PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Odo, Nnaemeka; Mandel, Jeffrey; Perlman, David; Alexander, Bruce; Scanlon, Paul</td>
</tr>
</tbody>
</table>

VERSION 1 - REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Philip Harber M.D., MPH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Professor of Public Health</td>
</tr>
<tr>
<td></td>
<td>Community, Environment, and Policy Division</td>
</tr>
<tr>
<td></td>
<td>Mel and Enid Zuckerman College of Public Health</td>
</tr>
<tr>
<td></td>
<td>Medical Research Building- Room 112</td>
</tr>
<tr>
<td></td>
<td>University of Arizona- MEZCOPH</td>
</tr>
<tr>
<td></td>
<td>1656 E. Mabel St. Tucson, AZ 85719 USA</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>28-Jan-2013</td>
</tr>
</tbody>
</table>

THE STUDY

The writing style needs improvement. With rewriting, this could become an important article.

GENERAL COMMENTS

GENERAL

The paper includes useful information about several factors which influence estimates of respiratory disease prevalence in worker populations—the optimal combination of multiple tests to define the presence of “disease” as well as the influence of inclusion/exclusion criteria for lung function test results. While these topics have been recognized, the authors’ approach adds new data and novel methods.

The paper is, however, limited by the writing style. Several suggestions might improve the readability and ability to interpret its scientific quality:

1) The paper should use clearer terminology for lung volumes. In particular, it refers to alveolar volume without distinguishing clearly whether it is based upon a single breath as part of the diffusing capacity (a.k.a. transfer factor) study or based upon a helium dilution technique with six minutes equilibration time.

2) The paper discusses “Total Lung Capacity” without always distinguishing whether referring to measurement as part of the six minute equilibration helium dilution technique or by body plethysmography. (In the latter case, they may wish to use the term thoracic gas volume.

3) The authors may wish to refer to a widely accepted guideline for both terminology and definition of methods for lung volume

4) The paper would benefit from use of appropriate subscripts for the physiologic terms. These should follow the guidelines of the journal to which submitted or, in the absence of such clear guidelines, the approaches used by the ATS/ERJ journals.

5) The distinction between the “class of physiologic abnormality” and a “disease” should be made more clear. The presence of a particular clinical physiologic findings (e.g., reduced FVC) does not define a “disease”. This is particularly important when the implicit relevance is to diagnosis of a pneumoconiosis (a disease) in the taconite miners.

6) The authors should confirm that they are using the term “percentile” referring to a population-based distribution rather than “percent of predicted” in many of their analyses.

SPECIFIC COMMENTS

Page 5, line 11: To which alveolar volume do the authors refer—single breath or helium dilution?

Page 5, line 16: the term “restrictive lung disease (RLD)” should be defined and differentiated from “restrictive physiologic abnormality”. They make this point more clearly later in the paragraph when they refer to “pattern of respiratory impairment”, but subsequently confound physiologic pattern and disease diagnosis in the rest of the paper.

Page 6 line 19: They refer to the ATS criteria. The ERS now has joined criteria with ATS.

Page 9, line 28: The paper mentions airflow obstruction in as a possible cause for differences between plethysmography and helium dilution measurements, but does not mention the role of non-communicating air spaces (e.g., cysts). The discussion does not mention the term “thoracic gas volume” when referring to plethysmography.

Page 9, line 49: The meaning of the sentence concerning “true values of spirometry estimates” is unclear; how do they define “true value”?

Page 10 (Table 2): The table does not include comparison of results across the groups (e.g., an analysis of variance about whether BMI actually differed). They use the term “covariate” in the table, but it actually has no statistical analysis treating these as covariates.

Tables (in general): The reader would be aided by using short descriptive titles rather than referring to “group 1”, etc. They do have short descriptors in the text itself.

Page 11, line 19: They state that test combinations would be “more accurate”. The definition of “accuracy” should be provided.
Page 11 (and elsewhere): The term “lung obstruction” should be defined somewhere. Do they really mean “airflow obstruction”? 

Page 11, line 30: They refer to the “percentile value for each participant”. Please confirm whether this is truly a population based percentile or is actually “percentage of predicted”.

Page 13, Table 3: The table has an inordinate number of footnotes. Perhaps the classification scheme could be defined elsewhere rather than repeat it here. In addition, the Clopper-Pearson method for determining confidence intervals is not described in the Methods section.

Page 14, line 4: The authors discuss BMI and related multivariable regressions. However, they do not provide any data. This is an important topic, and tables summarizing their analyses would be helpful.

Page 14 line 9: The authors should confirm whether they refer to “percentiles” or actually mean “percentage of predicted”.

Page 14, line 28: The first paragraph of the Discussion seems to address two distinct topics. If the last two sentences actually refers to the low participation rate, it should be in a separate paragraph.

Page 15, line 21: Discussion of the diffusing capacity (transfer factor) is useful. However, it seems to imply that deficits of gas transfer are more important or relevant than other causes of reduced diffusing capacity.

Page 16, line 20: They seem to imply that lung disease in Taconite miners is largely restricted to problems with gas exchange per se. The discussion might be strengthened if they considered the non-pneumoconiotic effect (i.e., airflow obstruction) that is increasingly recognized as a consequence of heavy dust exposure. In addition, the increasingly important role of longitudinal versus single point cross-sectional analysis in detecting lung disease in miners (“accelerated decliners”) might be mentioned to provide perspective to the reader.

In summary, the paper has useful data and suggests two important points-exclusion criteria affect the estimated prevalence, and test combinations may be preferable to single tests. Significant rewriting of the text could make this considerably clearer.

References:
The inclusion of outdated references (eg., 3,4, both of which have been supplanted by newer versions” is of no value). Conversely, the “classic” paper about exclusion biases was overlooked (Eisen EA, Dockery DW, Speizer FE, Fay ME, Ferris BG Jr. The association between health status and the performance of excessively variable spirometry tests in a population-based study in six U.S. cities. Am Rev Respir Dis. 1987

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>David Rees</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Head Occupational medicine division National Institute for Occupational Health, Professor Occupational health, University of the Witwatersrand</td>
</tr>
</tbody>
</table>
THE STUDY

The topic is important because workplace surveillance programmes and studies may use spirometry to identify subjects with restrictive lung function patterns. The limitations of using spirometry for this purpose may not be fully appreciated. The focus of the article was to investigate the range of estimates of disease (as identified by abnormal lung function tests) depending on how well subjects performed the lung function tests (using standard criteria of performance). The authors showed that estimates of disease prevalence were related to test performance.

