Prognostic utility of the 2011 GOLD classification and other multidimensional tools in Asian COPD patients: a prospective cohort study

Hiang Ping Chan1,2
Amartya Mukhopadhyay1,2
Pauline Lee Poh Chong1,2
Sally Chin1,2
Xue Yun Wong1,2
Venetia Ong1,2
Yiong Huak Chan3
Tow Keang Lim1,2
Jason Phua1,2

1Division of Respiratory and Critical Care Medicine, University Medicine Cluster, National University Health System, 2Department of Medicine, 3Biostatistics Unit, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Background: How well the 2011 Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification prognosticates for Asian patients with COPD is unknown.

Objective: The authors aimed to study the predictive utility of the GOLD 2011 classification for exacerbations and mortality as compared with other multidimensional tools in an Asian population.

Methods: In all, 1,110 COPD patients were prospectively followed between March 2008 and March 2013. They were classified using the 2011 and 2007 GOLD guidelines, modified Medical Research Council score, St. George’s Respiratory Questionnaire (SGRQ), and Body mass index, Obstruction, Dyspnea (BOD) index. Outcome measures were exacerbations and mortality. Multivariable survival analyses and receiver operating characteristic (ROC) curves were used to assess the different classification systems.

Results: Time-to-event analyses demonstrated earlier exacerbations in 2011’s GOLD D when compared with GOLD A (hazard ratio [HR] 0.54, 95% confidence interval [CI]: 0.31–0.95, P=0.032) and GOLD B (HR 0.62, 95% CI: 0.45–0.85, P=0.003) and higher mortality when compared with GOLD A (HR 0.37, 95% CI: 0.16–0.88, P=0.025) and GOLD B (HR 0.46, 95% CI: 0.31–0.70, P<0.001). The areas under the ROC curve for GOLD 2011, GOLD 2007, modified Medical Research Council score, St. George’s Respiratory Questionnaire, and BOD index were 0.62, 0.59, 0.61, 0.60, and 0.61, respectively, for the prediction of exacerbations and 0.71, 0.70, 0.71, 0.71, and 0.72, respectively, for the prediction of mortality (ROC comparator, P>0.05).

Conclusion: The 2011 GOLD classification predicts exacerbations and mortality moderately well in Asian COPD patients. Its prognostic utility is similar to that of other multidimensional systems.

Keywords: COPD, guideline, morbidity, mortality, prognosis

Introduction

Chronic obstructive pulmonary disease (COPD) is a complex condition with a heavy burden of disease and is one of the leading causes of mortality worldwide.1–2 The introduction of the multidimensional Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification system in 2011 recognized the complexity of this disease by recommending a more holistic approach than its predecessor from 2007, which focused primarily on forced expiratory volume in 1 second (FEV1). The current version also takes into consideration symptoms of patients through tools such as the modified Medical Research Council (mMRC) score, which is a simple grading system for severity of dyspnea, as well as exacerbation history, with two or more episodes of exacerbations over the past year deemed to be of high risk.3–5
Clearly, however, no classification system for COPD is perfect. GOLD 2011 has never been validated in Asian populations, despite its aim to be a global tool. The need for it to be validated is further underlined by differences in the epidemiology and pathophysiology of COPD in Asia compared to the rest of the world. Conflicting results have also emerged when outcomes in the GOLD groups B and C were compared, and available studies have not shown that GOLD 2011 is better than GOLD 2007 in terms of prognostication. Meanwhile, other tools for multidimensional assessment of patients that may predict morbidity and mortality in COPD patients exist. These include the use of health-related quality of life measures such as St. George’s Respiratory Questionnaire (SGRQ) and the Body mass index (BMI), Obstruction, Dyspnea, Exercise (BODE) index. We therefore aimed to study the utility of the GOLD 2011 classification in predicting exacerbations and mortality as compared with other multidimensional tools in a large Asian population of COPD patients.

