Racial discordance in spirometry comparing four commonly used reference equations to the National Health and Nutrition Examination Study III

Jacob Collen a,⁎, David Greenburg b, Aaron Holley c, Christopher King a, Stuart Roop a, Oleh Hnatiuk d

a Pulmonary/Critical Care Medicine, Walter Reed Army Medical Center, Washington, DC, USA
b General Internal Medicine, Madigan Army Medical Center, Tacoma, WA, USA
c Pulmonary/Critical Care/Sleep Medicine, Walter Reed Army Medical Center, Washington, DC, USA
d Pulmonary/Critical Care Medicine, National Institutes of Health, Bethesda, MD, USA

Received 8 June 2009; accepted 2 November 2009
Available online 19 November 2009

KEYWORDS
Spirometry; African—American; Discordance; Pulmonary function test

Summary
Diagnosing lung function abnormalities requires application of the appropriate reference equation for a given patient population. Current guidelines recommend the National Health and Examination Study III data set for evaluating patients in the United States. In Caucasian patients, relying on older reference equations, as opposed to those derived from the NHANES III data set, will often result in a different interpretation of a patient’s spirometry. The present study assessed whether similar discordance would occur in African—American patients.

A cross-sectional analysis of African—American patients undergoing spirometry testing at our hospital was performed. Patients were classified as normal, restricted, obstructed or mixed based upon the ATS/ERS guidelines, using Crapo, Knudson, Morris, Glindmeyer, and NHANES III prediction equations. Differences in classification were evaluated.

4463 subjects were identified, with a mean age of 49.6. Discordance in interpretation was most common when results from prediction equations by Morris, Knudson, and Glindmeyer were compared to NHANES III (24.6%, 26.4%, and 20.1%, respectively). Discordance was less common when comparing Crapo to NHANES III (12.8%). There was a tendency for Knudson, Morris and Glindmeyer to under classify restriction, and for Crapo, Morris, and Glindmeyer to over classify obstruction.

⁎ The views expressed in this paper are those of the authors and do not reflect the official policy of the Department of the Army, Department of Defense, or the US Government.

⁎ Corresponding author. Tel.: +1 703 966 1598; fax: +1 202 782 9032.
E-mail addresses: jacob.collen@amedd.army.mil (J. Collen), dgreenburg@usuhs.mil (D. Greenburg), aaron.holley@amedd.army.mil (A. Holley), christopher.king@amedd.army.mil (C. King), stuart.roop@amedd.army.mil (S. Roop), hnatiuko@nihlbi.nih.gov (O. Hnatiuk).
Introduction

In 2005, the American Thoracic Society (ATS) and the European Respiratory Society (ERS) published new, combined guidelines recommending the use of reference values from the National Health and Nutrition Examination Study III (NHANES III) data set for all patients in the United States between the ages of 8 and 80. Switching from older reference equations to NHANES III results in differences in disease classification and severity without an actual change in disease status in a Caucasian population. Our group recently performed a cross-sectional analysis of the non-Hispanic Caucasian patients in our hospital's spirometry database to assess for diagnostic reclassification when adopting the ATS/ERS guidelines. We found that reference equations from Crapo, Knudson, and Morris commonly under classify restriction ('restrictive pattern') and over classify obstruction in comparison to the NHANES III equations. This discordance in classification may create uncertainty for clinicians using spirometry to confirm or exclude pulmonary disease.

It is well known that differences in lung volumes exist between Caucasians and blacks in Africa. Historically, pulmonary laboratories have applied standard conversions to commonly used reference equations when testing African-American patients (multiplying the FVC and FEV1 by a conversion factor of 0.88). These standard conversions are based on prior studies showing that African-Americans have FEV1 and FVC values approximately 12% lower than Caucasians. These studies did not include patients older than 65 years, and a later study by Enright and colleagues found that African-Americans older than 65 years have FVC values approximately 6% lower than Caucasians. Reference equations for blacks have been created, but most are based on populations of blacks in Africa with only one set derived from a very specific, blue-collar African-American population in the United States. It is not clear whether any of these equations would be appropriate for use throughout the United States.

