and sensitivity, is the Difco ESP Culture System II, although the equipment required is almost twice as expensive as that required for the radiometric BACTEC system (over $70 000) and there have been problems reported with contamination[115]. The cost of the consumables required in 1997 was about $3 per test for either the BACTEC system or the Difco system[115].

Recently, a chromogenic culture system has been described[116] that is claimed to differentiate real growth from contamination. The method is said to be fast, relying as it does on colour changes sensitive to metabolic activity of the mycobacteria. One of these related systems has been evaluated [117], and found to have sensitivity similar to Lowenstein-Jensen medium and with a shorter mean time to detection of growth of only 15-17 days. The authors report that the system resulted in manipulation errors and that the system needs further development. One advantage of this test is that it can be performed without microscopic examinations, avoiding many of the quality issues that these raise.

NUCLEIC ACID AMPLIFICATION METHODS

Nucleic acid amplification (NAA) tests rely on polymerase chain reaction (PCR) techniques to amplify the quantity of mycobacterial nucleic acid present to the point where it becomes readily identifiable using a *M. tuberculosis* –specific DNA probe. The process is fairly complex and consists of a number of steps, so that quality assurance measures become particularly important if the results are to be reliable. Woods[118] has outlined some of the more important quality assurance steps in this process.

In the first place, areas for the preparation of reagent, for the processing of the specimens, and for the carrying out of the actual tests need to be physically separated in order to avoid or at least minimise cross-contamination and incorrect positive results. Secondly the area where the amplification and detection are carried out needs to monitored on a regular basis for contamination.

Thirdly, the laboratory technicians must be well-trained and regularly assessed as to their ongoing competence and the results of the assessment must be recorded. Fourthly, the equipment used must be regularly maintained and these maintenance activities must be documented as well. Fifthly, the test manufacturers must provide good quality control specimens to run with every assay. Finally, quality may be affected by the number of tests requested, by the ways in which specimens are collected and stored during transport to the laboratory as well as the way in which they are handled on reaching the laboratory. The big danger is cross-contamination since the slightest cross-contamination, from TB infected to non-infected specimens, will result in incorrect positives.

There are several commercially available NAA systems for the identification of *M. tuberculosis* in sputum specimens. In theory, these systems would be as well able to identify material from dead mycobacteria (perhaps remaining in scar tissue from a long past episode of TB) as from living micro-organisms indicating current disease. However, in a study reported in 2003, Rajalahti *et al*[119] , using the Amplicor system produced by Roche, were unable to detect any *M. tuberculosis* complex material from 25 war veterans who had pTB during the period 1940-1959, and who were considered to have been inadequately treated at the time, or from 19 subjects with cavitating disease who were properly treated during the period 1980-1983.

The Amplicor and LCx (Abbott) systems have been evaluated in terms of sensitivity and specificity and the results have been published. The Amplicor system has been evaluated by Tortoli *et al*[120] and by Mitarai *et al*[121], while the LCx system has been evaluated by Tortoli *et al* (*ibid*) and by Ruiz-Serrano *et al*[122]. Neither Tortoli *et al* nor Mitarai *et al* mentions how many of the specimens were obtained from HIV-infected individuals, although both studies were carried out in developed countries. Tortoli *et al*'s results (368 sputum specimens were analysed, of which approximately 15% were culture positive) report sensitivities and specificities for smear positive and smear negative pTB as follows:

		Amplicor	LCx
Smear +	Sensitivity	96.43	91.67
-	Specificity	63.89	55.56
Smear -	Sensitivity	34.62	46.15
	Specificity	98.37	95.64

Table III: Sensitivities and specificities for two NAA tests

The NAA positive detection rates were very similar to the combined culture results using Lowenstein-Jensen media and radiometric BACTEC assays. However, there were considerably more false positives as reflected in the lower specificities for the smear positive specimens. Mitarai *et al*[121] report very similar results for their study using the Amplicor system. In their study, they did not report separate parameters for smear negative and smear positive patients, rather a combined sensitivity of approximately 60% and a combined specificity of over 98%. Of interest in their study is that they report that of 77 results that were positive using conventional culture but negative using NAA, 37 were cultures of *M. tuberculosis* (they had 231 culture positive specimens for this organism) of which 8 were both smear and culture positive. The remainder were all NMTs. Ruiz-Serrano *et al* obtained roughly similar results for sensitivities to those reported by Tortoli *et al* using the LCx system in a group of patients

stated to include "many" HIV infected individuals, but were able to achieve much higher specificities of 100%. They did not estimate the parameters separately for smear positive and smear negative subjects. However they found that they were able to significantly improve the sensitivity of the method by omitting a step where the specimen is washed with distilled water prior to re-suspension. By directly re-suspending after decontamination, sensitivity improved from 53.8% to 85.2%; however, the specificity fell from 100% to 94.9%.

Finally, there have also been reports of NAA tests that are based on assay systems that have not yet become commercialised. These non-commercial in-house assays generally result in higher sensitivities than are achieved with commercially available systems (Dr.Karin Weyer, personal communication). For example, Araj *et al*[123] have published results for an in-house assay where the sensitivities for smear positive and smear negative specimens were 83% and 63% respectively. They reported specificities of 94% in both groups. The numbers in their study were, however, small (24 smear positive specimens and 11 smear negative specimens). The differences in the reported specificities for the commercially available test systems need to be further explored before one can make any general recommendations, and it is possible that there were differences in the populations from whom the specimens were obtained such as HIV infection rates.

ANTIBODY DETECTION METHODS

Antibody detection tests are tests that are aimed at identification of antibodies, in patients with active disease, that are absent, or present at lower levels, in healthy subjects. The difficulty is to develop a test that can distinguish between those with active TB, healed TB, TB infection, and TB non-infected, since antibodies may be present in all the first 3 categories[124]. The tests that have been described tend to be more successful in the identification of those with smear positive disease, whereas the diagnostic problem is more one of distinguishing between those with smear negative disease and other respiratory illnesses. Grange and Laszlo point out in their revue article[124] that even with *M. tuberculosis* -specific antibodies there is too much overlap in the values obtained for diseased and non-diseased subjects, resulting in poor likelihood ratios. In addition, unlike with acute illnesses, there is no possibility to make the diagnosis by demonstration of rising titres, although this is a possibility with HIV-infected patients. The detection of antibodies to lipoarabinomannan (LAM), a polysaccharide found in the mycobacterial capsule, has shown some initial promise. The different species of mycobacteria do not all have the same LAM structures, and the LAM of *M. tuberculosis* is sometimes referred to as "ManLAM" because it has mannitol residues in its "cap"[104]. The test is commercially available as the MycoDot™ test. It has been widely evaluated in African settings with varying levels of background HIV prevalence. In the MycoDot™ test kit, lipoarabinomannan is bound to plastic combs that are dipped into the subject's serum for 6 minutes, washed and then incubated with a signal reagent. The signal reagent reacts with vacant antigenic sites and a positive result appears as a dark red spot after 10 minutes[125]. Some reported sensitivities and specificities for LAM antibody detection are tabulated below:

Notes	Sensitivity	Specificity	Source
HIV+ TB vs. no TB (Thailand)?	31%-50%*	96% -98%*	Ratanasuwan et al[125]
HIV- TB vs. no TB (Thailand)?	28%-65%*	95%-100%*	Ratanasuwan et al[125]
4/13 cases were HIV+ (Tanzania)?	16%	84%	Somi et al[126]
29 Indian cases 115 US controls: no HIV data presented?	85%	93%-95%	Chan et al[127]
6 Ethiopian & 9 Swedish TB patients. No HIV status. Controls: 26 lab staff and 13 nurses (all Swedish)?	93%	96%	Hamasur et al[128]
12/37 adults with TB Guinea-Bissau HIV+?	50.0%	96.0%	Antunes et al[129]
34/232 adults with TB Guinea-Bissau HIV-?	67.6%	91.9%	Antunes et al[129]

*range depends on CD4+ lymphocyte counts   ? MycoDot™ on serum
? in-house assay used on serum   ? in-house test, urine specimens

Table IV: Sensitivities and specificities for a number of LAM tests

Tessema et al[130] have reported on the association between LAM levels in urine and LAM levels in serum using a commercially available ELISA kit. They studied 1000 adult patients referred to a respiratory clinic in Addis Ababa, of whom 200 were diagnosed as having pTB. They estimated that 30%-40% of the pTB patients were HIV co-infected. They found that the urine LAM test was positive in 94.2% of those with a positive serum LAM (but only 67.2% in smear negative patients) and that only 67% of those with a positive urine LAM test were positive on the serum LAM test. Unfortunately they did not report on the sensitivity and specificity of the urine LAM test for the diagnosis of active pTB, and it has not been possible to estimate these values from the data as presented in their article.

The MycoDot™ test has also been evaluated in South Africa (See Appendix 4) although the results have not been published (Dr. Karin Weyer, personal communication). The test was carried out on serum from 118 culture-confirmed TB patients, of whom 24 were HIV-infected, and 182 healthy controls. The sensitivity was 45.8%, and the specificity was 82.9%. When the specimens for the HIV-infected donors were considered separately the sensitivity was only 29.5%. However, as has already been discussed earlier in this review, the main expected impact of a mass screening programme on the incidence of TB is to be expected from the earlier detection of HIV non-infected subjects, since the HIV-infected people with TB are likely to present early. Therefore the poor sensitivity among HIV-infected individuals does not negate the potential use of this test for screening purposes. One unresolved problem, however, is that the potential cross-reactivity with environmental NTMs may be variable from area to area depending on which NTMs are prevalent in different regions, and this would have to be studied in South African mines: it could result in low specificities.

Even with a specificity of 80%, one would end up having to perform a definitive diagnostic test on 20% of the workforce and this may not be economically feasible, or operationally practical. There are several published reports of serological tests for the detection of antibodies to mycobacterial antigens. The sensitivities and specificities for the diagnosis of active pTB for some of these are summarised in the following Table (constructed from Chan *et al*[104], original articles not read).

Antigen	Sensitivity	Specificity
Antigen 5 (38 kDa antigen)	60% - 89%	94%-100%
A60 antigen	68% - 100%	80% - 100%
30 kDa antigen	48% - 80%	92% - 97%
30 kDa antigen (HIV+)	0% - 11%	
Antigen P90	70%	
19 kDa antigen	8% - 57%	
45/47 kDa antigen complex	40%	98%
P32 antigen	47%	
Cord factor	81%	96%
Glycolipid antigens	80% - 96%	91% - 98%
ICT tuberculosis test	68%	83%

Table V: Sensitivities and specificities for a number of serological tests

Most of these studies involved small numbers of subjects and were not carried out in South African gold miners, so that the values given should be treated with caution. The generally low sensitivities might be expected to be lower in populations with a high HIV prevalence rate, and in any event are generally little better (or worse) than the sensitivity of smear examination. Specificities are generally quite high, although lower than what is expected from smear examination, and the high exposure to NTMs in South African mines means that they might be lower under these conditions.

Some antibody tests have been assessed in South Africa. Further details are provided in Appendix 4, but the sensitivities and specificities are listed in the following Table (Dr. Karin Weyer, personal communication, all unpublished information).

Antigen	Sensitivity	Specificity
anda-TB GA test	97.1%	68.8%
Latex agglutination test	56.8%	81.9%
Immuzyme-TB	73.4%	83.2%
Kreatech TB kit	78.1%	85.2%

Table VI: Sensitivities and specificities for tests evaluated at the MRC, Pretoria

These results show generally higher sensitivities than one obtains using smear examination, but the lower specificities are disappointing since, for mass screening, these low specificities would result in an unacceptable number of false positives needing further testing. This makes their use logistically more difficult as well as more costly.

THE USE OF ANTIGEN DETECTION METHODS

Since, in general, higher sensitivity is obtained at the expense of specificity; one new approach has been to try to develop tests that detect *M. tuberculosis* antigens, since these tests would be very specific while permitting one to aim for higher sensitivity. In particular if the test is designed to detect any of a mixture of antigens then, in theory, sensitivity could be improved without, hopefully, any loss of specificity. Especially since the elucidation of the *M. tuberculosis* genome, it is now possible to identify areas of the genome that are not present in BCG or in NTMs[131], and then to describe the specific proteins that are coded for by these areas of the genome. Methods can then be developed to detect these specific antigens, either alone or in combination.

Andersen et al[131] go on to point out that for the 3 species that make up the *M. tuberculosis* complex, all three produce 3 antigens that are not found in this combination in any other mycobacterial species, although they may be present in other species individually or in pairs. These antigens are ESAT-6, CFP10 and MPT64 (this antigen, MPT64, is the same as that used in the patch test, described below, MPB64). ESAT-6 is especially specific, being found in only three other mycobacterial species.

These purified antigens can then be used either *in vivo* in the form of a skin test (see below) or *in vitro* for example through the stimulation of interferon gamma production, also described below.

PATCH SKIN TESTS USING MYCOBACTERIAL ANTIGENS

Nakamura *et al*[132] have reported on a promising new method of using antigens to detect TB disease by using a high dose skin patch test that incorporates the antigen MPB64. MPB64 was derived from Japanese cultures of *M. bovis* BCG. The skin patches were applied and left on the arm for 72 hours. They were then removed and any erythema induration or small red spots were regarded as "positive".

In their initial study, carried out in the Philippines on 53 patients with confirmed pTB and 43 healthy controls with positive PPD skin reactions, the test was found to have a sensitivity of 98.1% and a specificity of 100%. Most of the TB patients in this study had already been on treatment for 4 months and longer. The HIV infection status was not reported.

In a further study[133], also in the Philippines, the dermal patch was applied to 49 sputum positive patients with pTB, 13 who had completed TB treatment, and 28 non-TB but PPD positive controls. HIV infection status was not reported. Once again the patches were removed and a reading was made at 72 hours.

The sensitivity was found to be 87.8% and the specificity was 100%. In the subjects who had completed TB treatment, there was a gradual falling off in the number who were positive, until, by 7 months, all were negative. Numbers were, however, small.

The "patch test" as it has come to be known, is currently being evaluated further in field trials in the Philippines as well as in South Africa, and the planning for the initial review of results will be completed by the end of 2004 (Dr. Karin Weyer, personal communication).

ANTIGEN TESTS THAT EXPLOIT THE PRODUCTION OF INTERFERON GAMMA (IFN-?)

In these techniques whole blood is incubated overnight with the added antigens (such as ESAT or MPT64). Blood from patients with tuberculosis infection contains lymphocytes that respond by secreting IFN-?. This is detected by means of an ELISA test. The entire operation takes approximately 24 hours[134]. Newer versions of the tests incorporate the use of blood collection tubes that already contain *M. tuberculosis*-specific antigens (rather than PPD, thus improving the specificity) and other features that reduce labour and hence the possibilities of human errors. Since the results are on a continuous scale depending on how much INF-? is produced, receiver operating characteristic (ROC) curve analysis has been used to determine the optimal cut-off point to identify those who are most likely to have either a history of infection or current active disease.

Streeton *et al*[135] have determined the optimal cut-off point for identifying those with a history of past infection based on the PPD skin reaction as the gold standard. They analysed the results from 948 subjects of whom 1 was HIV-infected and 12 had current TB disease (the authors imply, but do not say, that these were all cases of pTB). Approximately half of their subjects had never received the BCG. They used the *Quanti*FERON®-TB test for their study.

Their results showed a sensitivity (for *infection* with *M. tuberculosis*) of 90% and a specificity of 98%.

For subjects with active disease, the sensitivity was 83.3% using the optimal cut-off point for a history of infection (95% confidence interval 62.2% to 100%). The specificity was not given, but, based on a further analysis of the results that they tabulated, would be approximately 58%. In fact, the mean IFN-? responses were higher for those with infection but no active disease as compared to those with active disease, indicating the inability of this test to adequately differentiate between these two groups, even although the cut-off value that was used had not been optimised for detecting those with active disease, and even although the study group would differ in substantial ways from the gold miners in South Africa (HIV sero-prevalence and BCG history).

Mazurek *et al*[136] have carried out a similar study using the same commercially available test as was used by Streeton *et al*. Their results are similar, but they go on to point out that there were fewer false positive results (when compared to PPD skin testing) due to BCG history or infection with NTMs. If the high prevalence of PPD skin test positives in South African gold miners can be shown to be significantly affected by BCG and/or NTMs, then the use of this test might be more specific than the skin test under South African conditions. However, it is

common to find that up to 30% or more of the miners have a past history of tuberculosis disease, so that its usefulness might be curtailed in this community.

PHAGE TECHNOLOGIES

In these tests, phage is added to the decontaminated and concentrated sputum specimen. It infects any mycobacteria present. Excess extracellular phage is then killed with a virucide. Thereafter, a mixture of *M. smegmatis*, which is rapid growing, and culture media is added to the specimen. The phage that has infected any mycobacteria in the original sputum specimen will have survived the virucide which only acts extracellularly, and now multiplies within those mycobacteria. It is then released and infects and kills those *M. smegmatis* organisms in its vicinity, so that, when the added *M. smegmatis* grows in the culture medium small clear circles are seen in the growth, indicating the presence of mycobacteria in the original sputum specimen[137]. Three reports have been found that describe the performance of this test. They all used a commercially prepared system known as *FASTPlaque*TB™ (FPTB).

The first report to be published was for a study of 1692 specimens from 853 pTB suspects in the Cape Town area of South Africa[138]. Subjects were all adults recruited from local clinics. The HIV-infection rate was not noted or commented on. Subjects who had a history of prior pTB during the past 3 years were excluded. Specimens were all subjected to FPTB, culture using Lowenstein-Jensen medium, and smear microscopy using auramine stain. Approximately 10% of the subjects were found to have pTB. The sensitivity and specificity of the test were found to be 75.2% and 98.7% respectively, as compared to 63.4% and 97.4% for microscopy. Although the performance of the test is not very much better than sputum microscopy, it was found to detect 54.1% of culture positive cases who had 2 consecutive negative smears. In fact the use of smear microscopy and FPTB permitted the diagnosis to be made within 2 days for 81.2% of all the culture positive cases. The authors recommend that the test be reserved for use in those with two consecutive negative smears.