Methods
Study design and patients
This is a prospective cohort study of all patients with COPD who were managed in our university hospital between March 2008 and March 2013. We diagnosed COPD in patients with dyspnea, chronic cough, and/or sputum production with at least 10 pack-years of smoking and persistent airflow limitation as evidenced by a postbronchodilator FEV₁/forced vital capacity ratio of <0.7. Trained COPD specialist nurses entered the following data into the institutional Central Clinical Research Database: age, sex, height, weight, BMI, and when clinically stable in the outpatient setting, spirometry results according to the American Thoracic Society guidelines, mMRC dyspnea score, and SGRQ. Outcome data on exacerbations, defined as an increase in symptoms not relieved by usual reliever medications and requiring either emergency department attendance and/or admission into the hospital, and mortality from the time of entry into the database until July 2013 were obtained from the national electronic medical record system.

Statistical analyses
We classified patients using five different systems: 1) current GOLD 2011 groups (A, B, C, or D) according to their FEV₁, mMRC score, and exacerbation history; 2) GOLD 2007 (GOLD 1: FEV₁ ≥80%; GOLD 2: 50%≤ FEV₁ <80%; GOLD 3: 30%≤ FEV₁ <50%; GOLD 4: FEV₁ <30%); 3) mMRC score in quartiles as described by Nishimura et al³ (Group 1: 0–1; Group 2: 2; Group 3: 3; Group 4: 4); 4) SGRQ total scores in quartiles as per Domingo-Salvany et al¹³ (Group 1: 0–20.94; Group 2: 20.95–30.62; Group 3: 30.63–44.91; Group 4: 44.92–100); and 5) BMI, Obstruction, Dyspnea (BOD) quartiles (BODE score without the 6-minute walk test) using a method similar to de Torres et al¹¹ (Group 1: BOD 0–2; Group 2: BOD 3; Group 3: BOD 4; Group 4: BOD 5–7).

We expressed categorical data as frequency and/or percentage and parametric data as mean ± standard deviation (SD). We compared groups using the chi-square test and analysis of variance, and applied the Bonferroni correction where appropriate. We performed survival analyses to determine if any significant differences in exacerbation and mortality were present between groups. We first used Kaplan–Meier analysis and the log-rank test, expressing results as mean ± standard error of mean. We then used Cox regression to adjust for possible confounding variables, with data expressed as hazard ratio (HR) and 95% confidence interval (CI). The adjusted predictive scores for mortality and exacerbation for receiver operating characteristic (ROC) analysis were obtained using logistic regression. We compared the resultant areas under the ROC curve (AUCs) using a method described by Hanley and McNeil. We used SPSS Statistics 17.0 (SPSS Inc., Chicago, IL, USA) and considered P<0.05 to be statistically significant.

Results
Among 1,110 recruited COPD patients, 188 (16.9%) were in GOLD A, 162 (14.6%) in GOLD B, 256 (23.1%) in GOLD C, and 504 (45.4%) in GOLD D. In all, 77.6% of subjects were Chinese, 16.4% were Malay, 5.0% were Indian, and 1.0% were of other ethnicities. There were significant differences in age, BMI, FEV₁, and SGRQ among the four GOLD groups (P<0.005, analysis of variance; Table 1).

Time to exacerbation was significantly shorter in GOLD C (3.49±0.16 years) and in particular in GOLD D (2.99±0.11 years) when compared with GOLD A (4.03±0.16 years) and GOLD B (3.68±0.16 years; P<0.005, log rank; Figure 1). After adjustment for confounders on Cox regression, earlier exacerbations were still noted in GOLD D.
Table 1 Baseline characteristics of the study population