A methodological problem is the use of alveolar volume (VA) instead of total lung capacity. An objective of the study was “to compare population estimates for the prevalence of lung restriction based on spirometry, alveolar volume (VA) and diffusing capacity (DLCO) in miners.” Total lung capacity is the standard for identifying restrictive lung disease and what we really need to know is how these three measurements compare to the standard. So a sound study design would be to compare the three measures of lung function (spirometry, VA and DLco) to the standard (TLC), even if the standard is not available in workplaces. Instead they are compared among each other, which has some value (as VA closely matches TLC in normal subjects) but is less than optimal since each of these methods has its own problems in reliably identifying restriction. This aspect should be considered in more detail in Limitations with more information on the literature evaluating VA as a measure of TLC.

Abstract

The first objective is “To assess the impact of acceptability criteria ...” But the criteria included repeatability as well. In much of the text “acceptability criteria” is used when I think it is both acceptability and repeatability.

Introduction

I think a more explicit statement of the objectives is required at the end of the Introduction.

Methods

It is not clear who did the measurements of DLCO and VA. Were these tests done by the technicians who did the spirometry? Table 1. I think it would be better to use “test” rather than “read”. The footnote is not clear to me. “In this particular grouping ..” does not specify the grouping and a QC grade is introduced without explanation.

Is it appropriate to adjust for age and gender (sex) in the multivariate analyses to determine the association between BMI and FVC? FVC was based on normal reference equations which adjust for age and gender already.

RESULTS & CONCLUSIONS

Results

Footnotes to Table 2. Only acceptability criteria are specified. Should it not be acceptable and repeatable?

I think it would be useful to include the results of the models showing decreasing FVC percentile with increased BMI.

Discussion

It is not surprising that test performance criteria affected the assessment of disease. Poor test performance creates apparent disease hence the wide use of performance criteria.

It seems to me that the proportion of subjects who did not meet all ATS criteria is unusually large (565/1084 = 524%). Is this larger than usual when groups of workers and former workers are tested outside
formal lung function labs? This high proportion suggests that there were problems in doing the tests and would affect the generalisability of the findings.

The “VA restriction” subjects could well have included some subjects with obstruction. A FEV1/FVC ≥ LLN occurs in obstruction if the FVC is underestimated and many subjects failed end of test criteria which may have reduced measured FVC.

The ATS/ERS article referenced to justify considering a test usable even if EOT criteria are not met actually states “However, early termination, by itself, is not a reason to eliminate all the results from such a manoeuvre from further consideration. Information on the FEV1 may be useful (depending on the length of exhalation) and could be reported from these early termination manoeuvres.” I think this should be clearer in the article.

In conclusion, I think the major findings are not novel: poor test performance will influence the categorisation of abnormality and BMI is known to affect spirometric parameters.

GENERAL COMMENTS

I feel that the article is more appropriate for an occupational health journal or one with a focus on pulmonology.

VERSION 1 – AUTHOR RESPONSE

Philip Harber, M.D., MPH

General Comments:

1) The paper should use clearer terminology for lung volumes. In particular, it refers to alveolar volume without distinguishing clearly whether it is based upon a single breath as part of the diffusing capacity (a.k.a. transfer factor) study or based upon a helium dilution technique with six minutes equilibration time.

Response: We recognize the need for clearer terminology for lung volumes in this paper. Three tests for lung volumes were included, body plethysmography (gold standard for diagnosis), spirometry and diffusing capacity (measuring alveolar volume). Body plethysmography was not conducted in this study. However, the other tests were used as an approximation for total lung capacity (TLC) measured using by body plethysmography. Terminologies from the ATS/ERS guidelines were applied throughout this article to help maintain clarity. Subscripts have now been used at all appropriate places, including subscripts for spirometry, diffusing capacity and lung volumes. These are FEV1 for forced expiratory volume in first second of expiration and TLCpleth for total lung capacity measured by body plethysmography. This distinction of method of alveolar volume measurement was stated in the methods section of the article (Page 7, paragraph 3). “…an Ultima PF system (Medical Graphics Corporation, St. Paul, MN) was used to conduct DLCO measurements. The latter uses single-breath helium dilution for the measurement of alveolar volume (VA).”

2) The paper discusses “Total Lung Capacity” without always distinguishing whether referring to measurement as part of the six minute equilibration helium dilution technique or by body plethysmography. (In the latter case, they may wish to use the term thoracic gas volume)

Response: We acknowledge the reviewer’s comment on the use of the term “Total Lung Capacity.” We have gone through the paper and clarified TLC throughout the paper. Given the suggestion by the ATS/ERS as to the non-specificity of the term, we decided not to use “thoracic gas volume.”

3) The authors may wish to refer to a widely accepted guideline for both terminology and definition of methods for lung volume determinations: The American Thoracic Society and the European
Respiratory Society published a Guideline, “Standardization of the Measurement of Lung Volumes” in The European Respiratory Journal, 2005:26:511. While their meanings can be discerned with some effort, use of the standardized approaches would make the paper easier to read and evaluate.


Terminologies from the ATS/ERS guidelines were used throughout this article. Subscripts have been used at all appropriate places, including subscripts for spirometry, diffusing capacity and lung volumes. We used standard approaches, as defined in our methods. This was a reflection of the ATS/ERS documents. We added a section title that reflects these approaches and another that addresses other definitions used.

4) The paper would benefit from use of appropriate subscripts for the physiologic terms. These should follow the guidelines of the journal to which submitted or, in the absence of such clear guidelines, the approaches used by the ATS/ERS journals.

Response: Terminologies from the ATS/ERS guidelines were used throughout this article. Subscripts have been used at all appropriate places, including subscripts for spirometry, diffusing capacity and lung volumes.