The second report [139] refers to a study carried out in Karachi. In that study 584 sputum specimens were assessed by means of FPTB, Lowenstein-Jensen culture and acid fast microscopy. HIV-infection status of the subjects was not mentioned or discussed. Complete results for all three tests were available for only 514 specimens, of which 245 were culture positive for *M. tuberculosis*. Seventy of the missing results were due to overgrowth with contaminants. The authors suspected that the contamination occurred during handling because *Bacillus spp.* and *Staphylococcus spp.* were the commonest organisms identified as contaminants. These authors obtained sensitivities and specificities of 87.4% and 88.2% respectively for smear positive specimens and 67.1% and 98.4% for smear negative specimens. The direct cost of the FPTB test was $8 in Pakistan in 2002.

In the third study to have been reported [140] FPTB was used along with PCR, culture using a radiometric BACTEC method and smear microscopy using Ziehl-Neelsen staining on 201

sputum specimens. Specimens were excluded if the suspects had a past history of pTB. Again, there was no mention of HIV-infection status. Contaminants precluded the FPTB results for 7, and culture results for an additional 2, leaving 192 sets of analysable results. Again, no mention was made of the HIV-infection rate among the participants. The results were similar to those obtained in the two previously reported studies, although the contamination rate was lower than reported by Ali Albay *et al*, and the sensitivity was (perhaps as a result) somewhat higher at 87.5%.

These results are promising, although the cost will have to be reduced considerably before FPTB testing can be considered for mass screening. However, the influence of past TB and HIV on the performance parameters needs to be established first, even if the test becomes available at very low prices.

In addition, the FPTB test does not differentiate between *M. tuberculosis* organisms and NTMs[137], an additional potential problem with using this test on South African mines.

BETA $_2$ -MICROGLOBULIN

Beta$_2$ -microglobulin (B2M) is a serological marker of immune activation that has been used in the monitoring of response to therapy for a number of conditions in the past. More recently it has been found to be raised in patients with TB disease and in patients with HIV-infection. It is even more markedly raised in patients with HIV-infection with TB disease. It has been investigated as a measure for monitoring the course of HIV disease as well as for prognostic purposes in HIV-infected individuals[141]. It has not been evaluated as a diagnostic or screening measure.

Flynn *et al*[142] demonstrated that mice that were lacking the gene responsible for B2M expression had a very high mortality very soon after infection with tuberculosis bacteria.

Piwowar et al [143] went on to measure B2M levels in 66 Ugandan patients and controls, 38 who were HIV non-infected and "asymptomatic", 12 who were HIV-infected and "asymptomatic" and 16 who were HIV-infected and "symptomatic". They found that the HIV-infected group had higher levels of B2M than those who were non-infected, and, among those who were HIV-infected, the symptomatic group had higher levels than those who were asymptomatic.

In extracting and reworking the data that they presented, it becomes apparent that B2M levels with an optimum cut-off value of 3.5 mg/litre had sensitivity for distinguishing between symptomatic and asymptomatic HIV-infected individuals of 75% and a specificity of 50%. They unfortunately did not include a group of HIV non-infected "symptomatics" in their study. The B2M levels among the healthy controls were considerably higher than previously reported for healthy westerners (approximately double in fact) which they presumed might be caused by a high background prevalence of tropical parasitic infections.

Unpublished data from South African gold miners suggests that the "normal" levels in South Africa are more similar to those found in western society (author's personal unpublished data).

Wanchu *et al*[144] conducted a similar study among Indian patients. Their study included 12 HIV-infected tuberculosis patients and 14 HIV non-infected healthy controls. Once again, extrapolating from the data that they presented, and using a cut-off value of 1.5 mg/litre, the sensitivity of this test would be almost 100% but the specificity would only be about 75%.

The reported study had small numbers, however, and hence these extrapolated values would have very wide 95% confidence intervals. Again, there was no comparison group in the study who were HIV non-infected but TB disease positive, so that these extrapolations may prove to be incorrect in reality, as will the optimal cut-off point.

None of these newer technologies has been subjected to a good quality published economic analysis that we were able to find in the literature search. In addition the studies are too few, too small, or too dissimilar for valid meta-analysis to be performed.

THE WORKSHOP VIEWPOINT

At the workshop held to discuss the literature review findings, these alternative newer methods were not presented in detail due to the illness of the main presenter. However, the workshop participants felt that there was probably little to be gained (at least at this stage) from further research in the mines into any of these methods as screening tools.

The main concerns were the possible lack of sensitivity due to the high prevalence of HIV in the South African mining setting and low specificities due to possible other infections in the South African setting, and which would result in too many false positives who would then require further definitive testing to exclude TB. Workshop participants were also doubtful that the costs of many of these alternative methods could be reduced to feasible levels for mass screening, at least in the medium term.

DISCUSSION AND CONCLUSIONS

Active case finding implies that health service workers search for additional cases among those who have not self-presented for diagnosis and care. This could involve mass screening of whole populations, or it could be confined to those considered at higher risk, such as close contacts of infectious cases, silicotics, HIV positive individuals and those with a history of TB.

Where mass screening is employed, if the prevalence of disease is 2% and the screening test has a sensitivity of 100% and specificity of 98% then 50% of the positives identified will be false positives. If the sensitivity is only 50% then 2/3 of those testing positive will be false positives. As a result, for the tests that are currently available, and given their test parameters, a second round of definitive testing needs to be carried out using a diagnostic

test that has high specificity (to avoid treating healthy people with 6 months of TB treatment unnecessarily) as well as a high sensitivity (to avoid further loss of cases). Currently, sputum culture is the best-performing diagnostic test available for this second round of testing. It should be remembered that this second round of diagnostic testing will need to be carried out on 2-3% of the workforce given the above scenario, and this may not always be logistically easy to do.

Pulmonary TB is a disease that might develop in relatively asymptomatic individuals who often only start treatment several months later. For a given screening test, the effectiveness of a screening programme that aims to detect cases early in order to reduce the period of infectiousness depends on the duration of infectiousness relative to the inter-screening interval.

Even if the screening test has a sensitivity of 100% and a specificity of 100%, if the duration of infectiousness is only 3 months, then an annual screening programme can only hope to achieve a reduction of 1/8 in the average duration of infectiousness. If, as is the case with currently available screening methods, the sensitivity of the screening test is only 50% then the reduction in the duration of infectiousness will only be by 1/16.

The benefits from mass screening become even more dubious if one considers that:

- Screening tests may have an even lower sensitivity in patients with HIV infection.

- Since smear positive patients are more likely to self-present, the screening test may detect a disproportionate number of those who are smear negative and less infectious anyway.

- Since 60%+ of all incident TB cases are caused by reactivation of dormant infections, the screening intervention will not have any short-term effect on these case numbers.

- Even if the screening test has a specificity of 98% it will identify very large numbers of people who do not have the disease (2% of the screened population), and who will have to be subjected to further definitive testing in order to exclude it. In situations where there is a high rate of BCG immunisation and, as in the South African mines, exposure to NTMs, many tests will have even lower specificities than are reported elsewhere in the world.

- Although up to 1/16 of transmission might be stopped through screening this may not translate into 1/16 fewer new infections since in a high prevalence situation there are often multiple sources of infection, not just the one contact who has been started on early treatment. People who become infected may become infected more than 1 time before becoming ill. Preventing 1/16 of these infections may not prevent the person becoming ill.

- Even if the mass screening programme were to work well by reducing the average period of infectiousness there would also be a reduction of the prevalence rate, changing the dynamics of the epidemic and reducing the yield from the screening programme.

✍ Increasing the frequency of the screening test would simply increase the net cost per case since the yield would decline as the screening programme develops.

On the other hand, it may also be argued that:

✍ The benefits of the screening programme need to be assessed in a more dynamic way. If 100 cases of pTB are detected early in 1 year then this would result in, possibly, 10 fewer cases over the next 10 years and 1 fewer subsequent to those 10 years as well. So we are not just detecting 100 cases early, we are also preventing 11 cases in the future. This may be difficult to quantify.

✍ Since HIV-infected people with TB have been shown to present earlier than those who are HIV non-infected and are more likely to be smear negative, from a transmission dynamics perspective it is not of much concern that a test may have poor sensitivity among those infected with HIV, as long as it is not used as a definitive diagnostic test (i.e. reserve it for screening only, and insist on a confirmatory test such as sputum culture).

✍ Even if a test has a sensitivity of only 70% this does not negate its use in this screening context since even this test will prevent 70% of the potential 25% reduction in transmission of a test with a sensitivity of 100%, and this may be cost-effective if the test is cheap enough to administer.

✍ Low specificities are only an issue if the definitive diagnostic test is either very expensive, carries a risk for the person being tested, or is culturally unacceptable. Sputum culture is a definitive diagnostic test that, although fairly expensive, is non-invasive and culturally acceptable.

✍ In contrast to screening tests for conditions like malignancies where the aim is to detect as many patients as early as possible, with dire consequences for the patient if the diagnosis is missed, and for which scenario a screening test should be highly sensitive even at the expense of specificity if necessary, the aim here is to reduce transmission in a cost-effective way. Therefore, a screening test with relatively lower sensitivity but higher specificity may be suitable. Tests that, because of their low sensitivity, may be inappropriate for definitive diagnostic purposes may be quite useful in this screening context.

✍ The fact that a screening programme will only be cost-effective for a limited number of years is not an argument against the programme: rather this is proof that it has done its job. The programme can simply be discontinued once the incidence rate has fallen.

It follows from the preceding discussion that the usual criteria used for assessing the quality of a screening test may be relative in the context of screening for pTB. It would be useful to have a reliable validated epidemiological model that incorporates screening parameters in its

structure, as well as economic study information to help resolve this debate. However, such a model and this economic information are not currently available.

SYMPTOMS QUESTIONNAIRES

Concerning the use of symptoms questionnaires as a means of mass screening, it is likely that the cost will be quite low since there are no significant testing costs other than the time taken to complete and analyse the questionnaires. However, there is little convincing evidence that they actually work. It is possible that miners may play down their symptoms when completing such questionnaires for fear of losing their jobs, especially if administered at the time of re-induction when new contracts may not have been signed yet.

On most South African mines, however, miners have several contacts with health professionals each year, and it would make sense, given the very high incidence rates of pTB in the mines, for all health professionals to be continually sensitised to the possibility that a miner who comes to see them, even for some routine purpose such as an audiometry test, might have pTB as well. All categories of staff might be trained to continually ask about cough, night sweats and weight loss, and to look for any visible evidence as well.

The very long periods of infectiousness described from Churchyard *et al*'s study suggest that opportunities for early diagnosis are being missed in the routinely accessed health services. This is born out by the results of the autopsy study of Murray *et a*l which showed that a high proportion of miners who die from TB do so without a pre-mortem diagnosis.

It is recommended that further research might be worthwhile, but a relatively low priority at this stage. If further research is carried out then it needs to address the timing of the questionnaires as well as the language used (must be "culturally competent"). One possibility is to evaluate a "scouting" programme whereby specially trained personnel are tasked with continually seeking out those with symptoms of pTB and referring them for further testing at the health station.

SERIAL WEIGHTS

Although this strategy was not discussed in the workshop and may be of limited use due to the high prevalence of HIV among mine workers, it's low cost means that it may prove useful as a method of identifying a sub-group for more focused screening using chest X-rays and sputum examination. This may be worthwhile examining further in a formal study.

ANNUAL CHEST X-RAYS

In general the use of chest X-rays as a screening tool for pTB is not supported by the published evidence, since the specificity, especially, is so low that one would end up having to subject a high proportion of the workforce to sputum culture. However, in the mines, because previous chest Xrays are available for comparison purposes, this might not be true. If the person reading the X-rays looks for new changes only then this might be a screening test with

a reasonably high specificity. In any event, since the miners continue to have annual chest X-rays as part of their occupational surveillance, it is essential that the screening focus be on making use of the opportunity to look for new changes that might be indicative of pTB. Indeed the mines report that a high proportion (currently over 30% in many cases) of their TB cases is detected by this method. This fits well with the long period of infectiousness that has been described on one mine.

Most of the published reports suggest that a mean duration of symptoms of perhaps 25 months is the norm in most parts of the world. If self-presenting cases are detected at a constant rate throughout the year and if the mean duration of disease prior to diagnosis and treatment is 3 months then one would not expect the annual chest X-ray to detect more than 25% of the cases, assuming that presence of symptoms is a good surrogate for "disease".

The fact that the annual chest X-ray is detecting over 30% in many cases suggests that the duration of disease is longer than 3 months. This may be especially true for the HIV non-infected miners, since a study on one mine has indicated that up to 50% of these pTB cases (in the HIV non-infected) are detected at the time of the annual chest X-ray.

This evidence could be used to advocate more frequent chest X-ray screening in which the reader looks for new changes on the X-ray (to improve the specificity). However, one needs to be careful because the addition of a second X-ray changes the cost-effectiveness dramatically. Where miners are being subjected to X-ray anyway, the cost is a sunk cost. But when an additional screening is introduced that would not otherwise take place then the entire cost of the additional screening must be included in the estimation of the cost per case detected. In their randomised controlled trial of twice yearly vs. annual chest X-ray screening, Churchyard *et al* found that there was no difference in the incidence rate at 1 year, but there might have been a lower case fatality rate in the group undergoing twice yearly screening.

In their study, they randomised individuals within the same mine to either arm, and this might have diluted any effect as far as the incidence rate is concerned, but would not have biased the measurement of the case fatality rate. The workshop attendees found these results intriguing and felt that the study should be repeated (with randomisation of sites this time) to see whether the findings could be repeated. They placed such a project joint second on the list of research priorities. It can also be argued that if more than 25% of the HIV non-infected cases are being detected at the annual chest X-ray then too many cases are being missed for too long by the routinely available health care services and that this needs to be addressed rather than introducing an additional and more costly screening step.

This argument raises an interesting question. Where there is a routinely used annual screening tool such as the annual chest X-ray, do health care workers, to some extent, abdicate their responsibility for the day-to-day detection of disease to the screening programme? It is of the utmost importance that health care workers understand that the screening tools that are used have poor sensitivity and should never be relied upon to detect

patients or to secure a definitive diagnosis in those who undergo mass screening (when there will be large numbers of false positives, even if the specificity is over 98%).

The screening programme is an additional intervention, and not there to replace good diagnostic clinical practice. Paradoxically, if health care workers do, even sub-consciously, abdicate their clinical responsibilities to the screening programme, which uses methods of poor sensitivity, the existence of any screening programme might make the case detection rates and transmission worse.

The workshop attendees strongly supported further research into the ways in which decisions and policies are implemented (or not implemented) regarding the routine TB services in the mines, and into ways in which to improve these services. They regarded this as the top priority for further research. It is recommended that the annual chest X-rays continue to be used as a source of case detection, followed by sputum bacteriology in those who screen positive in order to establish a diagnosis. Although in the medium term it is recommended that a further study be carried out of twice yearly vs. annual chest X-rays, the top priority is for research into the ways in which the routine services can be improved and strengthened. Health workers need to be educated concerning the limitations of the chest X-ray as a screening or diagnostic tool and encouraged to increase their levels of suspicion and the aggression with which they search for the TB bacillus when their suspicions are aroused.

MANTOUX SKIN TESTING

The use of the currently available Mantoux skin test, using tuberculins, is not useful for screening for active disease in adults in South Africa due to the very low specificity. Although the use of more specific sensitins in the future may improve the specificity there is no indication that the test will ever be useful for distinguishing between infection and disease in the South African mining population. It is recommended that no further research be funded into the use of existing technologies as potential screening tests for active disease in the South African mines at this stage.

SPUTUM BACTERIOLOGY

The sputum smear examination is the current recommended definitive diagnostic test for pTB in resource-poor countries. In these settings approximately 1/10 of suspects are found to have pTB. The test has a sensitivity of about 50% and a specificity of about 98%. This means that the LR+ is 25. Since 1/10 suspects actually have the disease the post-test odds of disease is 2.78. This means that the positive predictive value is 0.735. In other words, under typical field conditions, just over ¼ of those who are diagnosed as having pTB using this method will not actually have pTB. The situation is worse if the test is used as a screening test on the general population such as in a mining work-force where the prevalence of pTB might be only 5% of those who are able to produce a sputum specimen. Under these

circumstances the pre-test odds are 1:19 and the post-test odds become 1.32:1. The PPV is now 0.569. In other words, over 40% of those who test smear positive in such a screening programme would not actually have pTB. Thus it is imperative that the positive smear result is followed by a further definitive test such as culture and species identification. Smear examination alone should never be used as both a screening and definitive diagnostic test. In fact it might also be argued that even under clinical conditions the positive smear should always be followed by a culture and species identification, and treatment should only be started on the basis of the smear result if two consecutive smears are positive, if a culture is positive or if the patient is too ill to wait for the results of the sputum culture test.

Patients who are detected by means of a mass screening intervention are by definition asymptomatic, so the rule should be that the sputum smear microscopy result should be positive on two specimens, or else on one specimen followed by a positive culture.

In practice, smear examination is a potential screening tool for pTB, as long as it is not used to determine the definitive diagnosis, since it is unlikely that other screening tests will have specificities much higher than 98%. One issue, however, is that the specificity may be slightly lower in the mines due to the presence of NTMs. The costs of such a programme may be reduced by the fact that only a small proportion of miners will be able to produce a sputum specimen. Sputum induction needs to be further assessed as a way to improve sensitivity, since the published reports about this manoeuvre have been unconvincing.