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>GOLD A</th>
<th>GOLD B</th>
<th>GOLD C</th>
<th>GOLD D</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>188</td>
<td>162</td>
<td>256</td>
<td>504</td>
<td></td>
</tr>
<tr>
<td>Sex (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>90.4</td>
<td>85.2</td>
<td>93.0</td>
<td>87.3</td>
<td>0.041</td>
</tr>
<tr>
<td>Female</td>
<td>9.6</td>
<td>14.8</td>
<td>7.0</td>
<td>12.7</td>
<td></td>
</tr>
<tr>
<td>Race (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>80.3</td>
<td>85.8</td>
<td>73.4</td>
<td>75.6</td>
<td>0.1</td>
</tr>
<tr>
<td>Malay</td>
<td>14.9</td>
<td>8.0</td>
<td>18.8</td>
<td>18.5</td>
<td></td>
</tr>
<tr>
<td>Indian</td>
<td>3.7</td>
<td>4.3</td>
<td>6.6</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>1.1</td>
<td>1.9</td>
<td>1.2</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Age at entry</td>
<td>70.0±9.4</td>
<td>75.5±9.1</td>
<td>68.5±8.9</td>
<td>72.7±9.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.62±0.08</td>
<td>1.60±0.08</td>
<td>1.62±0.07</td>
<td>1.61±0.07</td>
<td>0.002</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>60.4±12.4</td>
<td>57.4±11.9</td>
<td>56.0±12.0</td>
<td>54.0±12.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>23.0±4.4</td>
<td>22.8±5.4</td>
<td>21.4±4.3</td>
<td>20.9±4.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FEV₁ (% predicted)</td>
<td>64.3±11.4</td>
<td>61.9±10.9</td>
<td>41.4±11.8</td>
<td>38.2±11.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FEV₁/FVC</td>
<td>56.3±7.7</td>
<td>55.9±8.2</td>
<td>47.4±10.5</td>
<td>45.9±10.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SGRQ score</td>
<td>23.2±11.1</td>
<td>34.5±16.2</td>
<td>27.8±14.1</td>
<td>40.5±17.8</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Notes: *Lower age vs GOLD B (P<0.001) and GOLD D (P=0.004). *Older age vs GOLD C (P<0.001) and GOLD D (P<0.001). *Shorter vs GOLD A (P=0.019) and GOLD C (P=0.003). *Lighter vs GOLD A (P<0.001). *Lighter vs GOLD A (P<0.001) and GOLD B (P=0.014). *Higher BMI vs GOLD C (P=0.016) and GOLD D (P=0.001). *Lower BMI vs GOLD A (P<0.001) and GOLD B (P<0.001). *Lower FEV₁ (% predicted) vs GOLD A (P<0.001) and GOLD B (P<0.001). Lower FEV₁, % vs GOLD A (P<0.001), GOLD B (P<0.001), and GOLD C (P=0.001). *Worse SGRQ score vs GOLD C (P<0.001). *Worse SGRQ score vs GOLD A (P<0.001), GOLD B (P<0.001), and GOLD C (P<0.001).

Abbreviations: GOLD, Global Initiative for Chronic Obstructive Lung Disease; BMI, body mass index; FEV₁, forced expiratory volume in 1 second; FVC, forced vital capacity; SGRQ, St. George’s Respiratory Questionnaire.

Figure 1 Adjusted Kaplan-Meier curves for exacerbation for GOLD 2011 (A), GOLD 2007 (B), mMRC quartiles (C), SGRQ quartiles (D), and BOD quartiles (E).

Abbreviations: GOLD, Global Initiative for Chronic Obstructive Lung Disease; mMRC, modified Medical Research Council; SGRQ, St. George’s Respiratory Questionnaire; BOD, Body mass index, Obstruction, Dyspnea.
compared to GOLD A (HR 0.393, 95% CI: 0.279–0.552, \(P<0.005\)), GOLD B (HR 0.670, 95% CI: 0.498–0.901, \(P=0.008\)), and GOLD C (HR 0.714, 95% CI: 0.556–0.917, \(P=0.008\)). The effects of age at entry (HR 1.010, 95% CI: 0.999–1.021, \(P=0.068\)), sex (\(P=0.078\)), BMI (\(P=0.900\)), and race (\(P=0.660\)) were nonsignificant.

Patients in GOLD D had the shortest time to mortality (4.24 ± 0.08 years) when compared with the other groups (GOLD A 4.90 ± 0.09, GOLD B 4.50 ± 0.10, GOLD C 4.65 ± 0.10; \(P<0.005\), log rank; Figure 2). GOLD B had a shorter time to mortality compared to GOLD C. On multivariable Cox regression, time to mortality remained shorter in GOLD D than in GOLD A (HR 0.31, 95% CI: 0.17–0.56, \(P<0.005\)) and GOLD C (HR 0.65, 95% CI: 0.45–0.95, \(P=0.025\)). Age at entry (HR 1.04, 95% CI: 1.03–1.06, \(P<0.005\)) and BMI (HR 0.95, 95% CI: 0.92–0.98, \(P=0.001\)) were significant confounders, but not race (\(P=0.59\)) and sex (\(P=0.35\)). There was also a trend toward worse mortality (HR 0.77, 95% CI: 0.52–1.14) when GOLD D was compared to GOLD B, albeit without reaching statistical significance.