Examples of these were FVC for forced vital capacity, FEV1 for forced expiratory volume in first second of expiration, TLC for total lung capacity in general (TLCpleth for total lung capacity measured by body plethysmography), DLCO for diffusing capacity and VA for alveolar volume.

5) The distinction between the “class of physiologic abnormality” and a “disease” should be made clearer. The presence of a particular clinical physiologic finding (e.g., reduced FVC) does not define a “disease”. This is particularly important when the implicit relevance is to diagnosis of pneumoconiosis (a disease) in the taconite miners.

Response: We agree with this comment and added the term “Restrictive Ventilatory Defect (RVD)” in the writing, especially when referring to abnormality identified using spirometry (See introduction, 1st paragraph). We describe the use of RLD and do explain the difference between RVD and RLD in the introduction. “RLD refers to impaired lung expansion, ventilation/oxygenation and is diagnosed using body plethysmography. It is a subset of what is actually measured, which is Restrictive Ventilatory Defect (RVD). In this paper, we measure RVD as an estimate for RLD.”

6) The authors should confirm that they are using the term “percentile” referring to a population-based distribution rather than “percent of predicted” in many of their analyses.

Response: This clarification about percentile assessment was included in the Methods section (Methods section, last paragraph). “This percentile is the FVC of each participant, standardized to the NHANES III population-based distribution of normal lung function. This is different from percent predicted which is a ratio of the FVC to the median predicted value for each participant.” We decided to use the term, “restrictive ventilatory defect” as a more representative term than implying the diagnosis of restrictive lung disease.

Specific comments:
Page 5, line 11: To which alveolar volume do the authors refer-single breath or helium dilution?

Response: This distinction was stated in the methods section of the article (Page 7, Paragraph 3). “…an Ultima PF system (Medical Graphics Corporation, St. Paul, MN) was used to
conduct DLCO measurements. The latter uses single-breath helium dilution for the measurement of alveolar volume (VA).

Page 5, line 16: the term “restrictive lung disease (RLD)” should be defined and differentiated from “restrictive physiologic abnormality”. They make this point more clearly later in the paragraph when they refer to “pattern of respiratory impairment”, but subsequently confound physiologic pattern and disease diagnosis in the rest of the paper.

Response: We included the term “Restrictive Ventilatory Defect (RVD)” as well as differentiating RLD from RVD throughout the document (See introduction, 1st paragraph). “RLD refers to impaired lung expansion, ventilation/oxygenation and is diagnosed using body plethysmography.”

Page 6 line 19: They refer to the ATS criteria. The ERS now has joined criteria with ATS.

Response: This was corrected throughout the document where applicable starting in the introduction section (3rd paragraph). “….. using the American Thoracic Society and European Respiratory Society (ATS/ERS) recommendations for spirometry.”

Page 9, line 28: The paper mentions airflow obstruction in as a possible cause for differences between plethysmography and helium dilution measurements, but does not mention the role of non-communicating air spaces (e.g., cysts). The discussion does not mention the term “thoracic gas volume” when referring to plethysmography.

Response: This clarification was included in the methods section (Paragraph 7 in the methods section). ”These obstructive scenarios include cysts, non-communicating bullae/air-spaces and pneumothorax and are not incorporated in the single-breath helium estimate of lung volume (VA).” Given the suggestion by the ATS/ERS as to the non-specificity of the term, we decided not to use “thoracic gas volume.” We use TLCpleth to refer to TLC measured using body plethysmography throughout the paper.

Page 9, line 49: The meaning of the sentence concerning “true values of spirometry estimates” is unclear; how do they define “true value”?

Response: This was referring to FVC values that would have been attained if these tests had reached plateau (EOT criteria). We clarify the meaning “true value” in the methods section, paragraph 11. “The assigned FVC values in this group (no plateau end-point reached) were likely close to true FVC values that would have been attained if plateau had been reached.”

Page 10 (Table 2): The table does not include comparison of results across the groups (e.g., an analysis of variance about whether BMI actually differed). They use the term “covariate” in the table, but it actually has no statistical analysis treating these as covariates.

Response: This comparison was done between the only two uniquely distinct groups, “excluded tests” vs. “acceptable tests” (groups 2 and 3). The findings were included in the 2nd paragraph of the results section of the article. “Comparison of groups 2 and 3 (“Exclusions” vs. “Met criteria”) showed a significantly higher mean testing age in group 2, a higher proportion of males, a higher mean FEV1 and a significantly higher proportion of ever smokers (current and former). All p-value estimates were <0.0001 and average BMI was not significantly different between the two groups.” The p-values were derived from chi-square and t-statistics. The label “variables” in table 2 was changed to “parameters”. And “covariates” in the footnotes of
table 2 was also changed to “parameters”.

Tables (in general): The reader would be aided by using short descriptive titles rather than referring to “group 1”, etc. They do have short descriptors in the text itself.

Response: Short descriptors were included for the groups in tables 2 & 3: “Total group” for group 1, “Exclusions” for group 2, “Met criteria” for group 3, and “Usable” for group 4.

Page 11, line 19: They state that test combinations would be “more accurate”. The definition of “accuracy” should be provided.

Response: More accurate identification was replaced with ‘increased likelihood’ for clarity. This was included in the methods section (Paragraph 11) for clarification. “Prevalence estimates of RVD determined by combining tests were presented to show increased likelihood of restriction determination (Table 3). The combination of these three tests also represented the lower bound for RVD prevalence estimates.”

Page 11 (and elsewhere): The term “lung obstruction” should be defined somewhere. Do they really mean “airflow obstruction”?

Response: This was clarified in the paper. In the methods section (Paragraph 7) we state, “A subject was identified as having ‘airflow/lung/spirometric obstruction’ if the FEV1/FVC ratio was < LLN.”

Page 11, line 30: They refer to the “percentile value for each participant”. Please confirm whether this is truly a population based percentile or is actually “percentage of predicted”.