If the sputum screening is performed 6 months after the routine chest X-ray screening then the impact of the intervention in terms of additional cases identified early might be expected to be greater. The outcome measure of interest is the duration of infectiousness, so that, if further research is carried out, there should be documentation of prevalence and incidence before and after the intervention. If the outcome of interest is the short-term incidence rate change then randomisation should be performed using mines or shafts as the units of randomisation rather than individuals. The workshop attendees did not see this kind of research as a priority, however, since it appears (admittedly with imperfect knowledge) that the greater priority is to improve the already routinely available services.

Although this is only indirectly related to the question of screening, further research is probably justified into ways of improving the sensitivity and quality of smear examinations along the lines described in this section of the literature review.

NEWER TECHNIQUES

Of the newer techniques described there are a few that appear promising as definitive diagnostic tests, such as the INF-? tests, and the patch test. The other newer tests that have been described suffer especially from low specificities (making them less useful either as definitive diagnostic tests or as screening tests). Where specificities appear to be high, the results are often based on small sample sizes or else have not yet been evaluated in the

South African mining (or similar) population. Tests that are based on the MPB-64, ESAT-6 or CFP-10 antigens are likely to be more useful due to the high background levels of NTMs in the South African mining industry. Additional issues when a test is to be used as a screening test are the cost (price plus administration) and acceptability.

The INF-? tests are likely to be too expensive for screening use since, unlike the case of a sputum-based test where only a small proportion of the population will be able to produce a specimen, the entire population being screened will be able to be tested. In addition, the test requires the taking of blood from asymptomatic people and there may be objections to this from a number of the work-force.

The high dose patch test, on the other hand, may be quite cheap once it is commercialised. Providing that the two trials currently under way confirm the earlier promise of this technique, this may turn out to be an affordable and acceptable test with good sensitivity and very high specificity. In that case (and the final results are due for release at the end of 2005), it might be worth funding research into the operationalising and economic evaluation of the test when used as a screening test in 2006.

Although the LAM tests showed initial promise, once they were used in tropical countries with a high prevalence of HIV infection their performance was not as good, certainly not as definitive diagnostic alternatives to the *status quo*. However, the fact that they might be used on urine samples makes them potentially useful for screening purposes if the specificity, especially, under South African conditions, can be improved.

The phage-based tests also show some initial promise. Although designed for definitive diagnostic purposes, they might be useful as screening tests if the technology can be simplified and the costs kept really low. One advantage from a cost perspective is that they are performed on sputum so tests would only be performed if subjects are able to produce sputum specimens.

RECOMMENDATIONS

A number of potentially useful research projects come to mind as a result of this literature review and the discussions held during the workshop. The workshop attendees recommended that there is a single top priority for research and that is presented as item number 1. Thereafter they felt that there were two secondary priority areas (neither higher than the other) that are listed as 2 a. and 2 b. Finally, some additional recommendations are made involving evidence and material that was either not presented at the workshop or that has come to light through reflection subsequent to the workshop.

1. A study of why it is that modern recommended methods for the diagnosis of pTB are not always practised on South African mines. This should include the decision and policy making processes on mines, the factors affecting implementation particularly ways to

improve primary care services in early detection of cases as well as the beliefs and attitudes of health workers and miners. This could comprise a number of sub-studies:

   a. An observational study to record the processes of the routine programme, including primary care services, against a set of benchmarks.

   b. Qualitative studies of the ways in which policy decisions are made and how they are implemented.

   c. KAP studies of health care workers and of miners.

2. The following two research topics were rated joint second as priorities by the workshop attendees:

   a. The effect on pTB case-fatality rates of twice yearly vs. once yearly chest X-rays.

   b. Laboratory quality assurance for sputum microscopy and culture.

3. Robust economic and epidemiological models need to be developed that can be used to assess the viability (in theory) of proposed interventions.

4. The patch test, if found at the end of 2005 to be as promising in field use as it appears, should be investigated as a potential screening tool in 2006.

5. The design, operationalising and evaluation of symptoms questionnaires to identify those at higher risk as a sub-group for further investigation.

6. The use of serial weight measurements (once per year, preferably mid-way between the routine annual chest X-rays) may also be worthwhile as a strategy for further investigation.

LIST OF REFERENCES

1. Lowe, R. E. and Murray, J. Tuberculosis case finding. *Proceedings of the Mining Medical and Other Health Care Professionals Association* 65, 31-38. 1992.

2. Director: health Systems research, Research Coordination and Epidemiology. TB in South Africa. *Epidemiological Comments* 4(3), 2-12. 2001.

3. Sonnenberg, P., Glynn, J. R., Fielding, K., Murray, J., Godfrey-Faussett, P., and Shearer, S. HIV and pulmonary tuberculosis: the impact goes beyond those infected with HIV. *AIDS* 18, 657-662. 2004.

4. Ross, M. H. and Girdler-Brown, B. V. *Corporate Audit: Tuberculosis and HIV/AIDS (Delmas, Matla, Douglas, Optimum and Zululand Anthracite Collieries).* 1999. Ingwe Coal Corporation.

5. Girdler-Brown, B. V. and Ross, M. H. *Corporate Audit: Tuberculosis and HIV/AIDS (Middelburg and Rietspruit Collieries).* 1999. Ingwe Coal Corporation.

6. Rothman KJ, Greenland S. *Modern Epidemiology.* Philadelphia: Lippincott, Williams and Wilkins, 1998.

7. Murray, J. and Nelson, G. Demographic data and disease rates in deceased South African miners. *Occupational Health Southern Africa* , 16. 2004.

8. Murray, J., Back, P., Lowe, P., and Coetzee, L. *Clinico-pathological Study of Pulmonary Tuberculosis.* SIMHEALTH 611. 2000. Johannesburg, SIMRAC.

9. Churchyard, G. J., Charalambous, S., Moloi, V., Seabi, O., Sehloho, D., Mangenene, N., Rankhakile, M. K., Lesholi, W., Corbett, E. L., and Grant, A. D. *The Use of Sputum Screening as Active Case Detection for Tuberculosis.* Health 705. 2003. Johannesburg, SIMRAC.

10. The Manager: National Tuberculosis Control Programme. *The South African Tuberculosis Control Programme Practical Guidelines 2000.* 2000. Pretoria, National Department of Health.

11. Chief Inspector of Mines. *Guidance Note for Occupational Medical Practitioners: Tuberculosis Control Programmes*. 2003. Pretoria, Department: Minerals and Energy.

12. Churchyard, G. J., Kleinschmidt, I., Corbett, E. L., Mulder, D., and de Kock, K. M. Mycobacterial disease in South African gold miners in the era of HIV infection. *International Journal of Tuberculosis and Lung Disease* 3(9), 791-798. 1999.

13. Murray, J., Kielkowski, D., and Reid, P. Occupational disease trends in black South African gold miners. *American Journal of Respiratory and Critical Care Medicine* 153, 706-710. 1996.

14. Kleinschmidt I,.Churchyard GJ. Variation in incidence of TB in sub-groups of SA gold miners. *Occupational and Environmental Medicine* 1997;54:636.

15. Hnizdo E. Risk of pTB relative to silicosis and exposure to silica dust in SA gold miners. *Occupational and Environmental Medicine* 1998;55:496-502.

16. Cowie, R. L. The epidemiology of tuberculosis in gold miners with silicosis. American *Journal of Respiratory and Critical Care Medicine* 150, 1460-1462. 1994.

17. van den Broek J, Borgdorff MW, Pakker NG, Chum HJ, Klokke AH, and Senkoro KP. HIV-1 infection a a risk factor for the development of tuberculosis: a case-control study in Tanzania. *International Journal of Epidemiology* 1993;22:1159-65.

18. Genewein, A., Telenti, A., and Bernasconi, C. Molecular approach to identifying route of transmission of tuberculosis in the community. *The Lancet* 342(841), 844. 1993.

19. Small, P. M., Hopewell, P. C., and Singh, S. P. The epidemiology of tuberculosis in San Francisco. *New England Journal of Medicine* 330(1703), 1709. 1994.

20. Alland, D., Kalkut, G. E., and Moss, A. R. Transmission of tuberculosis in New York city: an analysis by DNA fingerprinting and conventional epidemiological methods. *New England Journal of Medicine* 330(1710), 1716. 1994.

21. Behr , M. A., Warren, S. A., Salamon, H., Hopewell, P. C., Ponce de Leon, A., Daley, C. L., and Small, P. M. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *The Lancet* 353, 444-449. 2-6-1999.

22. van Deutekom, H., Hoijng, S. P., de Haas, P. E., Langendam, M. W., Horsman, A., van Soolingen, D., and Coutinho, R. A. Clustered tuberculosis cases: do they represent recent transmission and can they be detected earlier? *American Journal of Respiratory and Critical Care Medicine* 169(7), 806-810. 2004.

23. Godfrey-Faussett, P., Sonnenberg, P., Shearer, S. C., Bruchfeld, J., Mee, C., Morris, L., and Murray, J. Tuberculosis control and molecular epidemiology in a South African gold-mining community. *The Lancet* 356, 1066-1071. 2000.

24. Sonnenberg, P., Murray, J., Glynn, J. R., Shearer, S., Kambashi, B., and Godfrey-Faussett, P. HIV-1 and recurrence, relapse and reinfection of tuberculosis after cure: a cohort study in South African mineworkers. *The Lancet* 358, 1687-1693. 2001.

25. ten Asbroek, A. H. A., Borgdorff, M. W., Nagelkerke, N. J. D., Šebek, M. M. G. G., Devillé, W., van Embden, J. D. A., and van Soolingen, D. Estimation of serial interval and incubating period of tuberculosis using DNA fingerprinting. *International Journal of Tuberculosis and Lung Disease* 3(5), 414-420. 1999.

26. Sepkowitz, K. A. Tuberculin skin testing and the health care worker: lessons of the Prophit Survey. *Tubercle and Lung Disease* 77, 81-85. 1996.

27. Ward, H. A., Marciniuk, D. D., Pahwa, P., and Hoeppner, V. H. Extent of pulmonary tuberculosis in patients diagnosed by active compared to passive case finding. The *International Journal of Tuberculosis and Lung Disease* 6(5), 593-597. 2004.

28. Long, N. H., Diwan, V. K., and Winkvist, A. Difference in symptoms suggesting pulmonary tuberculosis among men and women. *Journal of Clinical Epidemiology* 55, 115-120. 2002.

29. El-Sony, A., Enarson, D. A., Khamis, A., Baraka, O., and Bjune, G. Relation of grading of sputum smears with clinical features of tuberculosis patients in routine practice in Sudan. *International Journal of Tuberculosis and Lung Disease* 6(2), 91-97. 2002.

30. Mori, T., Shimao, T., Byoung, W. J., and Sung, J. K. Analysis of case-finding process of tuberculosis in Korea. *Tubercle and Lung Disease* 73, 225-231. 1992.

31. Bah, B., Massari, V., Sow, O., Siriwardana, M., Camara, L. M., Larouzé, B., and Murray, J. F. Useful clues to the presence of smear-negative pulmonary tuberculosis

in a West African city. *International Journal of Tuberculosis and Lung Disease* 6(7), 592-598. 2002.

32. Samb, B., Sow, P. S., Kony, S., Maynart-Badiane, M., Diouf, G., Cissokho, S., Bâ, D., Sané, M., Klotz, F., Faye-Niang, M. A., Mboup, S., Ndoye, I., Delaporte, E., Hane, A. A., Samb, A., Coulaud, J.-P., Coll-Seck, A. M., Larouzé, B., and Murray, J. F. Risk factors for negative sputum acid-fast bacilli smears in pulmonary tuberculosis: results from Dakar, Senegal, a city with low HIV seroprevalence. *International Journal of Tuberculosis and Lung Disease* 3(4), 330-336. 1999.

33. Styblo K, Danková D, and Drápela J. Epidemiological and clinical study of tuberculosis in the district of Kolin, Czechoslovakia. *Bulletin of the World Health Organization* 1967;37:819-74.

34. Toman K. *Tuberculosis Case-finding and Chemotherapy.* Geneva: World Health Organization, 1979.

35. Palmieri, F., Giradi, E., Pellicelli, A. M., Rianda, A., Bordi, E., Busi Rizzi, E., Petrosillo, N., and Ippolito, G. Pulmonary tuberculosis in HIV-infected patients presenting with normal chest radiograph and negative sputum smear. *Infection* 30(2), 68-74. 2002.

36. Marciniuk, D. D., McNab, B. D., Martin, W. T., and Hoeppner, V. H. Detection of pulmonary tuberculosis in patients with a normal chest radiograph. *Chest* 115(2), 445-452. 1999.

37. Kanaya, A. M., Glidden, D. V., and Chambers, H. F. Identifying pulmonary tuberculosis in patients with negative sputum smear results. *Chest* 120, 349-355. 2001.

38. Steen, T. W. and Mazonde, G. N. Pulmonary tuberculosis in Kweneng district, Botswana: delays in diagnosis in 212 smear-positive patients. *International Journal of Tuberculosis and Lung Disease* 2(8), 627-634. 1998.

39. Lawn, S. D., Afful, B., and Acheampong, J. W. Pulmonary tuberculosis: diagnostic delay in Ghanaian adults. *International Journal of Tuberculosis and Lung Disease* 2(8), 635-640. 1998.

40. Wandwalo, E. R. and Mørkve, O. Delay in tuberculosis case-finding and treatment in Mwanza, Tanzania. *International Journal of Tuberculosis and Lung Disease* 4(2), 133-138. 2000.

41. Sherman, L. F., Fujiwara, P. I., Cook, S. V., Bazerman, L. B., and Frieden, T. R. Patient and health care system delays in the diagnosis and tratment of tuberculosis. *International Journal of Tuberculosis and Lung Disease* 3(12), 1088-1095. 1999.

42. Salaniponi, F. M., Harries, A. D., Banda, H. T., Kang'ombe, C., Mphasa, N., Mwale, A., Upindi, B., Nyirenda, T. E., Banerjee, A., and Boeree, M. J. Care seeking behaviour and diagnostic processes in patients with smear-positive pulmonary tuberculosis in Malawi. *International Journal of Tuberculosis and Lung Disease* 4(4), 327-332. 2000.

43. Salaniponi, F. M., Gausi, F., Kwanjana, J. H., and Harries, A. D. Time between sputum examination and treatment in patients with smear-negative pumonary tuberculosis. *International Journal of Tuberculosis and Lung Disease* 4(6), 581-583. 2000.

44. Yilmaz, A., Boga, S., Sulu, E., Durucu, M., Yilmaz, D., Baran, A., and Poluman, A. Delays in the diagnosis and treatment of hospitalized patients with smear-positive pulmonary tuberculosis. *Respiratory Medicine* 95, 802-805. 2001.

45. Creek, T. L., Lockman, S., Kenyon, T. A., Makhoa, M., Chimidza, N., Moeti, T., Sarpong, B. B., Binkin, N. J., and Tappero, J. W. Completeness and timeliness of treatment initiation after laboratory diagnosis of tuberculosis in Gaborone, Botswana. *International Journal of Tuberculosis and Lung Disease* 4(10), 956-961. 2000.

46. Liam, C. K. and Tang, B. G. Delay in the diagnosis and treatment of pulmonary tuberculosis in patients attending a university teaching hospital. *International Journal of Tuberculosis and Lung Disease* 1(4), 326-332. 1997.

47. Pronyk, P. M., Makhubele, M. B., Hargreaves, J. R., Tollman, S. M., and Hausler, H. P. Assessing health seeking behaviour among tuberculosis patients in rural South Africa. *International Journal of Tuberculosis and Lung Disease* 5(7), 619-627. 2001.

48. Fourie, P. B., Gatner, E. M. S., Glatthaar, E., and Kleeberg, H. H. Follow-up tuberculosis prevalence survey of Transkei. *Tubercle* 61, 71-79. 1980.

49. Fourie, P. B. and Knoetze, K. Prevalence of tuberculosis in Botswana and Lesotho: Results of two random sample surveys. *The Southern African Journal of Epidemiology and Infection* 1, 32-37. 1986.

50. Reddy, C. N. *Targeted Tuberculin Testing and the Prevalence of Latent Tuberculosis Infection (LTBI) in Health Care Workers at a TB Referral Hospital, Durban.* 2003. The College of Public Health Medicine of South Africa.

51. Borgdorff, M. W., Egwaga, S. M., Muwinge, H., Kalisvaart, N. A., and Nagelkerke, N. J. D. Density of tuberculin skin reaction and its association with tuberculosis infection as defined by dual tuberculin testing. *International Journal of Tuberculosis and Lung Disease* 4(11), 1060-1065. 2000.

52. Lifson, A. R., Grant, S. M., Lorvick, J., Pinto, F. D., He, H., Thompson, S., Keudell, E. G., Stark, M. J., Booth, R. E., and Watters, J. K. Two-step tuberculin skin testing of injection drug users recruited from community-based settings. *International Journal of Tuberculosis and Lung Disease* 1(2), 128-134. 1997.

53. Hecker, M. T., Johnson, J. L., Whalen, C. C., Nyole, S., Mugerwa, R. D., and Ellner, J. J. Two-step tuberculin skin testing in HIV-infected persons in Uganda. *American Journal of Respiratory and Critical Care Medicine* 155, 81-86. 1997.

54. Webster, C. T., Gordin, F. M., Matts, J. P., Korvick, J. A., Miller, C., Muth, K., Brown, L. S., Besch, C. L., Kumi, J. O., Salveson, C., Bincsik, A., Irvin, K., and Maiatico, G. Two-stage tuberculin skin testing in individuals with human immunodeficiency virus infection. *American Journal of Respiratory and Critical Care Medicine* 151, 805-808. 1995.

55. Johnsen, C. Evaluation of two-step tuberculin testing in a Massachusetts correctional facility. *American Journal of Infection Control* 23(3), 209-212. 1995.