Prior to publication of NHANES III, most reference equations were derived from small, homogenous populations that were tested in the 1970s and 80s. A study of pulmonary training programs in the United States, published in 1990, indicated that most labs were using correction factors when evaluating African-American patients, as opposed to using reference equations derived specifically from African-American populations. In 1995, Glindmeyer et al published a set of reference equations derived from African-Americans working at a paper plant in Louisiana. It is unclear whether or to what extent occupational respiratory inhalant exposure could have impacted their results, although the authors due attempt to account for this in the manuscript. To our knowledge, with the exception of NHANES III, this is the largest study on normal spirometry values for African-Americans in the US.

We hypothesized that NHANES III would produce significantly different reference ranges for normal spirometry in African-Americans in the US, when compared to the older equations. Our goal was to assess for discordance between normal values defined by NHANES III versus racially corrected normal values defined by Crapo, Knudson, and Morris. Because it is not known how often they are used in the United States, we compared the reference ranges defined by the Glindmeyer equations to NHANES III as well.

Materials and methods

We conducted a retrospective review of all adult, African-American patients undergoing spirometry testing at Walter Reed Army Medical Center, from January 2000 to May 2007. Referrals for symptoms, pre-operative screening, and routine follow-up after treatment are all common, and testing is ordered by physicians from all services (a pulmonary consult is not required). The presence or absence of symptoms during testing was not documented. African-American patients were identified by self-report at the time of testing.

All spirometry (Vmax, 6200 Autobox DL, Sensormedics, Yorba Linda, CA) was performed according to American Thoracic Society (ATS) standards. Prior to 2005, spirometric end of test criteria was based on the 1994 ATS standards. Measured spirometric indices for individual patients were interpreted using the equations published by Crapo, Knudson, Morris, Glindmeyer, and NHANES III. Expected values for forced vital capacity (FVC) and forced expiratory volume in 1 s (FEV1) were multiplied by a conversion factor of 0.88 for Crapo, Knudson, and Morris equations. No conversion factor was required for the Glindmeyer and NHANES III equations because they were both derived directly from African-American populations. All patients were (re)classified using an algorithm similar to the ATS/ERS recommended algorithm for spirometric interpretation.

This study was designed to assess differences in pattern classification. A restrictive pattern was defined by an FVC outside the lower end of the 95% confidence interval, or lower limit of normal (LLN), calculated using the methods described in the Crapo, Knudson, Morris, Glindmeyer, and NHANES III studies. Severity of restriction was graded based on degree of decrement in FVC below the LLN, in accordance with the 1986 ATS guidelines. As such, an FVC of 60–80% predicted was graded as "mild severity," 50–59% predicted was "moderate," and less than 50% predicted was "severe." An obstructive defect was defined by an FEV1/FVC ratio below the lower end of the 95% confidence interval (LLN) for each reference equation, with severity based on the degree of decrement in FEV1 according to
ATS/ERS 2005 guidelines. As such, an FEV1 > 70% predicted was graded as ‘’mild severity,’’ 60–69% predicted as ‘’moderate,’’ 50–59% predicted as ‘’moderately severe,’’ 35–49% predicted as ‘’severe,’’ and < 35% predicted as ‘’very severe.’’ Patients with both an obstructive and a restrictive defect were categorized as having a mixed pattern, with severity based on the same scale of decrement in FEV1 that was used for those categorized with an obstructive defect.

This protocol was approved by the Department of Clinical Investigation at our hospital.