Response: This clarification about percentile assessment was included in the article (Methods section, last paragraph). “This percentile is the FVC of each participant, standardized to the NHANES III population-based distribution of normal lung function. That is, the percentiles of the measured FVC in the normal population distribution of lung function. This is different from percent predicted which is a ratio of the FVC to the median predicted value for each participant (race, age, height and gender adjusted).”

Page 13, Table 3: The table has an inordinate number of footnotes. Perhaps the classification scheme could be defined elsewhere rather than repeat it here. In addition, the Clopper-Pearson method for determining confidence intervals is not described in the Methods section.

Response: We reduced the footnotes for tables 2 & as suggested. We also referred to the classification scheme for groups and lung function outcomes as seen in the methods section (Paragraphs 7 and 9 for lung function outcomes & paragraph 10 for groups). The clopper-pearson method was described in the methods section (Paragraph 12). “Crude prevalence estimates of obstruction and restriction by spirometry were determined. Exact (Clopper-Pearson) 95% confidence intervals were derived for all estimates of lung function impairment. This method has the advantage of calculating conservative confidence interval estimates when assessing binomial proportions.”

Page 14, line 4: The authors discuss BMI and related multivariable regressions. However, they do not provide any data. This is an important topic, and tables summarizing their analyses would be helpful.

Response: Results of this analysis were included in the document in form of a table, table 4 (just before the discussion section).
Page 14 line 9: The authors should confirm whether they refer to “percentiles” or actually mean “percentage of predicted”.

Response: We acknowledge this comment. We do refer to percentiles and do not mean percentage of predicted. This clarification about percentile assessment was included in the Methods section (Methods section, last paragraph). “This percentile is the FVC of each participant, standardized to the NHANES III population-based distribution of normal lung function. This is different from percent predicted which is a ratio of the FVC to the median predicted value for each participant (race, age, height and gender adjusted).”

Page 14, line 28: The first paragraph of the Discussion seems to address two distinct topics. If the last two sentences actually refer to the low participation rate, it should be in a separate paragraph.

Response: We separated these two points in this first paragraph in the discussion as suggested. See paragraphs 1 & 2 in the discussion section.

Page 15, line 21: Discussion of the diffusing capacity (transfer factor) is useful. However, it seems to imply that deficits of gas transfer are more important or relevant than other causes of reduced diffusing capacity.

Response: See related comments immediately below.

Page 16, line 20: They seem to imply that lung disease in Taconite miners is largely restricted to problems with gas exchange per se. The discussion might be strengthened if they considered the non-pneumoconiotic effect (i.e., airflow obstruction) that is increasingly recognized as a consequence of heavy dust exposure.

Response: We included comments on gas exchange factors in the discussion (Paragraph 3). “Our assessment showed a high prevalence of obstructive patterns in those not meeting acceptability criteria, compared to those fully meeting criteria (27.6% vs. 5%). Identifying lung obstruction in this mining population is relevant because of the recognized role of heavy dust exposure in causing lung obstruction (non-pneumoconiotic effect).”

In addition, the increasingly important role of longitudinal versus single point cross-sectional analysis in detecting lung disease in miners (“accelerated decliners”) might be mentioned to provide perspective to the reader.

Response: We included comments on the need for longitudinal monitoring in the discussion (last paragraph). “The common use of spirometry underscores its importance in longitudinal monitoring of workers in at-risk populations. This form of respiratory health surveillance is typically used because of its ease of administration, lower cost, and portability. Longitudinal monitoring has an advantage over single cross-sectional testing because of its ability to detect increasing abnormalities (trends) that gradually develop over time. In longitudinal epidemiologic investigations, each study participant has multiple data points, which provides better estimates of lung function.”

References:
The inclusion of outdated references (e.g., 3, 4, both of which have been supplanted by newer versions” is of no value). Conversely, the “classic” paper about exclusion biases was overlooked (Eisen EA, Dockery DW, Speizer FE, Fay ME, Ferris BG Jr. The association between health status and the performance of excessively variable spirometry tests in a population-based study in six U.S.
Dr. David Rees

The topic is important because workplace surveillance programmes and studies may use spirometry to identify subjects with restrictive lung function patterns. The limitations of using spirometry for this purpose may not be fully appreciated. The focus of the article was to investigate the range of estimates of disease (as identified by abnormal lung function tests) depending on how well subjects performed the lung function tests (using standard criteria of performance). The authors showed that estimates of disease prevalence were related to test performance.

General Comments:
A methodological problem is the use of alveolar volume (VA) instead of total lung capacity. An objective of the study was “to compare population estimates for the prevalence of lung restriction based on spirometry, alveolar volume (VA) and diffusing capacity (DLCO) in miners.” Total lung capacity is the standard for identifying restrictive lung disease and what we really need to know is how these three measurements compare to the standard. So a sound study design would be to compare the three measures of lung function (spirometry, VA and DLco) to the standard (TLC), even if the standard is not available in workplaces. Instead they are compared among each other, which has some value (as VA closely matches TLC in normal subjects) but is less than optimal since each of these methods has its own problems in reliably identifying restriction.

This aspect should be considered in more detail in Limitations with more information on the literature evaluating VA as a measure of TLC.

Response: In the discussion, we made clarifications about the limitations of VA as a measure of TLC. Example in paragraph 5, “Though some studies have shown TLCpleth to be comparable to lung volume measured by single-breath helium dilution (VA), current guidelines point out that VA underestimates TLC in the setting of moderate to severe obstructive disease.6, 9, 12-14 This is a key limitation of using VA as an approximate for TLC. For these reasons, use of VA (and DLCO) in this study was limited to participants without evidence of obstruction.” Clarification in paragraph 6, “While DLCO may be a sensitive indicator of early interstitial lung disease, it is not a measure of lung volume.”

Abstract
The first objective is “To assess the impact of acceptability criteria ...” But the criteria included repeatability as well. In much of the text “acceptability criteria” is used when I think it is both acceptability and repeatability.

Response: Acceptability criteria refer to the criteria for acceptable maneuver blows and for repeatable maneuver blows. This detail was added in the methods section, 3rd paragraph for clarity. “These criteria for spirometry testing included both criteria for acceptable blows, and criteria for maneuver repeatability.”