56. Bierrenbach, A. L., Floyd, S., Cunha, S. C., Dourado, I., Barreto, M. L., Pereira, S. M., Hijjar, M. A., and Rodrigues, L. C. A comparison of dual skin test with mycobacterial antigens and tuberculin skin test alone in estimating prevalence of Mycobacterium tuberculosis infection from population surveys. *International Journal of Tuberculosis and Lung Disease* 7(4), 312-319. 2003.

57. Hersh, A. L., Tosteson, A. N., and von Reyn, C. F. Dual skin testing for latent tuberculosis infection. A decision analysis. *American Journal of Preventive Medicine* 24(3), 254-259. 2003.

58. Tala-Heikkilä, M., Lemma, E., and Stanford, J. L. Comparitive study of skin testing with PPD and new tuberculins by the WHO Mantoux test. *Tubercle and Lung Disease* 73, 330-336. 1992.

59. Mathur, M. L. and Raman, S. Potential utility of BCG test in excluding tuberculosis in adults. *Respiratory Medicine* 97, 528-532. 2003.

60. World Health Organization. *Ninth Report of the WHO Expert Committee on Tuberculosis.* 52. 1974. WHO Technical Report Series.

61. Barnes PF, Verdegem TD, Vachon LA, Leedom JM, and Overturf GD. Chest roentgenogram in pulmonary tuberculosis: new data on an old test. *Chest* 1998;94:316-20.

62. Markey AC, Forster SM, Mitchell R, Larson E, and Smith H. Suspected cases of pulmonary tuberculosis referred from port of entry into Great Britain, 1980 - 1983. *British Medical Journal* 1986;292:378.

63. Garland, L. H. Studies on the accuracy of diagnostic procedures. *American Journal of Roentgenology, Radiation Therapy and Nuclear Medicine* 82, 25-38. 1959.

64. Nyboe J. Results of the international study on X-ray classification. *Bulletin of the International Union Against Tuberculosis* 1968;41:115-24.

65. Graham, S., das Gupta, K., Hidvegi, R. J., Hanson, R., Kosiuk, J., al Zahrani, K., and Menzies, D. Chest radiograph abnormalities associated with tuberculosis: reproducibility and yield of active cases. *International Journal of Tuberculosis and Lung Disease* 6(2), 137-142. 2002.

66. du Toit, G., Swingler, G., and Iloni, K. Observer variation in detecting lymphadenopathy on chest radiography. *International Journal of Tuberculosis and Lung Disease* 6(9), 814-817. 2002.

67. Kang, E-Y, Staples, C. A., McGuinness, G., Primack, S. L., and Müller, N. L. Detection and differential diagnosis of pulmonary infections and tumours in patients with AIDS: value of chest radiography versus CT. *American Journal of Roentgenology* 166, 15-19. 1996.

68. Gatner, E. M. S. and Burkhardt, K. R. Correlation of the results of X-ray and sputum culture in tuberculosis prevalence surveys. *Tubercle* 61, 27-31. 1980.

69. Balasangameshwara, V. H. and Chakraborty, A. K. Validity of case-finding tools in a national tuberculosis programme. *Tubercle and Lung Disease* 74, 52-58. 1993.

70. Johnson, J. L., Vjecha, M. J., Okwera, A., Hatanga, E., Byekwaso, F., Wolski, K., Aisu, T., Whalen, C. C., Huebner, R., Mugerwa, R. D., and Ellner, J. J. Impact of human immunodeficiency virus type-1 infection on the initial bacteriologic and radiographic manifestations of pulmonary tuberculosis in Uganda. *International Journal of Tuberculosis and Lung Disease* 2(5), 397-404. 1997.

71. Churchyard, G. J., Roux, S., Mahlatsi, N., Mosenogi, P., Khotle, L., Grant, A. D., Fielding, K., Corbett, E. L., and Hayes, R. J. *A Randomized Controlled Study Of The Effectiveness Of Annual And 6-Monthly Screening With Mass Miniature Radiography (MMR) For The Active Case-Finding Of Cardiopulmonary TB Patients.* GEN524. 2003. Johannesburg, SIMRAC.

72. Enarson DA, Jentgens H, Oberhoffer M, Rieder H, Rouillon A, and Salomao A. *Tuberculosis Guide for High Prevalence Countries.* Paris: International Union Against tuberculosis and Lung Disease, 1990.

73. Maher D, Chaulet P, Spinaci S, and Harries A. *Treatment of Tuberculosis: Guidelines for National Programmes.* Geneva: World Health Organization, 1997.

74. Foulds, J. and O'Brien, R. New tools for the diagnosis of tuberculosis: the perspective of developing countries. *International Journal of Tuberculosis and Lung Disease* 2(10), 778-783. 1998.

75. Perkins, M. D. New diagnostic tools for tuberculosis. *International Journal of Tuberculosis and Lung Disease* 4(12), s182-s188. 2000.

76. Wu, Z-L and Wang, A-Q. Diagnostic yield of repeated smear microscopy examinations among patients suspected of pulmonary TB in Shandong province of China. *International Journal of Tuberculosis and Lung Disease* 4(11), 1086-1087. 2000.

77. Harries, A. D., Mphasa, N. B., Mundy, C., Banerjee, A., Kwanjana, J. H., and Salaniponi, F. M. L. Screening tuberculosis suspects using two sputum smears. *International Journal of Tuberculosis and Lung Disease* 4(1), 36-40. 2000.

78. van Deun, A., Hamid Salim, A., Cooreman, E., Hossain, A., Rema, A., Chambugonj, N., Hye, A., Kawria, A., and Declercq, E. Optimal tuberculosis case detection by direct sputum smear microscopy: how much better is more? *International Journal of Tuberculosis and Lung Disease* 6(3), 222-230. 2002.

79. Karstaedt, A. S., Jones, N., Khoosal, M., and Crewe-Brown, H. H. The bacteriology of pulmonary tuberculosis in a population with high human immunodeficiency virus seroprevalence. *International Journal of Tuberculosis and Lung Disease* 2(4), 312-316. 1998.

80. Parry, C. M., Kamoto, O., Harries, A. D., Wirima, J. J., Nyirenda, T. E., Nyangulu, D. S., and Hart, C. A. The use of sputum induction for establishing a diagnosis in patients with suspected pulmonary tuberculosis in Malawi. *Tubercle and Lung Disease* 76, 72-76. 1995.

81. Li, L. M., Bai, L. Q., Yang, H. L., Xiao, C. F., Tang, R. Y. h, Chen, Y. F., Chen, S. M., Liu, S. S., Zhang, S. N., Ou, Y. H., and Niu, T. I. M. Sputum induction to improve the diagnostic yield in patients with suspected pulmonary tuberculosis. *International Journal of Tuberculosis and Lung Disease* 3(12), 1137-1139. 1999.

82. Merrick, S. T., Sepkowitz, K. A., Walsh, J., Damson, L., McKinley, P., and Jacobs, J. L. Comparison of induced versus expectorated sputum for diagnosis of pulmonary tuberculosis by acid-fast smear. *American Journal of Infection Control* 25(6), 463-466. 1997.

83. Ba, F. and Rieder, H. L. A comparison of fluorescence microscopy with the Ziehl-Neelsen technique in the examination of sputum for acid-fast bacilli. *International Journal of Tuberculosis and Lung Disease* 3(12), 1101-1105. 1999.

84. Harries, A. D., Nyirenda, T. E., Banerjee, A., Boeree, M. J., and Salaniponi, F. M. The diagnosis of smear-negative pulmonary tuberculosis: the practice of sputum smear examination in Malawi. *International Journal of Tuberculosis and Lung Disease* 3(10), 896-900. 1999.

85. Harries, A. D., Nyirenda, T. E., Banerjee, A., Mundy, C., and Salaniponi, F. M. District sputum smear microscopy services in Malawi. *International Journal of Tuberculosis and Lung Disease* 2(1), 914-918. 1998.

86. Ängeby, K. A. K., Alvarado-Gálvez, C., Pineda-GarcÍa, L., and Hoffner, S. E. Improved sputum microscopy for a more sensitive diagnosis of pulmonary tuberculosis. *International Journal of Tuberculosis and Lung Disease* 4(7), 684-687. 2000.

87. van Deun A, Chambugonj N, Hye A, and Hossain A. Rapid fading of carbolfuchsin stained AFB smears under extreme conditions of temperature and humidity (Letter). *International Journal of Tuberculosis and Lung Disease* 1997;1:384-6.

88. Banda, H. T., Harries, A. D., Boeree, M. J., Nyirenda, T. E., Banerjee, A., and Salaniponi, F. M. Viability of stored sputum specimens for smear microscopy and culture. *International Journal of Tuberculosis and Lung Disease* 4(3), 272-274. 2000.

89. van Deun, A., Roorda, F. A., Chambugonj, N., Hye, A., and Hossain, A. Reproducibility of sputum smear examination for acid-fast bacilli: practical problems met during cross-checking. *International Journal of Tuberculosis and Lung Disease* 3(9), 823-829. 1999.

90. Hurtig, A. K., Pande, S. B., Baral, S. C., Porter, J. D. H., and Bam, DS. Sputum examination for acid-fast bacilli in private laboratories, Kathmandu Valley, Nepal. *International Journal of Tuberculosis and Lung Disease* 3(11), 1009-1014. 1999.

91. van Deun, A. and Potaels, F. Limitations and requirements for quality control of sputum smear microscopy for acid-fast bacilli. *International Journal of Tuberculosis and Lung Disease* 2(9), 756-765. 1998.

92. Nguyen, T. N. L. n, Wells, C. D., Binkin, N. J., Pham, D. L., and Nguyen, V. C. The importance of quality control of sputum smear microscopy: the effect of reading errors on treatment decisions and outcomes. *International Journal of Tuberculosis and Lung Disease* 3(6), 483-487. 1999.

93. Lan, N. T. N., Wells, C. D., Binkin, N. J., Becerra, J. E., Linh, P. D., and Cô, N. V. Quality control of smear microscopy for acid-fast bacilli: the case for blinded re-reading. *International Journal of Tuberculosis and Lung Disease* 3(1), 55-61. 1999.

94. Aziz, M. and Bretzel, G. Use of a standardised checklist to assess peripheral sputum smear microscopy laboratories for tuberculosis diagnosis in Uganda. *International Journal of Tuberculosis and Lung Disease* 6(4), 340-349. 2002.

95. Rieder, H. L., Arnadottir, T., Tardencilla Gutierrez, A. A., Kasalika, A. C., Salaniponi, F. M. L., Ba, F., Diop, A. H., Anagonou, S., Gninafon, M., Ringdal, T., Trébucq, A., and Enarson, D. A. Evaluation of a standardized recording tool for sputum smear microscopy for acid-fast bacilli under routine conditions in low income countries. *International Journal of Tuberculosis and Lung Disease* 1(4), 339-345. 1997.

96. Long, R., Scalcini, M., Manfreda, J., Jean-Baptiste, M., and Hershfield, E. The impact of HIV on the usefulness of sputum smears for the diagnosis of tuberculosis. *American Journal of Public Health* 81, 1326-1328. 1991.

97. Githui, W., Nunn, P., and Juma, E. Cohort study of HIV positive and HIV-negative tuberculosis, Nairobi, kenya: comparison of bacteriological results. *Tubercle and Lung Disease* 73(4), 203-209. 1992.

98. Elink Schuurman, M. W., Srisaenpang, S., Pinitsoontorn, S., Bijleveld, I., Vaeteewoothacharn, K., and Methapat, C. The rapid village survey in tuberculosis control. *Tubercle and Lung Disease* 77, 549-554. 1996.

99. Harper, I., Fryatt, R., and White, A. Tuberculosis case finding in remote mountainous areas - are microscopy camps of any value? Experience from Nepal. *Tubercle and Lung Disease* 77, 384-388. 1996.

100. Kimerling, M. E., Shakes, C. F., Carlisle, R., Lok, K. H., Benjamin, W. H., and Dunlap, N. E. Spot sputum screening: evaluation of an intervention in two homeless shelters. *International Journal of Tuberculosis and Lung Disease* 3(7), 613-619. 1999.

101. Drobniewski, F. A., Caws, M., Gibson, A., and Young, D. Modern laboratory diagnosis of tuberculosis. *The Lancet Infectious Diseases* 3, 141-147. 2003.

102. Aber, V. N., Allen, B. W., and Mitchison, D. A. Quality control in tuberculosis bacteriology. Laboratory studies on positive cultures and the efficiency of direct smear examination. *Tubercle* 61(123), 133. 1980.

103. Apers, L., Mutsvangwa, J., Magwenzi, J., Chigara, N., Butterworth, A., Mason, P., and van der Stuyft, P. A comparison of direct microscopy, the concentration method and the Mycobacteria Growth Indicator Tube for the examination of sputum for acid-fast bacilli. *International Journal of Tuberculosis and Lung Disease* 7(4), 376-381. 2003.

104. Chan, E. D., Heifets, L., and Iseman, M. D. Immunological diagnosis of tuberculosis: a review. *Tubercle and Lung Disease* 80(3), 131-140. 2000.

105. Fodor, T. Detection of mycobacteria in sputum smears prepared by cytocentrifugation and sedimentation. *Tubercle and Lung Disease* 76, 273-274. 1995.

106. Perera, J. and Arachchi, D. M. The optimum relative centrifugal force and centrifugation time for improved sensitivity of smear and culture for detection of *Mycobacterium tuberculosis* from sputum. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 93, 405-409. 1999.

107. van Deun, A., Maug, A. K., Cooreman, E., Hossain, A., Chambuganj, N., Rema, V., Marandi, H., Kawria, A., and Portaels, F. Bleach sedimentatin method for increased sensitivity of sputum smear microscopy: does it work? *International Journal of Tuberculosis and Lung Disease* 4(4), 371-376. 2000.

108. Kim, S. Chan., Kang, S. I., Kim, D. W., Kim, S. Cheol., Cho, S-N., Hwang, J. H. H., Kim, Y., Song, S-D., and Kim, Y. H. Development and evaluation of an automated stainer for acid-fast bacilli. *Medical Engineering and Physics* 25, 341-347. 2003.

109. Veropoulos, K., Learmonth, G., Campbell, C., Knight, B., and Simpson, J. Automated identification of tubercle bacilli in sputum. A preliminary investigation. *Analytical and Quantitative Cytology and Histology* 21(4), 277-281. 1999.

110. Somoskövi, A., Györi, Z., Czoboly, N., and Magyar, P. Application of a computer-directed automated microscope in mycobacteriology. *International Journal of Tuberculosis and Lung Disease* 3(4), 354-357. 1999.

111. FIND. FIND Home Page. http://www.finddiagnostics.org/index.htm . 2004.

112. Casal, M., Gutierrez, J., and Vaquero, M. Comparative evaluation of the mycobacteria growth indicator tube with the BACTEC 460 TB system and Lowenstein-Jensen medium for isolation of mycobacteria from clinical specimens. *International Journal of Tuberculosis and Lung Disease* 1(1), 81-84. 1997.

113. Chien, H. P., Yu, M. C., Wu, M. H., Lin, T. P., and Luh, K. T. Comparison of the BACTEC MGIT 960 with Löwenstein-Jensen medium for recovery of mycobacteria from clinical specimens. *International Journal of Tuberculosis and Lung Disease* 4(9), 866-870. 2000.

114. Rivera, A. B., Tupasi, T. E., Grimaldo, E. R., Cardano, R. C., and Co, V. M. Rapid and improved recovery rate of *Mycobacterium tuberculosis* in Mycobacteria Growth Indicator Tube combined with solid Löwenstein Jensen medium. *International Journal of Tuberculosis and Lung Disease* 1(5), 454-459. 1997.

115. Woods, G. L., Fish, G., Plaunt, M., and Murphy, T. Clinical evaluation of Difco ESP culture system II for growth and detection of mycobacteria. *Journal of Clinical Microbiology* 35(1), 121-124. 1997.

116. TK-MEDIUM®. http://www.salubrisinc.com/medicaproductline.html . 2004.

117. Baylan, O., Kisa, O., Albay, A., and Doganci, L. Evaluation of a new automated, rapid, colorimetric culture system using medium for laboratory diagnosis of tuberculosis and determination of anti-tuberculosis drug susceptibility. The *International Journal of Tuberculosis and Lung Disease* 8(6), 772-777. 2004.

118. Woods, G. L. Quality assurance and cost effectiveness in diagnosing tuberculosis with new diagnostic tools (Symposium abstract). *International Journal of Tuberculosis and Lung Disease* 2(11), s173. 1998.

119. Rajalahti, I., Vuorinen, P., Järvenpää, R., and Nieminen, M. M. *Mycobacterium tuberculosis* complex is not detected by DNA amplification assay in sputum specimens of patients with lung scars due to past pulmonary tuberculosis. *International Journal of Tuberculosis and Lung Disease* 7(2), 190-193. 2003.

120. Tortoli, E., Tronci, M., Tosi, C. P., Galli, C., Lavinia, F., Natili, S., and Goglio, A. Multicenter evaluation of two commercial amplification kits (Amplicor, Roche and LCx, Abbott) for direct detection of *Mycobacterium tuberculosis* in pulmonary and extrapulmonary specimens. *Diagnostic Microbiology and Infectious Disease* 33, 173-179. 1999.

121. Mitarai, S., Kurashima, A, Tamura, A., Nagai, H., and Shishido, H. Clinical evaluation of Amplicor *Mycobacterium* detection system for the diagnosis of pulmonary mycobacterial infection using sputum. *Tuberculosis* 81(5/6), 319-325. 2001.

122. Ruiz-Serrano, M. J., Albadalejo, J., Martínez-Sánchez, L., and Bouza, E. LCx: a diagnostic alternative for the early detection of *Mycobacterium tuberculosis* complex. *Mycobacteriology* 32, 259-264. 1998.