The other classifications systems, namely, GOLD 2007, the mMRC score, SGRQ, and BOD index, were also able to demonstrate significant differences between grades (Figures 1 and 2). With the exception of mMRC (Group 1 – HR 0.62, 95% CI: 0.36–1.06, \(P=0.617\); Group 2 – HR 0.95, 95% CI: 0.56–1.62, \(P=0.853\); Group 3 – HR 1.09, 95% CI: 0.64–1.87, \(P=0.752\)), all the other systems demonstrated GOLD 4 or their equivalent groups to have significantly earlier exacerbations compared to milder groups (GOLD 2007: GOLD 1 – HR 0.21, 95% CI: 0.08–0.59, \(P=0.003\); GOLD 2 – HR 0.65, 95% CI: 0.48–0.89, \(P=0.008\); GOLD 3 – HR 0.91, 95% CI: 0.69–1.21; SGRQ quartiles: Group 1 – HR 0.58, 95% CI: 0.44–0.77, \(P<0.001\); Group 2 – HR 0.70, 95% CI: 0.53–0.92, \(P=0.010\); Group 3 – HR 0.95, 95% CI: 0.73–1.22, \(P=0.669\); BOD quartiles: Group 1 – HR 0.53, 95% CI: 0.40–0.70, \(P<0.001\); Group 2 – HR 0.60, 95% CI: 0.45–0.80, \(P<0.001\); Group 3 – HR 1.02, 95% CI: 0.78–1.33, \(P=0.885\)). In terms of mortality, GOLD/Group 4 had a significantly shorter time to mortality when compared to GOLD/Groups 1 and 2 in all systems except for GOLD 2007 (GOLD 2007: GOLD 1 – HR 0.00, 95% CI: 0.00–0.00,
the highest in GOLD D, followed by GOLD B. Meanwhile, increased from group A to group D, the risk of death was largely in keeping with several other studies done on Caucasian populations. The need to study the prognostic abilities of all five systems (GOLD 2011, GOLD 2007, mMRC quartiles, SGRQ, and BOD) for exacerbation (AUCs were 0.62±0.02, 0.59±0.02, 0.61±0.02, 0.60±0.02, and 0.61±0.02, respectively; (A)) and mortality (AUCs were 0.71±0.02, 0.70±0.02, 0.71±0.02, 0.71±0.02, and 0.72±0.02, respectively; (B)), with no significant difference demonstrated in the predictive abilities of all five systems \( (P>0.05) \).

**Discussion and conclusion**

While the 2011 GOLD guidelines are aimed at the global community, they have not previously been validated in Asian populations. The need to study the prognostic ability of the new classification is underscored by differences in the epidemiology and pathophysiology of COPD in Asia compared to the rest of the world.5,17 In our study, multivariable survival analysis showed that race did not affect exacerbations or mortality. Indeed, our findings are largely in keeping with several other studies done on Caucasian populations.8–10 Although the risk of exacerbations increased from group A to group D, the risk of death was the highest in GOLD D, followed by GOLD B. Meanwhile, there was no difference in the prognostic abilities of all the classification systems studied.

In the previous spirometry-based GOLD classification from 2007 and before, higher stages conferred higher mortality.18,19 However, recent studies using the new GOLD classification have shown differing results. Leivseth et al and Soriano et al reported higher mortality with higher GOLD grades.9,10 In the ECLIPSE study, mortality was worst in GOLD D, best in GOLD A, and similar in GOLD B and C.7 Our study shares similar findings to those of Lange et al, de Torres et al, and Johannessen et al who demonstrated that survival is better in the more severe (worse lung function but milder symptoms) GOLD C group when compared to the less severe GOLD B group.8,11,12 Interestingly, despite the fact that age was not an independent predictor of outcomes in our study, our patients in group B were older than those in group C, which was also the case in the studies by Lange et al and Johannessen et al.8,12 These findings suggest that it may be useful to consider a multidimensional approach like GOLD 2011 in the clinical setting even though spirometry-based classification has similar predictive utility given that GOLD 3 grade may not always be a portent of worse outcome compared to GOLD 2 grade. In sum, we postulate that increasing dyspnea is associated with poorer outcomes because it reflects not only the severity of COPD as well as its resultant effect of sarcopenia from reduced mobility but also other comorbidities in an aged population.8,20