Introduction
I think a more explicit statement of the objectives is required at the end of the Introduction.

Response: A statement of the objectives was included at the end of the introduction.
“There were three primary objectives in this study. The first was to assess the impact of ATS/ERS “acceptability” and “usability” criteria on estimates of restrictive ventilatory defect in a population of taconite miners. The second was to compare estimates of restrictive ventilatory defect with three different pulmonary function tests (spirometry, alveolar volume and diffusing capacity). The third was to assess the role of population characteristics on these estimates.”

Methods

It is not clear who did the measurements of DLCO and VA. Were these tests done by the technicians who did the spirometry?

Response: This clarification was included in the methods, 4th paragraph. “Screening spirometry was performed by technicians trained in a 2-day NIOSH-certified spirometry course. These technicians were also trained to and performed diffusing capacity testing.”

Table 1. I think it would be better to use “test” rather than “read”.

Response: For table 1, “read” was changed to “maneuver” to more specifically use the terms from the ATS/ERS document.

The footnote is not clear to me. “In this particular grouping ...” does not specify the grouping and a QC grade is introduced without explanation.

Response: We acknowledge your comments about this statement on “grouping.” This does not describe any terms in the table. For this reason it was removed from the footnotes. “QC grade” was described in the methods, paragraph 10. “These “usable” tests were 1) tests with quality control (QC) grades of ≥ ’B’ (at least two acceptable maneuvers with FEV1 values repeatable within 101 - 150 ml).” Since it does not directly describe table 1, it was removed from the footnotes.

Is it appropriate to adjust for age and gender (sex) in the multivariate analyses to determine the association between BMI and FVC? FVC was based on normal reference equations which adjust for age and gender already.

Response: Individual test performance is adjusted to a standard population-based on the individual’s age, gender, height and race in comparison to that standard population. This maneuver assumes that these characteristics are taken into account for all individuals tested. Given the likelihood of the test population being skewed by disproportionate categories of age and gender, it is important to control for these variables in the multivariate modeling. By doing this, the association between BMI and FVC percentile can be better understood within the test population. In effect, individual test performance adjustment is important in determining an individual’s FVC status, relative to the standard population. Adjusting for age and gender in the test groups is important in understanding the association between BMI and FVC unique to the population studied.

Results

Footnotes to Table 2. Only acceptability criteria are specified. Should it not be acceptable and repeatable?

Response: Acceptability criteria refer to the criteria for acceptable maneuver blows and for repeatable maneuver blows.
This detail was added in the methods section, 3rd paragraph for clarity. “These criteria for spirometry testing included both criteria for acceptable blows, and criteria for maneuver repeatability.”

I think it would be useful to include the results of the models showing decreasing FVC percentile with increased BMI.

Response: Results of this analysis was included in the document in form of a table, table 4.

Discussion
It is not surprising that test performance criteria affected the assessment of disease. Poor test performance creates apparent disease hence the wide use of performance criteria.

It seems to me that the proportion of subjects who did not meet all ATS criteria is unusually large (565/1084 = 52.4%). Is this larger than usual when groups of workers and former workers are tested outside formal lung function labs? This high proportion suggests that there were problems in doing the tests and would affect the generalizability of the findings.

Response: We acknowledge your comments about the large number of subjects who did not meet all ATS criteria. There is insufficient current literature on meeting spirometric criteria in mining studies to properly compare this population with others tested outside formal lung function labs. Studies in clinic populations suggest that this proportion (52.4%) is increased.

Paragraph 2 of the discussion section refers to population factors (like age, BMI, smoking) thought to affect test performance. There could be other effects not measured (uncontrolled confounding), but the ones used are the main ones that are described. They are also thought to affect accurate identification of restrictive ventilatory defect (RVD).

The “VA restriction” subjects could well have included some subjects with obstruction. FEV1/FVC ≥ LLN occurs in obstruction if the FVC is underestimated and many subjects failed end of test criteria which may have reduced measured FVC.

Response: We included comments to this effect in the discussion, paragraph 5. “VA underestimates TLC in the setting of moderate to severe obstructive disease. This is a key limitation of using VA as an approximate for TLC. For these reasons, use of VA (and DLCO) in this study was limited to participants without obstruction (FEV1/FVC ≥ LLN). With our inclusion of tests not meeting EOT criteria, VA restriction could have included some subjects with obstruction.

The ATS/ERS article referenced to justify considering a test useable even if EOT criteria are not met actually states “However, early termination, by itself, is not a reason to eliminate all the results from such a manoeuvre from further consideration.”

Response: We acknowledge the reviewer comment about early termination. For that reason, we included this group in tables 2 and 3 (Group 4).

Information on the FEV1 may be useful (depending on the length of exhalation) and could be reported from these early termination manoeuvres.” I think this should be clearer in the article.

Response: Average FEV1 values were higher in the acceptable test group. There were also more participants in the early termination maneuvers with mildly and moderately reduced FEV1 values. These results are included in the results section, in paragraph 3. “Comparison of groups 2 and 3 (“Exclusions” vs. “Met criteria”) showed a significantly higher mean
testing age in group 2, a higher proportion of males, a higher mean FEV1 and a significantly higher proportion of ever smokers (current and former).”

**VERSION 2 – REVIEW**

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Philip Harber</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Professor of Public Health</td>
</tr>
<tr>
<td></td>
<td>University of Arizona</td>
</tr>
<tr>
<td></td>
<td>USA</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>25-May-2013</td>
</tr>
</tbody>
</table>

**GENERAL COMMENTS**

**COMMENT TO AUTHORS:**

The paper is significantly improved both in terms of clarity and technical accuracy.