123. Araj, G. F., Talhouk, R. S., Itani, L. Y., Jaber, W., and Jamaleddine, G. W. Comparative performance of PCR-based assay versus microscopy and culture for the direct detection of *Mycobacterium tuberculosis* in clinical respiratory specimens in Lebanon. *International Journal of Tuberculosis and Lung Disease* 4(9), 877-881. 2000.

124. Grange, J. M. and Laszlo, A. Serodiagnostic tests for tuberculosis: a need for assessment of their operational predictive accuracy and acceptability. *Bulletin of the World Health Organization* 68(5), 571-576. 1990.

125. Ratanasuwan, W., Kreiss, J. K., Nolan, C. M., Schaeffler, B. A., Suwanagool, S., Tunsupasawasdikul, S., Chuchottaworn, C., Dejsomritrutai, W., and Foy, H. M. Evaluation of the MycoDot™ test for the diagnosis of tuberculosis in HIV seropositive and seronegative patients. *International Journal of Tuberculosis and Lung Disease* 1(3), 259-264. 1997.

126. Somi, G. R., O'Brien, R. J., Mfinanga, G. S., and Ipuge, Y. A. Evaluation of the MycoDot™ test in patients with suspected tuberculosis in a field setting in Tanzania. *International Journal of Tuberculosis and Lung Disease* 3(3), 231-238. 1999.

127. Chan, E. D., Reves, R., Belisle, J. T., Brennan, P. J., and Hahn, W. E. Diagnosis of tuberclosis by a visually detectable immunoassay for lipoarabinomannan. *American Journal of Respiratory and Critical Care Medicine* 161, 1713-1719. 2000.

128. Hamasur, B., Bruchfeld, J., Haile, M., Pawlowski, A., Bjorvatn, B., Källenius, G., and Svenson, S. B. Rapid diagnosis of tuberculosis by detection of mycobacterial lipoarabinomannan in urine. *Journal of Microbiological Methods* 45, 41-52. 2001.

129. Antunes, A., Nina, J., and David, S. Serological screening for tuberculosis in the community: an evaluation of the Mycodot procedure in an African population with high HIV-2 prevalence (Republic of Guinea-Bissau). *Research in Microbiology* 153, 301-305. 2002.

130. Tessema, T. A., Bjune, G., Hamasur, B., Svenson, S., Syre, H., and Bjorvatn, B. Circulating antibodies to lipoarabinomannan in relation to sputum microscopy, clinical features and urinary anti-lipoarabinomannan detection in pulmonary tuberculosis. *Scandinavian Journal of Infectious Diseases* 34, 97-103. 2002.

131. Andersen, P., Munk, M. E., Pollock, J. M., and Doherty, T. M. Specific immune-based diagnosis of tuberculosis. *The Lancet* 356, 1099-1104. 9-23-2000.

132. Nakamura, R. M., Velmonte, M. A., Kawajiri, K., Ang, C. F., Frias, R. A., Mendoza, M. T., Montoya, J. C., Honda, I., Haga, S., and Toida, I. MPB64 mycobacterial antigen: a new skin-test reagent through patch method for rapid diagnosis of active tuberculosis. *International Journal of Tuberculosis and Lung Disease* 2(7), 541-546. 1998.

133. Nakamura, R. M., Einck, L., Velmonte, M. A., Kawajiri, K., Ang, C. F., Delasllagas, C. E., and Nacy, C. A. Detection of active tuberculosis by an MPB-64 transdermal patch: a field study. *Scandinavian Journal of Infectious Diseases* 33(6), 405-407. 2001.

134. Summary of safety and effectiveness data. www.fda.gov/cdrh/pdf/p0 10033b.pdf . 2002.

135. Streeton, J. A., Desem, N., and Jones, S. L. Sensitivity and specificity of a gamma interferon blood test for tuberculosis infection. *International Journal of Tuberculosis and Lung Disease* 2(6), 443-450. 1998.

136. Mazurek, G. H., Lo Bue, P. A., Daley, C. L., Bernardo, J., Lardizabal, A. A., Bishai, W. R., Lademarco, M. F., and Rothel, J. S. Comparison of a whole blood interferon gamma assay with tuberculin skin testing for detecting latent Mycobacterium tuberculosis infection. *Journal of the American Medical Association* 286(14), 1740-1747. 2001.

137. Takiff, H. and Heifets, L. In search of rapid diagnosis and drug-resistance detection tools: Is the *FASTPlaque*TB test the answer? (Editorial). *International Journal of Tuberculosis and Lung Disease* 6(7), 560-561. 2002.

138. Albert, H., Heydenrych, A., Brookes, R., Mole, R. J., Harley, B., Subotsky, E., Henry, R., and Azevedo, V. Performance of a rapid phage-based test, *FASTPlaque*TB™, to diagnose pulmonary tuberculosis from sputum specimens in South Africa. *International Journal of Tuberculosis and Lung Disease* 6(6), 529-537. 2002.

139. Muzaffar, R., Batool, S., Aziz, F., Naqvi, A., and Rizvi, A. Evaluation of the *FASTPlaque*TB assay for direct detection of *Mycobacterium tuberculosis* in sputum specimens. *International Journal of Tuberculosis and Lung Disease* 6(7), 635-640. 2002.

140. Albay, A., Kisa, O., Baylan, O., and Doganci, L. The evaluation of FAST*Plaque*TB™ test for the rapid diagnosis of tuberculosis. *Diagnostic Microbiology and Infectious Disease*. 2003.

141. Marchant, A. Serological markers of disease activity in tuberculosis and HIV infection (Editorial Revue). *Clinical and Experimental Immunology* 122(10), 12. 2000.

142. Flynn, J. L., Goldstein, M. M., Triebold, K. J., Koller, B., and Bloom, B. R. Major histocompatibility complex class I-restricted T cells are required for resistance to *Mycobacterium tuberculosis* infection. *Proceedings of the National Academy of Science* USA 89(24), 12013-12017. 1992.

143. Piwowar, E. M., Tugume, S. B., Grant, R. M., Lutalo, T., Pattishall, K., and Katongole-Mbidde, E. Beta-2 microglobulin values among human immunodeficiency virus (HIV)-negative, HIV-positive asymptomatic, and HIV-positive symptomatic Ugandans. *Clinical and Diagnostic Laboratory Immunology* 2(2), 236-237. 1995.

144. Wanchu, A., Arora, S., Bhatnagar, A., Sud, A., Bambery, P., and Singh, S. Decline in beta-2 microglobulin levels after anti-tubercular therapy in tubercular patients with HIV. *The Indian Journal of Chest Diseases and Allied Sciences* 43(4), 211-215. 2001.

APPENDIX 1

Pre- and Post- workshop questionnaires

PRE-WORKSHOP OPINION POLL

NICKNAME: _____

Are you employed in the SA mining industry at least 30% of your time?

YES	NO

In you opinion, which of the following TB diagnostic methods warrant(s) further investigation as **screening** method(s) for TB on South African mines? Please choose an option for each method listed by marking the appropriate boxes beneath each method.

1.   Questionnaires about symptoms

STRONGLY AGREE					STRONGLY DISAGREE

2.   Skin tests

STRONGLY AGREE					STRONGLY DISAGREE

3.   Chest X-rays

STRONGLY AGREE					STRONGLY DISAGREE

4.   Sputum examinations (smears and/or cultures)

STRONGLY AGREE					STRONGLY DISAGREE

5.   New molecular methods

STRONGLY AGREE					STRONGLY DISAGREE

POST-WORKSHOP OPINION POLL

NICKNAME: _____

In you opinion, which of the following TB diagnostic methods warrant(s) further investigation as **screening** method(s) for TB on South African mines? Please choose an option for each method listed by marking the appropriate boxes beneath each method.

1.   Questionnaires about symptoms

STRONGLY AGREE						STRONGLY DISAGREE

2.   Skin tests

STRONGLY AGREE						STRONGLY DISAGREE

3.   Chest X-rays

STRONGLY AGREE						STRONGLY DISAGREE

4.   Sputum examinations (smears and/or cultures)

STRONGLY AGREE						STRONGLY DISAGREE

5.   New molecular methods

STRONGLY AGREE						STRONGLY DISAGREE

Analysis of the responses to the pre- and post- questionnaires

24 people completed the pre-questionnaire. 11/23 were involved in the mining industry for more than 30% of their time (the 24th did not respond to the item).

21 people completed the post-questionnaire. All could be linked to their responses to the pre-questionnaire through their nicknames.

Median scores for the 5 items:

Question item:	1	2	3	4	5
Pre-questionnaire	1	4	2	2	2
Post-questionnaire	1	5	2	2	5

None of the changes in scores were statistically significant using Wilcoxon's signed rank test except for the changes in responses to questions number 2 ($p = 0.02$) and 5 ($p = 0.0008$).

In other words, the respondents:

- ✍ felt even more strongly at the end of the workshop that further research into skin tests was not warranted and

- ✍ changed their views about novel methods from being fairly in favour at the start of the workshop but firmly against at the end.

The data was then stratified according to whether the respondent spent more than 30% of his/her time working in the mining industry, but there were no differences in the medians for the two strata either in the pre-questionnaire or in the post-questionnaire.

APPENDIX 2

Screening terminology
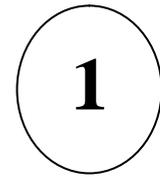
UNPACKING SOME OF THE SCREENING TERMINOLOGY

The tuberculosis screening workshop will be attended by participants with a wide range of prior epidemiological knowledge, from novice researchers and public health students in training, through to experienced practitioners with limited formal training in epidemiology, and experienced researchers of international repute with considerable knowledge and skills.

This set of 1-2 page fact sheets of background information is provided in advance of the workshop mainly to orientate the novice participants prior to the workshop. The fact sheets will, however, be sent out electronically to all participants so that everyone is aware of what has been covered.

As a result, all participants are requested to read through the documentation. For experienced researchers, all that will be required is to glance, briefly, at the headings on each page. For others, whose background in epidemiology is superficial, some time should please be spent studying the contents and making sure that the concepts are understood.

Some of these fact sheets will be included in the final report as appendices for readers of the report who might want some background information of this nature. Others might be incorporated into the actual report at appropriate places. As a result, any comments that participants might have concerning the accuracy of the fact sheets (errors, misinterpretations etc…), style ("unclear" "confusing" etc…) or the amount of detail (too much, too little), or omissions will be very welcome. Please send these comments to hsct@icon.co.za.

SCREENING DEFINITION AND PURPOSE

Screening for a disease is a set of activities carried out in apparently healthy (i.e. asymptomatic) populations (with regard to the disease) in order to classify the population into those who might, and those who probably do not, have the disease. In practice many of the people who take part in a screening programme are in fact symptomatic and may not be correctly called "apparently healthy": they will, however, usually be undiagnosed prior to the screening.

The following definition is given in Rothman and Greenland, 1998[1]:

> *"Screening for disease control can be defined as the examination of asymptomatic people in order to classify them as likely or unlikely to have the disease that is the object of the screening".*

Those individuals who, after screening, are classified as likely to have the disease, are then usually invited to undergo additional, more definitive, examination and or investigation (referred to as the "definitive" test) in order to determine whether or not they really do have the disease (a proportion will turn out to be disease-free, the so-called "false positives"). These "definitive" tests are also sometimes referred to as "diagnostic" tests. However, since screening tests are also usually diagnostic tests, the term "definitive" will be used in this workshop.

In most textbooks the only purpose of screening is stated to be to detect disease early and, thereby, to improve the outcome for the individual with the disease, by starting treatment earlier (an individual benefit).

However, another purpose might be to identify cases of an infectious disease earlier than would be the case if one were to simply wait for such people to present to a treatment centre, and, thereby, through institution of earlier treatment, limit the spread of the disease in the population (a population benefit). In the context of screening for tuberculosis, both these purposes are relevant.

The rationale behind using a screening test first followed by a definitive diagnostic assessment of those found (as a result of the screening test) to be "likely" to have the disease ("screening test positive"), is: It makes intuitive sense to reserve an expensive, invasive, unpleasant diagnostic test, which may also be difficult to perform, for use in the small sub-set of people who screen positive, rather than try to apply such a test to an entire population.

---

[1] Rothman, K.J., Greenland, S., 1998. *Modern Epidemiology* (2nd edition), page 499. Philadelphia: Lippincott, Williams and Wilkins.

It follows that screening tests should preferably be inexpensive, easy to perform, non-invasive, and acceptable to people who make up the population being screened.

CRITERIA FOR INSTITUTING A SCREENING PROGRAMME

Beaglehole *et al*[2] have described a list of criteria that ought to be met before instituting a screening programme, based on (it is abridged) a much-cited publication by Wilson and Jüngner[3]. This list is presented below:

DISEASE	Serious
	High prevalence of pre-clinical stage
	Natural history understood
	Long period between first signs and overt disease
DIAGNOSTIC (SCREENING[a]) TEST	Sensitive and specific
	Simple and cheap
	Safe and acceptable
	Reliable
	Resources are adequate [a]
DIAGNOSIS AND TREATMENT	Facilities are adequate
	Effective, acceptable, and safe treatment available

a: my additions

Some of these terms, namely "sensitivity", "specificity", and "reliable", will be explained in subsequent fact sheets. Of note here, is the term "effective" that is used in the context of the "Diagnosis and Treatment" part of the above list. Effective is usually taken to mean that something works as intended under operational conditions. "Efficacy" may be excellent under clinical trial conditions, curing 90%+ of patients, but the "Effectiveness" under operational conditions may only be a 65% cure rate. The term "effective" should probably also be applied in the context of the "Diagnostic screening test" in the above table.

The list of criteria above says nothing of the cost-effectiveness of the screening programme. Introduction of a screening programme always entails an opportunity cost, since resources are never limitless. Thus one may want to add that the screening intervention must be a cost-effective use of available resources. Furthermore, if early institution of treatment offers few or no benefits for either the patient or the community, then there may be little point in trying to detect the presence of the disease early through a mass screening programme. Thus one may want to add that delayed treatment of the disease carries unacceptable consequences for the individual or for the community. On the next page, there is a brief account of each of the criteria given in Beaglehole *et al*'s list.

---

[2] Beaglehole, R., Bonita, R., Kjellström, T., 1993. *Basic Epidemiology*, (updated and reprinted 2000), page 93. Geneva: World Health Organization.
[3] Wilson, J.M.G., Jüngner, G., 1968. *Principles and Practice of Screening for Disease.* (*Public Health Papers*, Number 34). Geneva: World Health organization.

Serious

It goes without saying that if one is to commit major resources to screening for a disease then one would only do this if the disease is serious, resulting in significant morbidity or mortality for those who have the disease, as well as adverse consequences for society as a whole.

High prevalence of pre-clinical stage

The prevalence of the pre-clinical stage is higher if the prevalence of the disease is high, and/or if the interval between the onset of the disease and clinical symptoms is long. If a disease has very serious consequences if not detected and treated early, such as some congenital metabolic abnormalities, then this requirement may be considered less important.

Natural history understood

We need to know, if a disease is not treated, or is only treated late, what the adverse consequences will be. Otherwise it will not be possible to weigh the costs (financial and non-financial) of the screening programme against its benefits (financial and non-financial). Furthermore, we need to know what the period is between the time when an average case becomes detectable by the screening test and the time that the patient becomes overtly symptomatic and seeks treatment. Knowledge about the duration of this sub-clinical period will also enable us to determine how frequently we should carry out the screening test.

Long period between first signs and overt disease

If the period between first detectable signs, using the screening test, and the emergence of overt disease, is very short, then screening would have to be carried out very frequently in order to be able to detect a reasonable proportion of cases and intervene before the disease becomes symptomatic. This would make the screening programme impracticable.

Sensitive and specific (two important measures of validity)

The sensitivity of the screening test is simply the proportion of pre-clinical cases that is detected as "positive" by the test. Similarly, the specificity of the screening test is the proportion of non-cases that is identified as "negative" by the screening test. We might be unhappy with a test that only identified 60% of pre-clinical cases (sensitivity = 0.6), or that only identifies 65% of non-cases as being non-cases (specificity = 0.65). If the sensitivity is low then a screening programme will miss many cases and would be less effective. If the specificity is low then we would identify a large number of false positives who would then be needlessly subjected to definitive diagnostic measures, with a loss of efficiency. Either scenario will result in a less attractive screening programme and possible wasted resources. For a particular, given, test, where one can adjust the level of the result that is considered "positive", there is a trade-off between sensitivity and specificity (we cannot improve both simultaneously). Often, for screening tests, it is preferred to have greater sensitivity and accept a loss in specificity. Sensitivity and specificity will be discussed in more detail later.

Simple and cheap

Screening tests that are used in population-based programmes need to be administered to large numbers of people, usually on a regular basis (e.g. once a year or once every 2-3 years. Hence they should be inexpensive. The tests also have to be administered by competent testers and the results have to be read or interpreted by competent people. These personnel may be needed in large numbers, and may be relatively unskilled. Hence, there is a need for simplicity in administration and interpretation of the test.

Safe and acceptable

The screening test is being administered to large numbers of people who are asymptomatic and consider themselves to be healthy. Most of those who have the test will usually be perfectly well. It is important that the test does not, therefore, carry significant risks or side-effects. In addition, the testing procedure should not cause undue discomfort and should be culturally and politically acceptable, or else many of those who need to be screened might refuse to take part.

Reliable

Reliability of a test is a test property whereby the same test, if repeated on the same individual, whose disease state has not changed, will give the same result. Relatively high sensitivity and specificity do not necessarily imply high reliability. This property will be discussed further in a later fact sheet.

Resources are adequate

There should be adequate facilities and competent human resources to carry out the screening programme, the definitive testing and evaluation of those who screen positive, and the treatment of those who are determined to be suffering from the disease. This includes managerial resources for organizing, training, budgeting and monitoring of the programme. This need is often overlooked and when it is overlooked, may lead to the failure of the programme to deliver the expected results.