Statistics

Relationships between continuous and categorical variables were assessed using Student’s t test and chi-square test respectively. Agreement on interpretation of spirometry (normal, obstructive, restrictive, or mixed) between predictive models was assessed using kappa. Agreement on severity of obstruction and restriction was assessed using quadratic kappa. Reported p-values are two-sided. Statistical significance was defined as p < 0.05. All analyses were performed using STATA ver. 9.2 (StataCorp LP, College Station, TX). Multivariatic logistic regression was used to calculate adjusted odds ratios for discordance between interpretative strategies using equations by Crapo, Knudson, Morris, and Glindmeyer against those from NHANES III. Covariates included in logistic regression models included sex, age (<25, 25–34.9, 35–49.9, 50–64.9, 65–79.9, >80), BMI (<18.5, 18–24.9, 25–29.9, 30–40, >40), short stature (height in the lowest 2.5% for gender), or tall stature (height in the highest 2.5% for gender).

Results

14,390 pulmonary function tests were performed at our facility during the study period. The majority of subjects were Caucasians (n = 8733; 60.7%) with the remainder categorized as African–American (n = 4463; 31.0%), Hispanic (n = 667; 4.6%), and Asian (n = 527, 3.7%). This study is limited to exploring discordance among African–Americans. Demographic information on the patients included in our study is listed in Table 1.

Concordant classification was observed in 87.2%, 73.6%, 75.4%, and 80% when comparing Crapo, Knudson, Morris, and Glindmeyer, respectively, to NHANES III. The two most common patterns of reclassification were from normal by Knudson, Morris, and Glindmeyer to restricted by NHANES III, and from obstructed by Crapo, Morris, and Glindmeyer to normal by NHANES III. A third major category of reclassification was from restricted by Crapo to normal by NHANES III (Figs. 1 and 2).

The severity of restriction was mild for the majority of patients who were reclassified from normal by Knudson, Morris, or Glindmeyer to restrictive by NHANES III. For patients who were reclassified from obstructed by Crapo, Morris, or Glindmeyer to normal by NHANES III, the majority of defects were also mild.

Evaluation of agreement in diagnostic classification between NHANES III and Crapo was excellent (Kappa = 0.80), fair-to-good with Knudson (Kappa = 0.47), fair-to-good with Morris (Kappa = 0.53), and good-to-excellent with Glindmeyer (Kappa = 0.67). Measures of obstruction and restriction severity were excellent between Crapo, Knudson, Morris, and Glindmeyer and NHANES III (Kappas ranged from 0.81 to 0.95).

Classification of discordance between Crapo, Knudson, Morris, Glindmeyer, and NHANES III was also evaluated by gender. Agreement between classifications was excellent between Crapo and NHANES III for both males and females, but was higher in males (Kappa = 0.85), than in females (Kappa = 0.73). Females (n = 223, 41.3%) were more likely than males (n = 69, 13.2%) to be reclassified from restrictive by Crapo to normal by NHANES III. Agreement with Knudson was identical for males and females (Kappa = 0.47), while females were slightly less likely to be discordant for both Morris (kappa females = 0.54; kappa males = 0.52) and Glindmeyer (kappa females = 0.69; kappa males = 0.64).

Predictors of discordance included gender, older age, short stature, and obesity (Table 2). Females were less likely to be reclassified than males when comparing Knudson (OR = 0.66; 95% CI = 0.57–0.76), Morris (OR = 0.68; 95% CI = 0.59–0.78), and Glindmeyer (OR = 0.67; 95% CI = 0.57–0.78) to NHANES III. Females were more likely to be reclassified when comparing interpretations from Crapo to those from NHANES III (OR = 1.78; 95% CI = 1.48–2.13). Age greater than 65 predicted higher rates of discordance for all equations. Short stature was protective against discordance when using the Knudson, Morris and Glindmeyer equations. Being overweight (BMI > 30) predicted discordance when using the equations by Knudson and Morris.