The authors may wish to consider whether the paper should be limited to spirometry alone rather than including the results of the single breath diffusing capacity and single breath helium dilution lung volume test. Several considerations include the following:

- Spirometry is widely used in worksites, whereas diffusing capacity and alveolar volume are not.
- When spirometry is complemented by additional tests, the six minute helium dilution lung volumes are generally measured along with the diffusing capacity. (This is commonly done since the equipment is generally constructed to perform this).
- As the authors point out in their discussion, estimating lung volumes (e.g., TLC) based upon the single breath used in the diffusing capacity test is frequently very inaccurate. It is particularly inaccurate in the presence of airflow obstruction, which is one of their major classifications. (The actual dilution TLC is occasionally inaccurate when compared with plethysmography as they point out, but this is much less common than estimating TLC from the single breath Va).
- When classifying physiologic patterns by spirometry, the authors follow well accepted definitions. Conversely, when they add the diffusing capacity test result, their classification differs from the more usual approach which considers the six minute dilution or plethysmography volume.
- Including these data makes the results and the discussion much more difficult for the reader to follow.

Abstract: They use the term “lung restriction”, but they probably mean “restrictive physiologic patterns”.

Page 29 line 19: Do they mean “restrictive lung disease” or disease with restrictive physiologic pattern?

Page 29 line 22: (important): Some might not agree with their concepts in this paragraph. For example, body plethysmography is not often used because of its limited availability, but rather six minute equilibration helium dilution is employed. In addition, they define “RLD” by impairment of “ventilation/oxygenation”, whereas ventilation perfusion mismatch often occurs from either obstructive airway disease or from pulmonary vascular disease rather than “restrictive” abnormalities per se.

Page 32 line 14: (minor): The sentence is poorly worded, suggesting that “quality checks” is a “common error”. They should rearrange the
Page 9. line 30: This paragraph is ambiguous when it discusses whether “lung volume estimation by helium dilution” refers to the usual six minute equilibration technique or the single breath method that is incidentally incorporated in diffusing capacity testing. They should clarify which term they mean every time they reference this.

Page 11 line 29 (and elsewhere): The author should more carefully follow standard conventions when they refer to “VA” and “DLCO”. For example, these require subscripting. (If the journal to which submitted prefers full capitalization approach, this should of course be used)

Table 3: The formatting is markedly improved from the prior version.

Table 4: (Important): When they refer to “percentiles”, are they truly referring to percentiles within a distribution or do they really mean “percent of predicted” (which is the more commonly employed technique in PFTs)? (e.g., an individual whose result is 80% of predicted value is not at the 80th percentile within a population). If they truly mean percentiles, this should be described carefully under methods.

Page 40 line 17 (minor): they mean “ventilator”, not “ventilator”.

Page 41, line 18: The authors provide a nice discussion of the inaccuracy of the Va method in comparison to actual TLC measurement. This reviewer believes this supports focusing on the spirometry per se rather than confusing matters by classifying with the Va while knowing that the error is not random, but rather is confounded by the presence of disease.

Page 42 line 47 and elsewhere: The use of the term “lung restriction” should be carefully employed. For example, a low FVC will occur in the presence of diseases affecting muscles even though the lungs per se may be normal. They should consider using “restrictive physiologic pattern” or even use their term “RVD” rather than the less technically correct term of “lung restriction”.

In summary, the paper is significantly improved and has useful information to convey. Attention to some of these technical details and consideration of the relative utility/obfuscation of including diffusing capacity and single breath lung volume data should be considered. (Certainly, if the authors strongly believe that these additional data are an inherent part of their paper, the editors should consider respecting the authors’ opinion)

REVIEWER
David Rees
Head: Occupational medicine National Institute for Occupational Health; Professor Occupational Health, University of the Witwatersrand. South Africa

I have no competing interests.

REVIEW RETURNED 22-May-2013

THE STUDY
The objectives in the abstract and text (page 6) differ. I think it is better to have a single uniform set of objectives.
RESULTS & CONCLUSIONS

I am concerned about the statement "Overall, the prevalence estimates determined using Group 4 reflect a reasonable compromise in the application of ATS/ERS criteria." Page 17 line 37. The logic is fair enough (older and sicker participants may be excluded if ATS/ERS criteria were strictly applied). But spirometric obstruction prevalences were very different for Groups 4 and 3: 17.6% and 5.0%. The assumption is I suppose that the differences are explained by sicker and older people, who "failed" the criteria, were included in Group 4. But we do not know this is true. Without knowing this one could also conclude that Group 4 has been shown by the data to be inappropriate as it overestimates obstruction. The statement should be moderated or supported by "sickness" data.

VA restriction prevalences are greater than spirometry restriction prevalences for all 4 groups. It is unclear to me why this is the case, as VA restriction excluded participants with obstruction and hence those with a mixed pattern, and VA restriction would exclude participants with a spuriously low FVC. So I would have expected VA restriction prevalences to be lower. This finding should be discussed.

GENERAL COMMENTS

I have reviewed this article before and many questions have been answered so I have minor points only.

The article is well written and interesting.

Table 1. The difference between categories "Two highest acceptable not repeatable" and "Not repeatable (3 acceptable maneuvers)" is not clear to me. In the former were there only 2 acceptable maneuvers?

Table 3. Footnotes and superscripts need revision. (E.g. in the Table there are 10 numerical superscripts but the footnotes go to 8.)

The first sentence of Discussion: restrictive ventilator should ventilatory.

Page 16 line 29 of Discussion. I am unsure what is meant by "This approach also improves generalizability of our results." It would reduce generalizability to other settings if ATS/ERS criteria were strictly applied in the other settings.

Page 17 line 50. "The prevalence of RVD determined with either spirometry or VA in combination (&) with isolated reduction in DLCO represented RVD likely caused by interstitial lung disease." took me a while to understand. I think it would help the reader if you referred to the footnotes in Table 3 to identify the test combinations.

The last paragraph of the Discussion is confusing to me. I am unsure of the basis of the statement that spirometry is a reasonable approach for the identification of lung restriction in this setting. It does not follow from the preceding sentences and is somewhat in conflict with the Conclusions.

I think that the last 3 sentences of the Discussion should be omitted. They are not pertinent to the objectives of the article and are not logical in my view. For example, because spirometry is common does not underscore its importance in longitudinal monitoring.

VERSION 2 – AUTHOR RESPONSE
Dr. David Rees

A. The objectives in the abstract and text (page 6) differ. I think it is better to have a single uniform set of objectives.