Facilities are adequate

Where identified patients will need to be admitted to a facility for treatment, are there adequate facilities available for these admissions? If not, then the screening programme may simply detect people with the disease in question, and there would be no possibility to then admit them for their treatment.

Effective, acceptable, and safe treatment available

Whether treatment for the disease in question is carried out in a facility or not, there needs to be an effective treatment option available that is safe and acceptable to the person with the disease. Otherwise the purpose of the screening programme will fail to be achieved.

SENSITIVITY AND SPECIFICITY

Consider a hypothetical situation in which a screening test was administered to 400 people, of whom 200 really do have the disease in question and 200 really do not. The results might be:

RESULTS OF SCREENING FOR 200 PEOPLE WHO HAVE THE DISEASE:			RESULTS OF SCREENING FOR 200 WHO DO NOT HAVE THE DISEASE:	
Screen test +ve =	180		Screen test +ve =	30
Screen test –ve =	20		Screen test –ve =	170
Total of those screened =	200		Total of those screened =	200
SENSITIVITY =	180/200 = 0.90		**SPECIFICITY** =	170/200 = 0.85

In other words, 90% of those who really do have the disease test positive with the screening test (this is the sensitivity of the tests) and 10% (20) are false negatives.

In addition, 85% of those who really do not have the disease test negative with the screening test (this is the specificity of the test) and 15% (30) are false positives.

This information is usually presented in a combined 2X2 table:

		IS DISEASE REALLY PRESENT?		
		YES (+)	NO (-)	
SCREEN TEST RESULT	+VE	a	b	$(a + b)$
	-VE	c	d	$(c + d)$
		$(a + c)$	$(b + d)$	$(a + b + c + d)$

The **SENSITIVITY** of the test is then given by $a/(a + c)$ and

The **SPECIFICITY** of the test is given by $d/(b + d)$

FALSE POSITIVES = b and **FALSE NEGATIVES** = c

POINTS TO REGISTER:

1. It stands to reason that the values for the sensitivity and specificity cannot be < 0 or > 1.

2. If sensitivity is high (i.e. close to 1) the screening test will be able to detect most of the true cases.

3. If sensitivity is low (i.e. close to 0) the test will miss many cases, and screening will not be effective.

4. If the specificity is high (i.e. close to 1) then very few people will be subjected, unnecessarily, to definite investigation for the disease. The test will have excluded them from further consideration.

5. If the specificity is low (i.e. closer to 0) then there will be a high number of false positives who will need to undergo unnecessary definitive testing, a waste of resources and time, a loss of efficiency.

6. The results of a study to determine the sensitivity and specificity of a screening test are calculated from two randomly selected samples of people (known cases and known non-cases) who were then subjected to the screening test. Intuitively, (and this can be justified theoretically), it should be clear that the larger these two samples of testees, the more likely that the resulting estimates for sensitivity and specificity will be close to the true values for the test. But since the estimates are always based on a study involving a sample of those with the disease and a sample of those without the disease, one never knows whether the resulting estimates are exactly correct (this almost never happens) or deviated above or below the real value. As a result, it is desirable that estimates for the sensitivity and specificity should be accompanied by 95% confidence intervals:

    > Sensitivity = 0.9 (95% confidence interval is 0.848 to 0.936)
    > Specificity = 0.85 (95% confidence interval is 0.791 to 0.895)

    This means that there would be 95% certainty that the true sensitivity of the test lies between 0.848 and 0.936. Also, we may be 95% certain that the true test specificity lies between 0.791 and 0.895.

7. Sensitivity and specificity estimates are only truly unbiased if the samples of people tested are chosen at random from a defined population.

8. Sensitivities and specificities are not constant values, except for the defined populations in which they were estimated. The sensitivities and specificities of some of the early HIV ELISA tests, for example, were high in western Europe and north America, but the specificities were often poor in central Africa where exposure to tropical antigens had led to the widespread development of cross-reacting antibodies.

In conclusion:

?   We would like to use a screening test that has high sensitivity and high specificity. It is often not possible to have both as there is a trade-off involved (this will be explained in a later fact sheet). For screening tests, higher sensitivity is usually preferred at the cost of lower specificity. This ensures that we do not miss cases (recall that we only screen for serious diseases).

?   Sensitivities and specificities should, ideally, be estimated using random samples of cases and non-cases drawn from the population (or one that is very similar) that is to undergo the screening. Results should be available and quoted with 95% confidence intervals.

PREDICTIVE VALUES

A test also has two other properties when used in practice, namely the positive and negative predictive values. These properties are of interest when we consider the value of the definitive diagnostic test that is administered to those who are identified by the screening process as being likely to be suffering from the disease in question.

The positive predictive value of a test is the probability that a person who tests positive really does have the disease being tested for. The negative predictive value is the probability that a person who tests negative really does not have the disease in question. As with sensitivities and specificities, these predictive values can only have values between 0 and 1. Unlike sensitivity and specificity these values depend on the prevalence of the disease in the population being tested. The relationship is explained and described mathematically with the use of the same 2X2 table that was used for the explanation of sensitivity and specificity:

IS DISEASE REALLY PRESENT?

		YES (+)	NO (-)	
SCREEN TEST RESULT	+VE	*a*	*b*	*(a + b)*
	-VE	*c*	*d*	*(c + d)*
		(a + c)	*(b + d)*	*(a + b + c + d)*

RECALL THAT:

The **SENSITIVITY** of the test is then given by a/(a + c) and

The **SPECIFICITY** of the test is given by d/(b + d)

FALSE POSITIVES = b and **FALSE NEGATIVES** = c

IN ADDITION, NOW, WE DEFINE THE FOLLOWING TWO MEASURES:

The **POSITIVE PREDICTIVE VALUE (PPV)** = *a/(a + b)*

The **NEGATIVE PREDICTIVE VALUE (NPV)** = *d/(c + d)*

Here is a worked example using the hypothetical results presented in fact sheet number 4, where, if you recall, 200 people with the disease in question, and 200 without the disease, were subjected to a screening test. The results are presented in the table below:

IS DISEASE REALLY PRESENT?

		YES (+)	NO (-)	
SCREEN TEST RESULT	+VE	180	30	210
	-VE	20	170	190
		200	200	400

RECALL THAT:

The **SENSITIVITY** of the test is then given by $a/(a + c)$ = 180/200 = 0.9 and

The **SPECIFICITY** of **the** test is given by $d/(b + d)$ = 170/200 = 0.85

The 95% confidence intervals (I have calculated these using Epi Info) are

Sensitivity = 0.9 (95% confidence interval is 0.848 to 0.936)
Specificity = 0.85 (95% confidence interval is 0.791 to 0.895)

FALSE POSITIVES = b = 30 and **FALSE NEGATIVES** = c = 20

The **POSITIVE PREDICTIVE VALUE (PPV)** = $a/(a + b)$ = 180/210   = 0.8571

The **NEGATIVE PREDICTIVE VALUE (NPV)** = $d/(c + d)$ = 170/190   = 0.8947

The 95% confidence intervals (I have calculated these using Epi Info) are

PPV = 0.8571 (95% confidence interval is 0.801 to 0.900)
NPV = 0.8947 (95% confidence interval is 0.840 to 0.933)

NOTE that the **PREVALENCE** of disease in our combined sample in this case is:

PREVALENCE   =   200/(200 + 200)   =   0.5.

POINTS TO REGISTER ABOUT THE PREDICTIVE VALUES OF A TEST:

1. Once again, predictive values are estimated from testing samples and so we should always present the results alongside their confidence intervals.

2. A high PPV means that if someone tests positive we can be fairly certain that they have the disease in question.

3. A high NPV means that if a person tests negative we can be fairly certain that the person does not have the disease in question.

4. PPV and NPV are not constant values even for a defined population, unless the ratio of those being tested with the disease to those being tested without the disease is constant. In other words, the PREVALENCE of the disease affects the estimate for PPV and NPV.

5. You are invited to convince yourself about the above statement (4) by filling in your own values in the 2X2 table.

   Assuming constant sensitivity of 0.9 and constant specificity of 0.85, start with a prevalence of 20% by making (a + c) = 20 and (b + d) = 80. Then calculate the PPV and NPV.

   Then repeat the exercise for a prevalence of 80% by making (a + c) = 80 and (b + d) = 20. Recalculate the PPV and NPV.

   You will see that as prevalence increases, from 20% to 80%, the PPV increases and the NPV decreases (NPV and PPV always move in opposite directions).

6. When we are told that a test has a "high" PPV we need to ask ourselves what the prevalence of disease was in the group tested. It may be that the researchers have used equal numbers of diseased and non-diseased people (prevalence = 0.5 or 50%). However, in the population that we are screening the prevalence of disease may be 1% or less, so that in practice we would have a much lower PPV and a much higher NPV.

Conclusion:

$\scriptstyle\diagup$ We need to consider the prevalence of disease in the group being tested before we can determine the predictive values of the test.

$\scriptstyle\diagup$ Since the prevalence of the disease in the population is usually low (1%-2% perhaps), even for "common" diseases, the PPV of the screening test will always be low.

$\scriptstyle\diagup$ Screening identifies a sub-population of positives in whom the prevalence of the disease in question will be higher than in the general population.

$\scriptstyle\diagup$ Thus, when the definitive diagnostic test is administered and comes back positive, the PPV will be reasonably high.

3

LIKELIHOOD RATIOS

Thus far we have been introduced to the concepts of the sensitivity and the specificity of a screening test. These two parameters determine the quality of the test, and are constant for a given test used in a defined population. Under operational conditions, however, we would also like to know something about the performance of the test, given the existing prevalence of the disease in the population being screened. We have seen how we can use the concepts of positive predictive value (PPV) and negative predictive value (NPV) in order to do this.

If we know these three values, namely the sensitivity, the specificity, and the prevalence of the disease, then we can work out the PPV and NPV and we will be able to determine the number of people who will have to be subjected to a definitive diagnostic test and also what proportion of these people will eventually be found to have the disease. We will also be able to work out the number of false negatives – the people with the disease who were missed.

However, this is all a bit laborious. Therefore a method has been devised whereby we can, by knowing only 2 values, the so-called likelihood ratio for a positive test result (LR+, defined on page 3), which is a composite value derived from the sensitivity and specificity of the test, in our screened population, and the odds of disease in the population (the "pre-test odds"), simply multiply the two together to obtain the post-test odds of having the disease:

$$\text{pre-test odds} * \text{LR+} = \text{post-test odds}.$$

The pre-test odds of disease is related to the prevalence of the disease, and the post-test odds to the PPV, as follows:

```
                          DISEASE

                      +           -

             +    ┌─────────┬─────────┐
                  │    a    │    b    │   a + b
   TEST           ├─────────┼─────────┤
             -    │    c    │    d    │   c + d
                  └─────────┴─────────┘

                    a + c       b + d     a + b + c + d
```

Pre-test odds	=	(a + c)/(b + d) : 1	=	(a + c) / (b + d)
Prevalence	=	(a+c)/[(a +c) + (b + d)]	=	Pre-test odds / (1 + Pretest odds)
Post-test odds	=	a/b : 1	=	a/b
PPV	=	a/(a + b)	=	Post-test odds / (1 + Post-test odds)

Because we are now simply estimating 1 variable, the post-test odds, from two others, it is possible to construct a simple nomogram (see 4th page) which can be readily used to read off the PPV for any given combination of prevalence and LR+ (where, for convenience and familiarity of use, the pre-test odds is usually replaced by the corresponding pre-calculated prevalence rate and the post-test odds is replaced by the corresponding pre-calculated PPV).

It is argued that this use of a composite of sensitivity and specificity, the LR+, makes life simpler, and means having to know only 2 numbers in order to come to a conclusion about the PPV[4].

It must be stressed that knowing the LR+ value of a test, rather than the Sensitivity and Specificity from which it is derived, adds no new information about the test or its performance. It is simply an <u>*alternative*</u> *way of getting to the PPV.*

A disadvantage of the method using LR+ is that the nomogram cannot incorporate confidence intervals for the PPV value, since these would depend on sample size and sample size is variable between the studies that are used to determine the parameters of the test.

Alternative nomenclatures for sensitivity and specificity

The sensitivity of a screening test is the proportion of those with the disease that test positive using that screening test. Thus the sensitivity could also be called the "true positive rate".

The specificity of a screening test is the proportion of those who do not have a disease that test negative. This could also be called "the true negative rate".

The proportion (1-sensitivity) is the proportion of people with the disease who test negative using the screening test: this could also be called "the false negative rate".

The proportion (1-specificity) is the proportion of people without the disease who test positive using the screening test: this could also be called "the false positive rate".

We may be more satisfied by a test that has a higher true positive rate and a lower false positive rate. The ratio of these two rates would then be high. This ratio is what is termed the "Likelihood ratio for a positive test result" (LR+).

The following definitions are from Sackett *et al*[5]:

THE LIKELIHOOD RATIO FOR A POSITIVE TEST RESULT (LR+)
(Some authors, e.g. Fleiss[6] and Altman[7] refer to this simply as the "Liklihood ratio")

LR+   = True positive rate/False positive rate   = Sensitivity/(1-Specificity)

If a test has a high sensitivity <u>and</u> a high specificity, then the LR+ will be high.

THE LIKELIHOOD RATIO FOR A NEGATIVE TEST RESULT (LR-)

LR-   = False negative rate/True negative rate   = (1-Sensitivity)/Specificity

---

[4] Sackett, D.L., Straus, S.E., Richardson, W.S., Rosenberg, W., Haynes, R.B., 2000. *Evidence-based Medicine* (2nd edition), page 73. Edinburgh: Churchill Livingstone.
[5] Sackett, D.L., Straus, S.E., Richardson, W.S., Rosenberg, W., Haynes, R.B., 2000. *Evidence-based Medicine* (2nd edition), page 73. Edinburgh: Churchill Livingstone.
[6] Fleiss, J, Levin, B., Paik, M. C., 2003. *Statistical Methods for Rates and Proportions* (3rd edition), page 8. Hoboken: Wiley-Interscience.
[7] Altman, D.G., 1991. *Practical Statistics for Medical Research* (1st edition), page 417. Boca Raton: Chapman and Hall/CRC.

A high LR- ratio will result if test Sensitivity and Specificity are both low.

Most of what is currently presented in the literature concerns the use of the LR+ to convert pre-test odds (related to prevalence) to post-test odds (related to PPV). A large number of nomograms, often inter-active, have been developed and are available for a variety of diseases on the www (none found for tuberculosis so-far). Some have also been published in journals and in textbooks.

A nomogram relating NPV to LR- and prevalence has been prepared manually and is presented on page 5. It must be stressed that this is a manual preparation and should not be relied upon for accurate estimations.

The following nomogram for LR+ and PPV is pasted from the www (see nomogram for the address). Note the use of the terms "Pre-test probabilities" and "Post-test probabilities". These are similes for Prevalence and PPV in our context.

Nomogram for converting Pre-test Probabilities to Post-test Probabilities for a diagnostic test result with a given Likelihood Ratio.
(http://www.cebm.net/likelihood_ratios.asp)

A poor screening test with sensitivity of 0.5 and with specificity of 0.5 would have a LR+ of 1. Hence it is suggested that, for our screening test assessment, we confine ourselves to consideration of tests with LR+ values >1.

Pre-Test Probability (%)   **Likelihood Ratio**   **Post-Test Probability (%)**
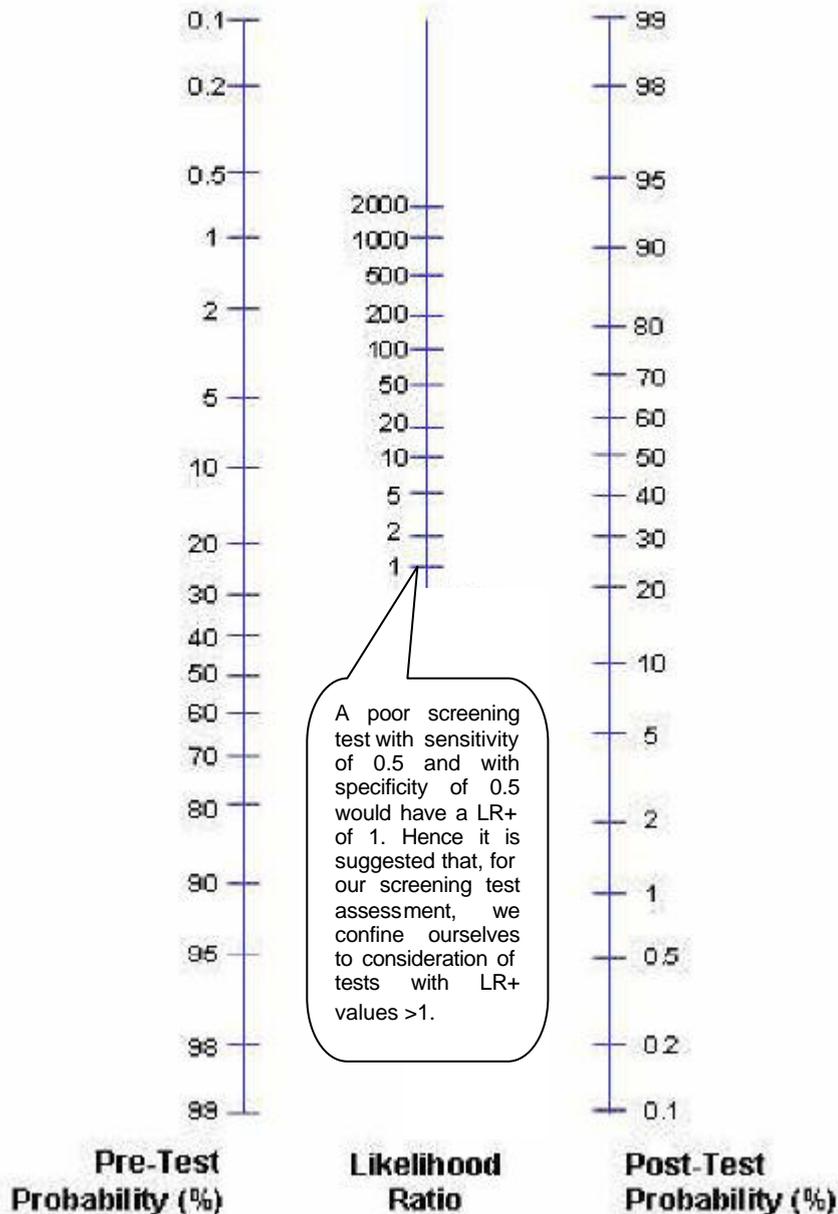
For example, if a test has LR+ of 10 and the prevalence of disease is 1% then by connecting these two points and extending the line we reach a value of about 7% for the "Post-test probability". This means that if 100 people test positive with this test then only 7 of them will be found eventually to have the disease. The other 83 will be false positive results.