Discussion

In this population of 4463 African–American patients, we found significant discordance in the spirometric interpretations between older reference standards and NHANES III. Discordant classification was found in over 20% of the

<table>
<thead>
<tr>
<th>Table 1 Patient demographics.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Males n = 2283 (51.2%)</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Height (cm)</td>
</tr>
<tr>
<td>Weight (kg)</td>
</tr>
<tr>
<td>BMI</td>
</tr>
<tr>
<td>FEV1</td>
</tr>
<tr>
<td>FVC</td>
</tr>
<tr>
<td>FEV1/FVC</td>
</tr>
<tr>
<td>Females n = 2180 (48.9%)</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Height (cm)</td>
</tr>
<tr>
<td>Weight (kg)</td>
</tr>
<tr>
<td>BMI</td>
</tr>
<tr>
<td>FEV1</td>
</tr>
<tr>
<td>FVC</td>
</tr>
<tr>
<td>FEV1/FVC</td>
</tr>
</tbody>
</table>
overall population, ranging from 12.8% (Crapo) to 26.4% (Knudson), and was higher in specific groups. Gender, age greater than 65, obesity (BMI > 30), and short stature were all predictors of discordance.

Both the number of reclassifications and the change in interpretation patterns are clinically important, with the potential for altered diagnoses without a change in clinical status, and flawed evaluation and management. For example, Crapo showed a tendency to over classify females as having a restrictive pattern in our population. This could lead to inappropriate evaluations for restrictive lung disease, to include inappropriate imaging and radiation exposure. Knudson, Morris and Glindmeyer tended to under classify restrictive patterns in our population, with the potential for missed diagnoses. Unless all laboratories are currently using NHANES III, individuals undergoing spirometry as part of an initial work-up may experience unnecessary referrals, or may not be referred when needed. With groups now recommending screening and office based spirometry for different clinical scenarios, results will increasingly be read by physicians not specifically trained in the nuances of spirometric interpretation. In this case, unexpected changes in reference standards could have a particularly large impact on missed diagnoses, and inappropriate referrals, tests, and treatments.

From a researchers’ perspective, widespread implementation of the new ATS/ERS guidelines may result in different patient populations being studied. The results from populations defined by the older criteria may not be applicable to current patient populations. Additionally, spirometric measures presented as percent predicted could potentially affect disability ratings. Lung transplantation referrals and pre-operative risk assessments may be affected when FEV₁ or FVC values fall near cut-offs required for eligibility.

The reasons for discordance between older studies and NHANES III cannot be determined with certainty. The populations studied by Crapo and Morris consisted mainly of members of The Church of Jesus Christ of Latter Day Saints (Mormons), who were selected for their decreased exposure to air pollution and tobacco smoke. The patients studied by Glindmeyer all worked at a paper processing plant in Louisiana and were of a similar socioeconomic status. In both cases, the populations would be considered homogenous in comparison to NHANES III, making their equations less generalizable to different groups across the US.

Technical and procedural differences could account for significant discordance. NHANES III required a five-maneuver minimum, and when they re-analyzed their own data using 1994 ATS guidelines (three acceptable curves

![Figure 1](image1.png)

**Figure 1** Discordant classifications.

![Figure 2](image2.png)

**Figure 2** Patterns of spirometric discordance.
with a reproducible test), predicted FVC and FEV₁ dropped by 62.5 cc and 52 cc respectively. Knudson, Glindmeyer, and Crapo used three acceptable curves, while Morris required only two, which could account for many patients being reclassified from normal to restrictive by NHANES III. There were also different methods used to calculate the appropriate start- and end of test criteria, as the ATS recommended using different techniques over time. Lastly, the Crapo reference equation utilized the largest FVC and FEV₁ sum rather than the largest value from separate efforts, as recommended by ATS. This may result in a 50-mL reduction in the FVC.¹² Other factors, including the smaller sized data sets, different age ranges, and older dates of publication (cohort effect) could also contribute to discordance.