Response: The objectives in the abstract were corrected and now read:
1) To assess the impact of ATS/ERS “acceptability” and “usability” criteria for spirometry on the estimates of restrictive ventilatory defect in a population of taconite miners
2) To compare estimates of restrictive ventilatory defect with three different pulmonary function tests (spirometry, alveolar volume (VA) and diffusing capacity (DLCO))
3) To assess the role of population characteristics on these estimates
These are the same as expressed in the introduction.

B. I am concerned about the statement "Overall, the prevalence estimates determined using Group 4 reflect a reasonable compromise in the application of ATS/ERS criteria.” Page 17 line 37. The logic is fair enough (older and sicker participants may be excluded if ATS/ERS criteria were strictly applied). But spirometric obstruction prevalences were very different for Groups 4 and 3: 17.6% and 5.0%. The assumption is that the differences are explained by sicker and older people, who “failed” the criteria, were included in Group 4. But we do not know this is true. Without knowing this one could also conclude that Group 4 has been shown by the data to be inappropriate as it overestimates obstruction. The statement should be moderated or supported by “sickness” data.

Response: We agree with the reviewer’s comment. In the text, we added a statement expressing this assumption, “Overall, the prevalence estimates determined using Group 4 reflected a reasonable compromise in the application of ATS/ERS acceptability criteria. This assumes that the differences in obstruction prevalences observed between Groups 3 and 4 represent the sicker, older miners, as suggested by the differences in mean age, BMI and current smoking proportions.”

C. VA restriction prevalences are greater than spirometry restriction prevalences for all 4 groups. It is unclear to me why this is the case, as VA restriction excluded participants with obstruction and hence those with a mixed pattern, and VA restriction would exclude participants with a spuriously low FVC. So I would have expected VA restriction prevalences to be lower. This finding should be discussed.

Response: We agree with the reviewer that the prevalence of VA restriction after excluding obstruction should be less than overall spirometric restriction. In table 3, when looking at “All spirometric restriction,” this is the case, as expected. However, in the same population at risk, i.e., after excluding obstruction from the denominator, VA restriction proportion may exceed spirometric restriction proportion. This is possible because VA has enhanced sensitivity compared to spirometry when true restriction is present. A comment has been added in page 17, in the first paragraph, to this effect.
We refer to the following article:
I have reviewed this article before and many questions have been answered so I have minor points only.

The article is well written and interesting.

D. Table 1. The difference between categories "Two highest acceptable not repeatable" and "Not repeatable (3 acceptable maneuvers)" is not clear to me. In the former were there only 2 acceptable maneuvers?
Response: We acknowledge the comment about the need for more clarity on these two entries in this table. We added the following descriptions in the footnotes, **"Attained 3 acceptable maneuvers but, as per criteria, the two highest values were not repeatable. The lower two were repeatable and their values were used. **"Attained 3 acceptable maneuvers but, as per criteria, none were repeatable.” These were for *Two highest acceptable not repeatable and **Not repeatable (3 acceptable maneuvers) entries.

---

E. Table 3. Footnotes and superscripts need revision. (E.g. in the Table there are 10 numerical superscripts but the footnotes go to 8.)

Response: Thank you for pointing out this error. It has been corrected in the edits.

F. The first sentence of Discussion: restrictive ventilator should ventilatory.

Response: It has been corrected in the edits.

---

G. Page 16 line 29 of Discussion. I am unsure what is meant by “This approach also improves generalizability of our results.” It would reduce generalizability to other settings if ATS/ERS criteria were strictly applied in the other settings.

Response: We acknowledge and agree with the reviewer’s comment regarding the generalizability of findings. We removed the sentence from the paragraph as it appeared to confuse the preceding point about the need for including tests not meeting EOT criteria.

---

H. Page 17 line 50. “The prevalence of RVD determined with either spirometry or VA in combination (&) with isolated reduction in DLCO represented RVD likely caused by interstitial lung disease.” took me a while to understand. I think it would help the reader if you referred to the footnotes in Table 3 to identify the test combinations.

Response: We have included the direction to refer to the footnotes in table 3. “The prevalence of RVD determined with either spirometry or VA in combination (‘&’) with isolated reduction in DLCO represented RVD likely caused by interstitial lung disease (See footnotes in Table 3 for test combinations).”

---

I. The last paragraph of the Discussion is confusing to me. I am unsure of the basis of the statement that spirometry is a reasonable approach for the identification of lung restriction in this setting. It does not follow from the preceding sentences and is somewhat in conflict with the Conclusions.

I think that the last 3 sentences of the Discussion should be omitted. They are not pertinent to the objectives of the article and are not logical in my view. For example, because spirometry is common does not underscore its importance in longitudinal monitoring.

Response to previous two comments: We took out the entire paragraph in agreement with the reviewer’s comment and replaced it with what we believe is a more accurate statement: “Since spirometry is commonly used in assessing lung health in occupational settings, it is important to characterize estimates of varying conditions in cross-sectional studies. The further understanding of its performance in longitudinal settings will also enhance its use.”

Philip Harber, M.D., MPH
The paper is significantly improved both in terms of clarity and technical accuracy.

A. The authors may wish to consider whether the paper should be limited to spirometry alone rather than including the results of the single breath diffusing capacity and single breath helium dilution lung volume test. Several considerations include the following:

- Spirometry is widely used in worksites, whereas diffusing capacity and alveolar volume are not.
- When spirometry is complemented by additional tests, the six minute helium dilution lung volumes are generally measured along with the diffusing capacity. (This is commonly done since the equipment is generally constructed to perform this).
- As the authors point out in their discussion, estimating lung volumes (e.g., TLC) based upon the single breath used in the diffusing capacity test is frequently very inaccurate. It is particularly inaccurate in the presence of airflow obstruction, which is one of their major classifications. (The actual dilution TLC is occasionally inaccurate when compared with plethysmography as they point out, but this is much less common than estimating TLC from the single breath Va).
- When classifying physiologic patterns by spirometry, the authors follow well accepted definitions. Conversely, when they add the diffusing capacity test result, their classification differs from the more usual approach which considers the six minute dilution or plethysmography volume.
- Including these data makes the results and the discussion much more difficult for the reader to follow.