This study also attempted to examine if there are any other classification systems that can predict exacerbations...
and mortality with greater accuracy than the current GOLD 2011 categories. Multiple studies have so far shown that GOLD 2011 has similar capacity in prognosticating COPD patients as GOLD 2007.\(^7\)\(^,\)\(^11\)\(^,\)\(^12\)\(^,\)\(^21\) However, Lange et al reported that GOLD 2011 is better at predicting exacerbation, while Leivseth et al reported that GOLD 2007 is better able to predict mortality.\(^8\)\(^,\)\(^9\) There have also been studies utilizing other multidimensional tools in assessing outcomes. Domingo-Salvany et al demonstrated that health-related quality of life measures can be used to prognosticate mortality in COPD as patients with worse SGRQ and 36-item short form health survey (SF-36) scores were found to have higher mortality.\(^13\) Nishimura et al reported that mortality increased with increasing mMRC scores.\(^5\) A recent study showed that patients with higher BODE score or COPD Comorbidity Test index had higher mortality.\(^11\) In our study, all the multidimensional systems evaluated had moderate ability to predict exacerbations and mortality, although the AUCs were better for the latter than the former, while none were found to be superior or inferior to the GOLD 2011 system.

This study has several strengths. We followed a large cohort of Asian patients over a period of 5 years. The population included a good mix of patients from different GOLD groups, in comparison to previous studies in which group C was underrepresented.\(^8\)\(^,\)\(^12\)\(^,\)\(^21\) We prospectively recorded not only the GOLD but also other classification systems. We used the national electronic medical record system that enabled us to retrieve clinical data pertaining to a patient’s attendance at emergency departments and/or admissions to all the public hospitals in Singapore. Our study also has several limitations. Attendances at private hospitals may have been missed, although this is rare in our setting and should not affect the findings. Our definition of exacerbations required either emergency department attendance or hospitalization, and thus, excludes mild exacerbations. This may have underestimated the ability of GOLD 2011 when compared to GOLD 2007 in predicting exacerbations given that patients’ history of exacerbations had been demonstrated to be the best predictor of future exacerbations by Hurst et al.\(^22\) The COPD Assessment Test score was not used to categorize patients according to GOLD 2011 as we only recorded the mMRC score prospectively. The BOD index was used rather than the BODE index as 6-minute walk tests were not routine. Further, our study population appears to be lean and predominantly male. This can be explained by the trend that in Asia, a higher proportion of smokers tend to be male and also that Asians tend to have lower BMI of up to three to four units compared to their Caucasian counterparts.\(^23\)\(^-\)\(^25\)

Finally, data on other comorbidities and the exact causes of death were unavailable.

In conclusion, the GOLD 2011 classification has a modest ability to predict outcomes in COPD in an Asian population. While the risk of exacerbations increased from group A to group D, the risk of death was the highest in GOLD D, followed by GOLD B. It must be emphasized that the new GOLD categories are used first and foremost as a guide to facilitate disease management. While simpler systems such as GOLD 2007 and mMRC scores have similar predictive utility as GOLD 2011, they are currently not recommended to be used in isolation for disease management. Therefore, given that no other systems, including GOLD 2007, the mMRC score, SGRQ, and the BOD index, were found to be superior, we suggest that it is appropriate to use GOLD 2011 concurrently for prognostication while it is being used in its primary role of guiding COPD management.

**Acknowledgment**

No funding was received for this study.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


23. Deurenberg P, Deurenberg-Yap M, Guricci S. Asians are different from Caucasians and from each other in their body mass index/body fat percent relationship. Obes Rev. 2002;3(3):141–146.