The use of a single, standard correction factor to adjust for race has been questioned. In keeping with prior studies,⁸⁻¹²,¹⁶⁻¹⁸ the NHANES III authors found that African–American patients have a lower FEV₁ and FVC and an FEV₁/FVC ratio that is similar to or greater than that found in Caucasians. They agreed that the adjustment factor of 12–15% does approximate the difference between the two groups, with potential for some error at certain ages and heights. Enright and colleagues also showed that the commonly used 12% adjustment factor may not apply to elderly African–American patients (>65 years), and may overestimate predicted values in this age group.¹⁵ Others feel the wide variation in racial differences seen across studies makes using a single conversion factor inappropriate.¹⁹ The 2005 ATS/ERS consensus recommendations state that in certain situations a racial adjustment factor may be used, but that specific race/ethnic equations are preferred.

We anticipated that the more recently published, race specific equations from Glindmeyer’s study would have superior concordance to NHANES III than older reference equations that used a racial correction factor. As such, we were surprised to find that when we compared Crapo with a correction factor to NHANES III, the agreement was excellent, with less discordance than Glindmeyer. Glindmeyer only evaluated patients up to the age of 65, while Crapo and NHANES III evaluated older patients. We found that age greater than 65 was predictive of discordance (Table 2), which may explain why Glindmeyer had unexpectedly less agreement with NHANES III than Crapo, despite using equations based on African–Americans. Therefore, while the discordance in our study may in part be attributable to using standard correction factors, the above discrepancy between Crapo and Glindmeyer points to a greater impact of age.

Ultimately, racial differences in biomedical data can be difficult to interpret due to the necessary assumptions involved.²⁸,²⁹ Identity is not fixed or easily determined (i.e. biracial, multiracial; self-identification can be dynamic).²⁹ Categorizations based on skin color, language spoken, or country of origin, are imprecise and may change over time. Individual biases may also affect classification.²⁸ We lack an adequate understanding of the complex factors that account for differences in perceived groups (socioeconomic status, access to healthcare, quality of healthcare, racism, and exposure to occupational or environmental hazards).²⁹ A study by Harik-Khan et al.³⁰ failed to completely account for racial differences in lung function in African–American patients based on physical or socioeconomic factors alone. With regards to standardized medical reference equations and race, multiple factors can make an apparently homogenous population of “African–Americans” heterogeneous.³¹ This can limit the reliability of race based equations and racial conversion factors.

Our study has several limitations. This is a retrospective analysis of all screening pulmonary function tests done in our clinic over a seven year period. Our hospital serves

### Table 2

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Crapo 95% CI</th>
<th>Knudson 95% CI</th>
<th>Morris 95% CI</th>
<th>Glindmeyer 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>18–24</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>25–35</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>35–45</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>45–55</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>55–65</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>65–75</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>75–85</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>85–95</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>95–105</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < 0.05; |p < 0.001.
military beneficiaries, which include active duty service members, retirees, and their dependants. Our hospital is located in a large US city, and many of our beneficiaries live in the surrounding area. However, we also serve patients referred from other military facilities around the country as well as from overseas. Additionally, our population of African-Americans had an average body mass index of 28.9, which may be higher than the body mass index found in other populations, and potentially not generalizable to other US African-Americans. It is unclear what impact these factors had on our results.

In summary, we have presented the first study showing that discordance and spirometric misclassification (with all of the potential for incorrectly directed evaluation and management) exists in African-Americans when older reference equations are applied as opposed to NHANES III. The NHANES III data set is large and includes a broad demographic range sampled from across the United States, with specific equations derived from Caucasian, Hispanic, and African-American patients. As we have shown, there are significant and clinically relevant differences in prediction equation results for African-Americans when using NHANES III, compared to older reference equations. In the absence of reference values derived specifically from the patient population expected to utilize a given lab, drawing reference values from a large and racially diverse population, like the one used in NHANES III, seems prudent. Physicians need to be aware of the potential for spirometric discordance when interpreting spirometry in African-American patients.

Conflict of interest

The authors have no conflicts of interest to disclose.

References