Response: The authors acknowledge the reviewer’s comment about the improved clarity that might be gained by limiting the focus to spirometry. However, we believe that the approach taken by utilizing spirometry, diffusing capacity and alveolar volume, though more challenging to grasp, is an important and unique aspect of the paper and should not be removed for that reason. We have tried to incorporate editorial suggestions whenever feasible to simplify the paper.

B. Abstract: They use the term “lung restriction”, but they probably mean “restrictive physiologic patterns”.

Response: While editing the objectives of the abstract, we used the more precise term, “restrictive ventilatory defect.”

C. Page 29 line 19: Do they mean “restrictive lung disease” or disease with restrictive physiologic pattern?

Response: We corrected this to the more precise term of “restrictive ventilatory defect” and have used this terminology in a consistent fashion throughout.

D. Page 29 line 22: (important): Some might not agree with their concepts in this paragraph. For example, body plethysmography is not often used because of its limited availability, but rather six minute equilibration helium dilution is employed. In addition, they define “RLD” by impairment of ventilation/oxygenation, whereas ventilation perfusion mismatch often occurs from either obstructive airway disease or from pulmonary vascular disease rather than “restrictive” abnormalities per se.

Response: We acknowledge this comment. The concepts in this paragraph were based on screening for lung disease in an occupational setting. In this setting, a key lung function test is spirometry. We refined the definition of RLD to be more specific and have defined it as, “Restrictive lung disease (RLD) refers to a decrease in total volume of the lungs, due to impaired expansion from decreased lung elasticity.”
E. Page 32 line 14: (minor): The sentence is poorly worded, suggesting that “quality checks” is a “common error”. They should rearrange the sentence.

Response: We removed the qualifier, “common” in the sentence and rearranged the sentence to read as follows: “Precautions were taken to avoid errors. These included carrying out regular quality checks of equipment and monitoring the procedural performance of technicians.”

F. Page 9, line 30: This paragraph is ambiguous when it discusses whether “lung volume estimation by helium dilution” refers to the usual six minute equilibration technique or the single breath method that is incidentally incorporated in diffusing capacity testing. They should clarify which term they mean every time they reference this.

Response: We corrected this by added the description “single-breath” before helium dilution to clarify the technique used. We made this clarification throughout the article. This is also described in the methods section as previously advised.

G. Page 11 line 29 (and elsewhere): The author should more carefully follow standard conventions when they refer to “VA” and “DLCO”. For example, these require subscripting. (If the journal to which submitted prefers full capitalization approach, this should of course be used)

Response: “VA” and “DLCO” are generally accepted terms for the expression of alveolar volume and diffusing capacity respectively. We acknowledge that expressions of these terms are varied in literature. We have used the same expressions as used in 2005 ATS/ERS documents pulmonary function testing. We replaced “DLCO” with “DL,CO” and “TLCpleth” with “VL,pleth” (Total lung capacity measured with body plethysmography) throughout the article.

H. Table 3: The formatting is markedly improved from the prior version.

Response: None Required.

I. Table 4: (Important): When they refer to “percentiles”, are they truly referring to percentiles within a distribution or do they really mean “percent of predicted” (which is the more commonly employed technique in PFTs)? (e.g., an individual whose result is 80% of predicted value is not at the 80th percentile within a population). If they truly mean percentiles, this should be described carefully under methods.

Response: We have added a footnote describing the conversion to percentiles. It is also described in the methods section. It reads, “Percentiles of spirometry performance are expressed as the FVC of each participant, standardized to the NHANES III population-based distribution of normal lung function. This is different from percent predicted which is the ratio of the FVC to the median predicted value for each participant (race, age, height and gender adjusted).”

J. Page 40 line 17 (minor): they mean “ventilator”, not “ventilator”.

Response: This has been corrected in the edits.

K. Page 41, line 18: The authors provide a nice discussion of the inaccuracy of the Va method in comparison to actual TLC measurement. This reviewer believes this supports focusing on the spirometry per se rather than confusing matters by classifying with the Va while knowing that the error is not random, but rather is confounded by the presence of disease.
Response: (See comments above under our response to Comment A). In addition, the authors believe that including multiple tests provides unique perspectives for spirometry use. Given its ease of use, and low cost, we thought it was important and of interest to compare estimates from spirometry with more complicated, expensive and less available tests. Our results provide insights that would be otherwise not discussed with limiting the discussion to spirometry.

L. Page 42 line 47 and elsewhere: The use of the term “lung restriction” should be carefully employed. For example, a low FVC will occur in the presence of diseases affecting muscles even though the lungs per se may be normal. They should consider using “restrictive physiologic pattern” or even use their term “RVD” rather than the less technically correct term of “lung restriction”.

Response: We have thoroughly gone over the article, taking care to use the more precise term, “RVD.”

M. In summary, the paper is significantly improved and has useful information to convey. Attention to some of these technical details and consideration of the relative utility/obfuscation of including diffusing capacity and single breath lung volume data should be considered.(certainly, if the authors strongly believe that these additional data are an inherent part of their paper, the editors should consider respecting the authors’ opinion)

Response: As above, it is our feeling that these tests provide context on how spirometry performs, relative to more complicated tests. These tests, when used in the proper context, (removing obstruction, combining to improve accuracy) suggest that spirometry estimates of restriction appear to be reasonable. We feel this is an important message to convey.
Estimates of restrictive ventilatory defect in the mining industry. Considerations for epidemiological investigations: a cross-sectional study

Nnaemeka U Odo, Jeffrey H Mandel, David M Perlman, Bruce H Alexander and Paul D Scanlon

BMJ Open 2013 3:
doi: 10.1136/bmjopen-2013-002561

Updated information and services can be found at:
http://bmjopen.bmj.com/content/3/7/e002561

These include:

References
This article cites 38 articles, 9 of which you can access for free at:
http://bmjopen.bmj.com/content/3/7/e002561#BIBL

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
- Epidemiology (2161)
- Occupational and environmental medicine (289)
- Public health (2275)
- Research methods (619)
- Respiratory medicine (351)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/