NOMOGRAM FOR ESTIMATING NPV FROM LR- AND PREVALENCE

STEP 1	STEP 3	STEP 2
PREVALENCE (%)	READ OFF THE NPV	LR-
50.00	0.50000	1.000
44.26	0.61333	0.794
38.69	0.71523	0.631
33.33	0.80000	0.500
28.32	0.86503	0.395
24.01	0.90210	0.316
20.00	0.94118	0.250
16.67	0.96154	0.200
13.64	0.97564	0.158
11.19	0.98437	0.126
9.09	0.99010	0.100

5
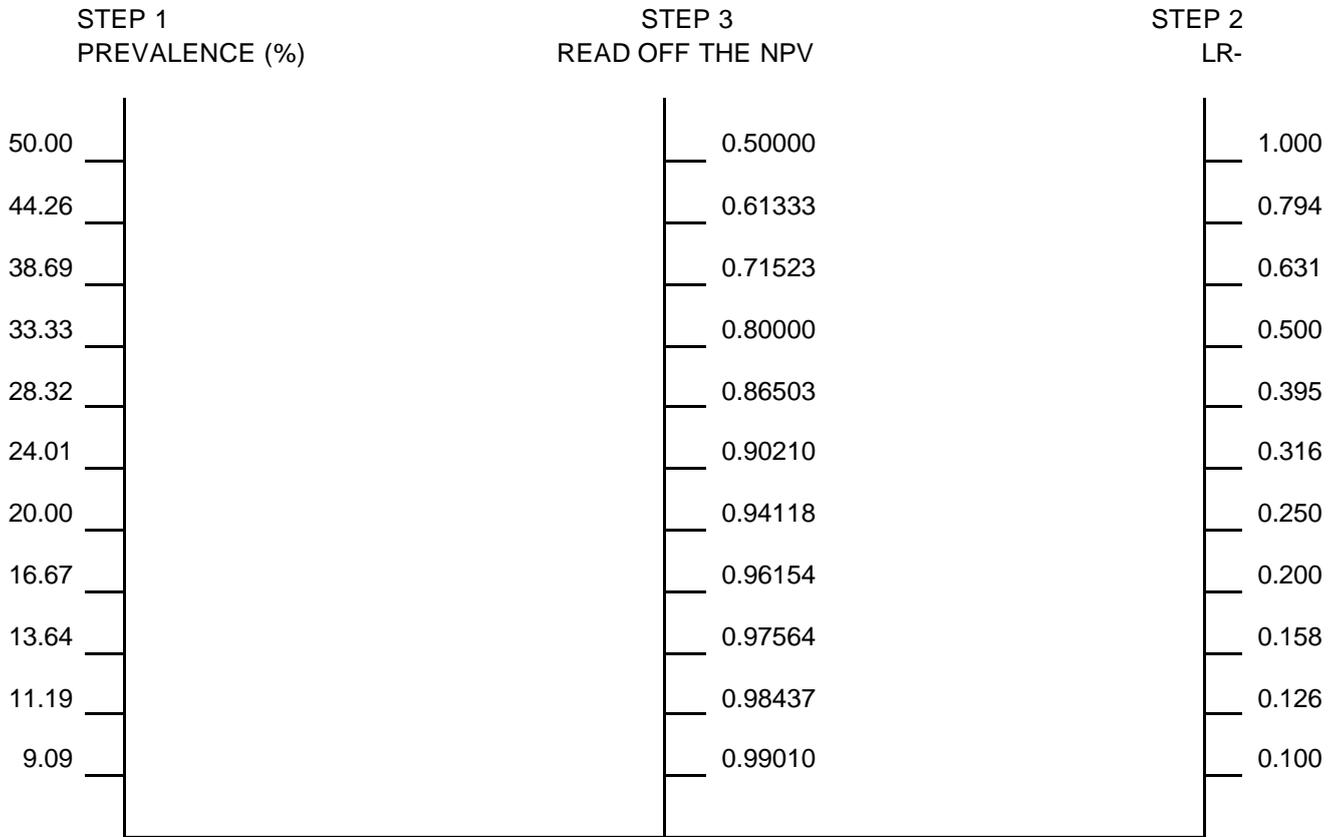
ROC CURVES

The purpose of diagnostic testing, whether used in a screening context or in a definitive diagnostic work-up, is to influence our thoughts about the probability of disease being present. For example, if a person has been coughing for a week we might think that pTB is a possibility, but we would have no reason to think that it is a strong possibility, since many people might have a cough lasting 1 week and most of these would not have pTB.

We might be in possession of statistics that show that only 1% of people who have a cough for 1 week actually have pTB. However, if we subsequently perform a sputum smear examination and it comes back positive for AFBs then our position will change. It may be that 98% of people with sputum smears that are positive for AFBs have active pTB (the other 2% having mycobacteria other than tuberculosis in their sputum).

The process of testing has thus changed our perceptions about the likelihood that the patient has pTB from 1% (the pre-test probability) to 97% (the post-test probability). The previous two fact sheets dealt with the use of either 2X2 table analysis or likelihood ratios in order to work out the post-test probabilities for diagnostic tests. This should be straightforward for tests that are either unequivocally positive or negative, since all those being tested can then be classified as diseased or not on the basis of the test result.

A problem arises, however, when a particular diagnostic test may give a result on a spectrum from one that is definitely negative to one that is definitely positive, with an overlapping in-between. We need to set a "cut-off" value for the test result in such a way that we would classify all those to one side as "positive" and all those to the other side as "negative", otherwise we would be left with a large number of unclassified results.

The cut-off value for a positive PPD (purified protein derivative) skin test for tuberculosis infection offers an example relevant to tuberculosis. The skin test may react subsequent to having had a BCG vaccination in infancy, or subsequent to environmental exposure to other cross-reacting mycobacteria. Or it might react because the person being tested has had an actual infection with *M. tuberculosis*.

The diameter of the skin reaction to the PPD is usually measured in mm. If we were to set the cut-off for those to be considered as infected with *M. tuberculosis* at 2mm we would probably identify a large number of false positives (very high sensitivity and very high negative predictive values would result). On the other hand if we were to set it at 20mm we would undoubtedly miss a very large number of the infected subjects (very high specificity and positive predictive values would result). The rational thing to do might be to set the cut-off value at that level where the smallest number of classification errors will occur. The receiver operating characteristic (ROC) curve and its analysis are used for precisely this purpose.

Consideration of the theory allows us to understand the literature but also to understand the trade-off between the sensitivity and specificity of a test.

ROC curves and their associated analyses are frequently encountered in the literature about diagnostic tests. The term is used to describe the process because ROC curve analysis was first used to establish "cut-off" values to assist in the discrimination between audio signals and noise in such a way as to commit the fewest classification errors while so-doing[8]. A "receiver" (e.g. an air traffic controller) is required to operate (act) according to the classification.

Let us consider a hypothetical diagnostic test, for tuberculosis, that has a wide range of possible numerical results. For simplicity this test has been administered to only 30 people, 16 of whom have the disease and 14 of whom do not. (Normally ROC curve analysis should be performed on results from larger numbers of subjects). A scatter-plot of the results is presented on page 3 as Figure 1.

Referring to Figure 1 it is seen that the 16 subjects with the disease gave results ranging from 1.2 to 2.7 units. The 14 people without the disease gave results ranging from 0.2 to 1.5 units. There is clearly an overlapping range, from 1.2 to 1.5, where both diseased and non-diseased subjects have similar test results.

If we set the cut-off value at point A (1.57 units) then our test will have successfully identified all those who are disease-free as being healthy. However, it will also have incorrectly identified 4/16 diseased subjects as being non-diseased. There will be 4 false negative results.

If, instead, we set the cut-off value at point C (1.08 units), then the test will identify all those who have disease as being diseased, with no false negatives. However we will now have acquired 5 false positives! If the cut-off is set at C (1.37 units) we will have 2 false negatives and 2 false positives. The sensitivities, specificities and predictive values that result for each of these 3 different cut-off points are presented in the following table:

	FALSE +	FALSE -	SENSITIVITY	SPECIFICITY	PPV	NPV
A	0	4	0.75 (12/16)	1.000 (14/14)	1.000 (12/12)	0.778 (14/18)
B	2	2	0.875 (14/16)	0.857 (12/14)	0.875 (14/16)	0.857 (12/14)
C	5	0	1.000 (16/16)	0.643 (9/14)	0.762 (16/21)	1.000 (9/9)

The results in this table clearly demonstrate the trade off between adjusting the cut-off in an attempt to achieve a higher sensitivity (lower specificity) or a higher specificity (lower sensitivity). The "errors" of classification are the sums of the false negatives and false positives. This process is relatively straightforward where we are dealing with such small numbers.

---

[8] Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression* (2nd edition), page 160. New York: John Wiley and Sons.

Figure 1: Scatter-plot of hypothetical test results (14 non-diseased, 16 diseased individuals)
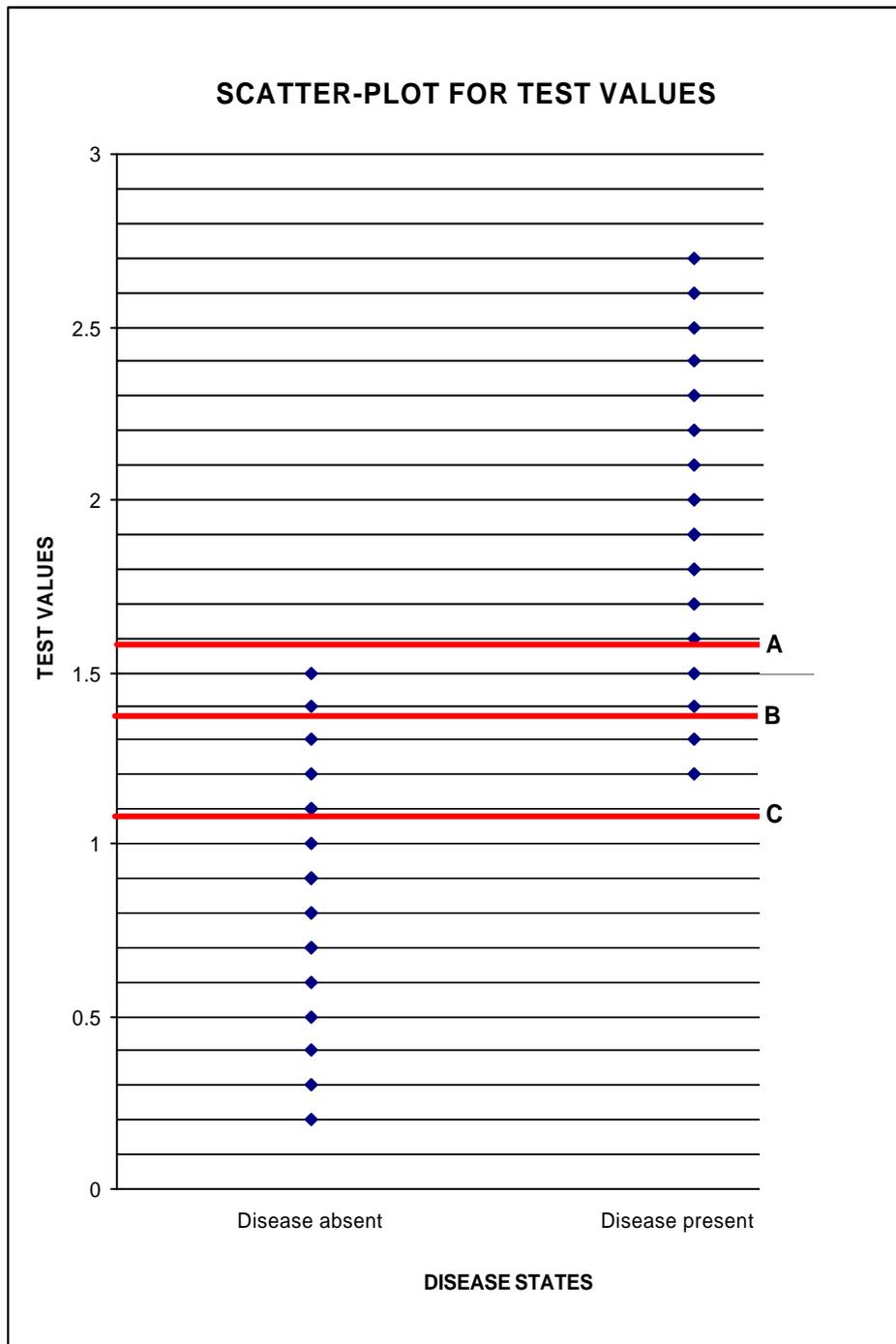
When, as is more usual, we are dealing with large numbers of test results and a more complex overlap between those with the disease, and those without the disease, a method of analysis has been developed that is called the "ROC curve" analysis. In this curve the sensitivity (recall also known as the "true positive rate") is plotted against (1-specificity) (recall also known as the "false positive rate") for each and every possible cut-off value for the data set. A perfect test would be one where there is a possible cut-off value (we will call it C) at which all those who are diseased are to one side (say, above) and all those who are non-diseased are to the other (say, below). Then there would be no false positives and no false negatives if this cut-off (C) is used, and Sensitivity = Specificity = PPV = NPV = 1. LR+ would = ? in such a case, and LR- would = 0.

If we were to plot sensitivity against (1-specificity) for such an ideal data set, then the sensitivity would be 1 for all cut-off points below C, and since specificity would be 1 for all cut-off points above C, (1-specificity) would be zero for all points above C. The resulting ROC "curve" of sensitivity vs. (1-specificity) would thus look like this (dashed rectangular line):

Figure 2: ROC curve for a perfectly discriminating test
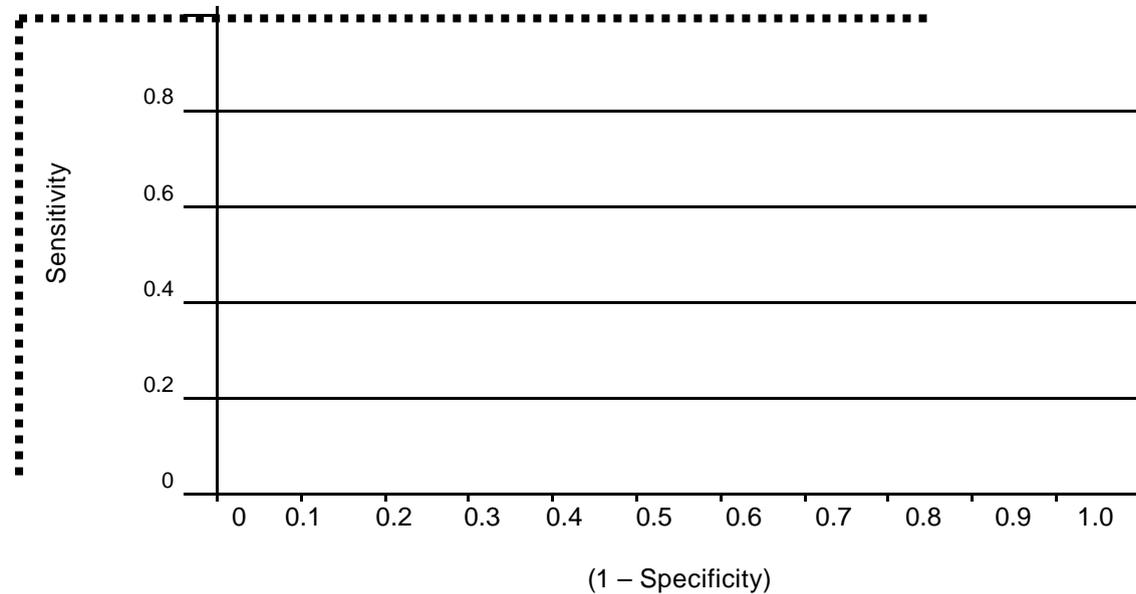
In reality, however, there will be an overlap of the results for those with and without the disease, so that the curve will look more like that in the next Figure (3):

Figure 3: ROC curve for a less-than-perfectly discriminating test
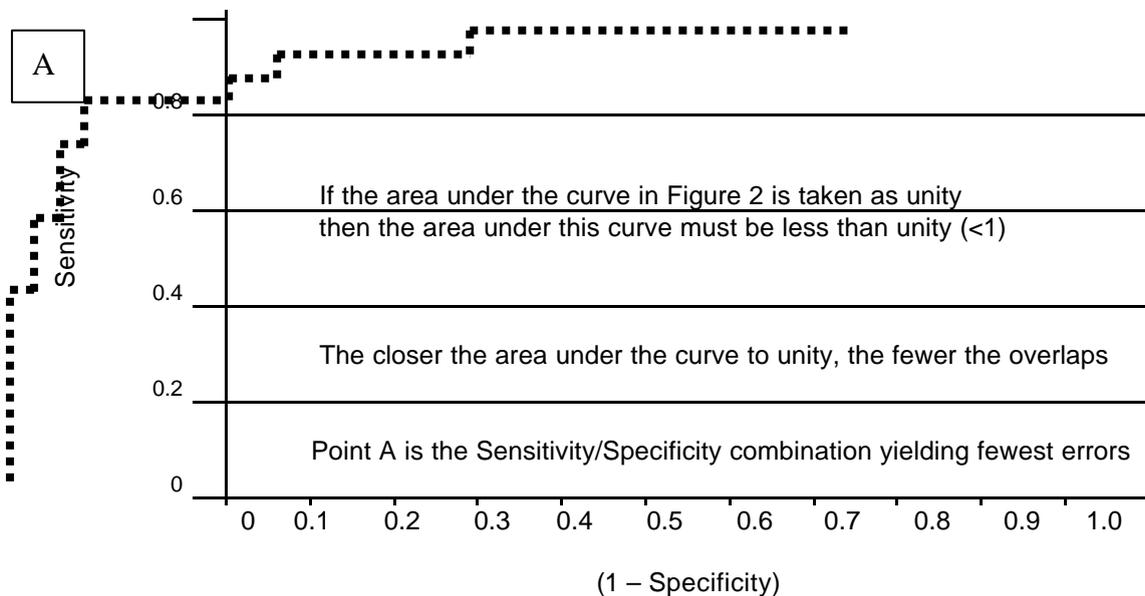
In order to demonstrate this application, the following set of results are presented for a hypothetical data set consisting of overlapping test results for 150 people with a certain disease and 150 people without the disease. The data was analyzed using Stata statistical software.

Figure 4.   Scatter-plot for test results on 150 people with a certain disease and 150 without the disease
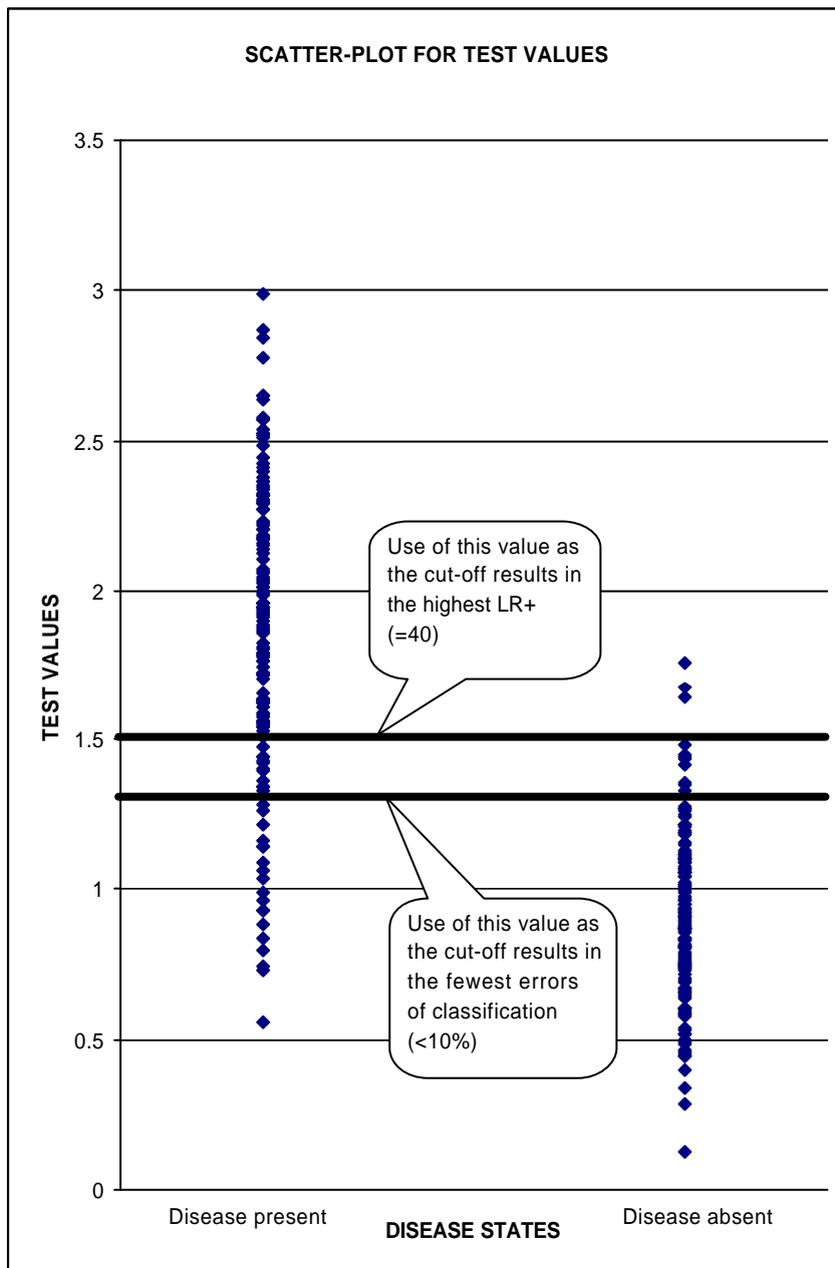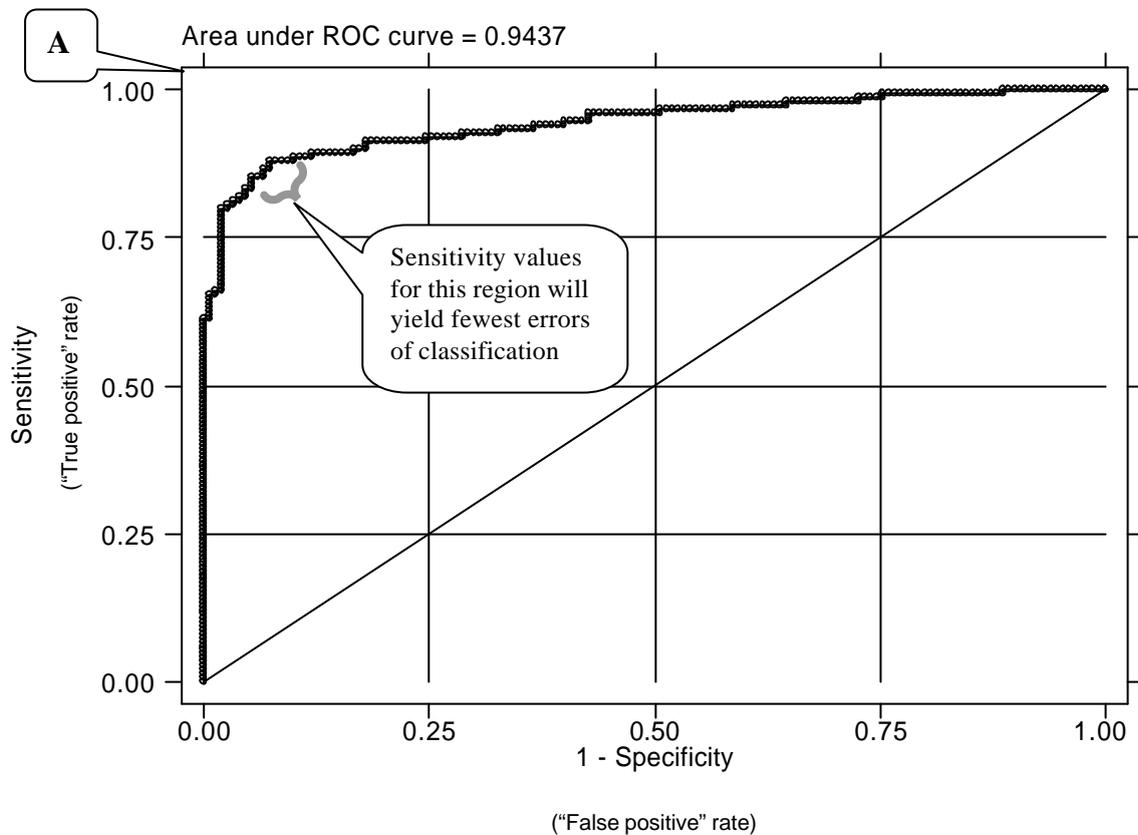
Figure 5.   ROC curve for the hypothetical data set for 150 people with a certain disease and 150 people with the disease portrayed in the scatter-plot of Figure 4

Obs	Area	Std. Err.	[95% Conf. Interval]	
300	0.9437	0.0135	0.91724	0.97022

Stata also provides a table of sensitivities, specificities, percentage of values that are correctly classified, LR+ s and LR- s for each of the different cut-off points used to plot the graph (in fact the software uses as many points as possible given the variation in the data set). A much abridged extract of this table is presented on page 7. One can then examine this table and identify the Point A from Figure 5 that has the lowest possible number of errors. In fact the point A is found to correspond to a cut-off point of 1.2818 units, with a sensitivity of 0.88 a specificity of 0.93, and with 90.33% of subjects correctly classified. The highest LR+ value, however (40.00), was obtained for a cut-off point of 1.5092 units, at which point only 89% of subjects would have been correctly classified. This example illustrates the point that analysis using the likelihood ratio and ROC curves do not necessarily yield the same results. The reason is that the LR+ method attempts to identify that cut-off point that will result in the highest PPV. If adjusting the cut-off point (downwards in our case) results in a greater increase in the specificity than decrease in sensitivity, then the LR+ method will identify a cut-off point that is lower than the one that results in the fewest errors.

CUT-OFF POINT	SENSITIVITY	SPECIFICITY	CORRECTLY CLASSIFIED	LR+	LR-
(>= 1.216131687164307)	89.33%	88.00%	88.67%	7.4444	0.1212
(>= 1.218450903892517)	88.67%	88.00%	88.33%	7.3889	0.1288
(>= 1.238641738891602)	88.67%	88.67%	88.67%	7.8235	0.1278
(>= 1.245455384254456)	88.67%	89.33%	89.00%	8.3125	0.1269
(>= 1.261659145355225)	88.67%	90.00%	89.33%	8.8667	0.1259
(>= 1.262915730476379)	88.00%	90.00%	89.00%	8.8000	0.1333
(>= 1.265987992286682)	88.00%	90.67%	89.33%	9.4286	0.1324
(>= 1.26666533946991)	88.00%	91.33%	89.67%	10.1538	0.1314
(>= 1.278373956680298)	88.00%	92.00%	90.00%	11.0000	0.1304
(>= 1.281816840171814)	**88.00%**	**92.67%**	**90.33%**	**12.0000**	**0.1295**
(>= 1.326415300369263)	87.33%	92.67%	90.00%	11.9091	0.1367
(>= 1.330667972564697)	86.67%	92.67%	89.67%	11.8182	0.1439
(>= 1.339103817939758)	86.67%	93.33%	90.00%	13.0000	0.1429
(>= 1.341539978981018)	86.00%	93.33%	89.67%	12.9000	0.1500
(>= 1.347755193710327)	85.33%	93.33%	89.33%	12.8000	0.1571
(>= 1.352781057357788)	85.33%	94.00%	89.67%	14.2222	0.1560
(>= 1.360898613929749)	85.33%	94.67%	90.00%	16.0000	0.1549
(>= 1.393806099891663)	84.67%	94.67%	89.67%	15.8750	0.1620
(>= 1.404150485992432)	84.00%	94.67%	89.33%	15.7500	0.1690
(>= 1.41346800327301)	83.33%	94.67%	89.00%	15.6250	0.1761
(>= 1.421349167823792)	83.33%	95.33%	89.33%	17.8571	0.1748
(>= 1.427397131919861)	82.67%	95.33%	89.00%	17.7143	0.1818
(>= 1.432774662971497)	82.00%	95.33%	88.67%	17.5714	0.1888
(>= 1.441804647445679)	82.00%	96.00%	89.00%	20.5000	0.1875
(>= 1.442908116116333)	81.33%	96.00%	88.67%	20.3333	0.1944
(>= 1.444351315498352)	81.33%	96.67%	89.00%	24.4000	0.1931
(>= 1.44958758354187)	80.67%	96.67%	88.67%	24.2000	0.2000
(>= 1.475230932235718)	80.67%	97.33%	89.00%	30.2500	0.1986
(>= 1.481273531913757)	80.00%	97.33%	88.67%	30.0000	0.2055
(>= 1.509162664413452)	*80.00%*	*98.00%*	*89.00%*	*40.0000*	*0.2041*
(>= 1.516967535018921)	79.33%	98.00%	88.67%	39.6667	0.2109
(>= 1.524313926696777)	78.67%	98.00%	88.33%	39.3333	0.2177
(>= 1.538057327270508)	78.00%	98.00%	88.00%	39.0000	0.2245
(>= 1.543705344200134)	77.33%	98.00%	87.67%	38.6667	0.2313
(>= 1.548669815063477)	76.67%	98.00%	87.33%	38.3333	0.2381

This ROC curve analysis has identified the relevant point where the classification errors are at a minimum. Recall however that the classification errors are the sum of false negatives and false positives, and in this analysis these are all bundled together. In a screening programme we may be less concerned about false positive results than about false negatives. The false positives will be identified in the next round of definitive diagnostic testing, whereas the false negatives will have slipped through the net. From a screening perspective, therefore, it is still a good idea to consider all the evidence available before reaching a decision about the best policy to follow.

APPENDIX 3

Quality check-list for economic evaluation studies

CHECK LIST FOR ASSESSING ECONOMIC EVALUATION STUDIES

		Yes	No
1	Was a well-defined question posted in answerable form?		
1.1	Did study examine both cost and effects of programs?		
1.2	Did study involve a comparison of alternatives?		
1.3	Was view point and context for the analysis stated		
2	Was a comprehensive description of competing alternatives given?		
2.1	Where any important alternatives omitted?		
2.2	Was (should) a do-nothing alternative considered?		
3	Is there evidence of the program's effectiveness?		
3.1	Is it from randomised trials, meta-analysis or expert opinion?		
4	Were all relevant costs and consequences for each alternative identified?		
4.1	Was the range wide enough for the research question?		
4.2	Were capital costs and operating costs included?		
5	Were consequences measured correctly in appropriate physical units?		
6	Were costs and consequences valued credibly?		
6.1	Were sources of values clearly identified?		
6.2	Were market values employed?		
7	Were costs and consequences adjusted for differential timing?		
7.1	Was there discounting to present value?		
7.2	Is there a justification for the discount rate used?		
8	Was an incremental analysis performed of the costs/consequences of alternatives?		
9.	Was a sensitivity analysis performed?		
10.	Did the presentation and discussion of study results include all issues of concern?		
10.1	Were study results sensitive to changes to key assumptions?		
10.1	Were the results compared to previous research?		
10.2	Did the study discuss generalisability to other settings?		
10.3	Did it discuss issues of ethics, equity and feasibility?		

Reference:

Adapted by Dr Seloacoe Thooe from: Drummond, M.F., Stoddart, G.L., Torrance, G.W., 1996. *Methods for the Economic Evaluation of Health Care Programmes*. London: Oxford University Press.

APPENDIX 4

Screening results for diagnostic tests – unpublished data kindly supplied by Dr Karin Weyer from the South African Medical Research Council.

anda-TB GA test (anda Biologicals)

Test principle
Blood gelification time in the presence of aldehyde (eg. glutaraldehyde) due to increases in fibrinogen and/or polyclonal gamma-globulins. Blood from TB patients should have shorter (<10 min) gelification time than blood from non-TB patients (> 12 min)

74 cases with confirmed (sputum smear-positive) pulmonary TB, 14 HIV positive
21 cured TB cases (initially sputum smear-positive, 6-months treatment completed)
20 cases with other illnesses (eg. malignancies), TB excluded

Results (based on low numbers, interpret with caution)
Sensitivity: 97.1%
Specificity: 68.8%
Positive Predictive Value: 85.7%
Negative Predictive Value: 71.4%

Test not affected by HIV (p=0.624)
Test still positive in 76.2% of cured TB patients
High false-positive rates (55%) in patients with malignancies

2. Mycodot (DynaGen)

Test principle
Detection of anti-mycobacterial serum antibodies through lipoarabinomannan (LAM) antigen bound to plastic combs

118 culture-confirmed TB patients, 24 HIV-positive
182 healthy controls, TB excluded

Results
Sensitivity: 45.8%
Specificity: 82.9%
Positive predictive value: 63.5%
Negative predictive value: 70.2%

Sensitivity reduced to 29.5% in HIV-positive specimens

3. Latex agglutination test (Murex Diagnostics)

Test principle
Latex-coupled IgG antibodies resulting in agglutination of blood from TB patients

71 confirmed TB patients (sputum microscopy positive)
88 confirmed TB patients (sputum culture positive)
177 negative controls, TB excluded

Results

	Smear as reference	Culture as reference
Sensitivity	73.2%	56.8%
Specificity	83.9%	81.9%
Positive predictive value	61.2%	61.0%
Negative predictive value	90.1%	79.2%

4. Immuzyme-TB (Omega Diagnostics)

Test principle
Enzyme-immunoassay to detect IgG antibodies to *M. tuberculosis* in blood

A. 75 culture-confirmed TB patients
B. 75 culture-confirmed patients on at least 2 months of TB treatment
C. 75 culture confirmed patients who have completed TB treatment
D. 75 healthy negative controls, TB excluded

Results (group A only)
Sensitivity:	73.4%
Specificity:	83.2%
Positive predictive value:	68.3%
Negative predictive value:	81.2%

5. Kreatech TB kit (Kreatech Diagnostics)

Test principle
Assay based on IgA antibody titers to mycobacterial Kp-90 immuno cross-reactive antigenic compound

A. 75 culture-confirmed TB patients
B. 75 culture-confirmed patients on at least 2 months of TB treatment
C. 75 culture confirmed patients who have completed TB treatment
D. 75 healthy negative controls, TB excluded

Results (group A only)
Sensitivity:	78.1%
Specificity:	85.2%
Positive predictive value:	74.9%
Negative predictive value:	81.2%