Translation of Berlin Questionnaire to Portuguese language and its application in OSA identification in a sleep disordered breathing clinic

A.P. Vaz*, M. Drummond, P. Caetano Mota, M. Severo, J. Almeida, J.C. Winck

Department of Pulmonology, H.S. Joao, Porto, Portugal
Department of Pulmonology, H.S. Joao, University of Porto Medical School, Porto, Portugal
Department of Hygiene and Epidemiology, University of Porto Medical School, Institute of Public Health of the University of Porto, Porto, Portugal

Received July 11, 2010; accepted September 12, 2010

KEYWORDS
Berlin Questionnaire; Obstructive sleep apnea; Screening; Diagnosis

Abstract
Background: Berlin Questionnaire (BQ), an English language screening tool for obstructive sleep apnea (OSA) in primary care, has been applied in tertiary settings, with variable results.
Aims: Development of BQ Portuguese version and evaluation of its utility in a sleep disordered breathing clinic (SDBC).
Material and methods: BQ was translated using back translation methodology and prospectively applied, previously to cardiorespiratory sleep study, to 95 consecutive subjects, referred to a SDBC, with OSA suspicion. OSA risk assessment was based on responses in 10 items, organized in 3 categories: snoring and witnessed apneas (category 1), daytime sleepiness (category 2), high blood pressure (HBP)/obesity (category 3).
Results: In the studied sample, 67.4 % were males, with a mean age of 51 – 13 years. Categories 1, 2 and 3 were positive in 91.6, 24.2 and 66.3 %, respectively. BQ identified 68.4 % of the patients as being in the high risk group for OSA and the remaining 31.6 % in the low risk. BQ sensitivity and specificity were 72.1 and 50 %, respectively, for an apnea-hipopnea index (AHI) > 5, 82.6 and 44.8 % for AHI > 15, 88.4 and 39.1 % for AHI > 30. Being in the high risk group for OSA did not influence significantly the probability of having the disease (positive likelihood ratio [LR] between 1.44-1.49). Only the items related to snoring loudness, witnessed apneas and HBP/obesity presented a statistically positive association with AHI, with the model constituted by their association presenting a greater discrimination capability, especially for an AHI > 5 (sensitivity 65.2 %, specificity 80 %, positive LR 3.26).
Introduction

Obstructive Sleep Apnea (OSA) syndrome is a common disorder that affects 9 to 24% of adult middle-aged population, characterized by repeated episodes of upper airway obstruction during sleep, intermittent arterial oxygen desaturation, increasing respiratory efforts and sleep disruption.  

Diagnosis suspicion of OSA is of particular relevance, since it has been implicated with high medical and peri-operative morbidity and mortality and an effective treatment available.  

Polysomnography is the gold standard test for OSA diagnosis, although its availability, as also the cardiorespiratory sleep study, may be limited. In fact, with the growing recognition of OSA epidemiological relevance and pathophysiological consequences, the medical community has been confronted with a rise in requests for the diagnostic test, with implications regarding resources rationalization. With this in mind, screening tools, such as questionnaires and clinical models, have been developed, trying to combine different OSA risk factors that in a cost-effective way could help clinicians in identifying patients that should be referred or have priority on the waiting list for diagnostic testing.

Berlin Questionnaire (BQ), an outcome from the Sleep in Primary Care Conference, in April 1996 in Berlin, is one of the most recognized screening tools used in this area. It includes 10 items organized in 3 categories: snoring and witnessed apneas (category 1), daytime sleepiness (category 2), high blood pressure (HBP)/obesity (category 3).

Conclusions: The BQ is not an appropriate screening tool for OSA in a SDBC, although snoring loudness, witnessed apneas, HBP/obesity have demonstrated being significant questionnaire elements in this population.

'2010 Published by Elsevier Espaæa, S.L. on behalf of Sociedade Portuguesa de Pneumologia.
gender, weight, height, neck circumference and ethnicity. Predetermination of high or lower risk for OSA is based on responses to each category of items.\(^{14}\)

Netzer and colleagues validated the BQ as a screening tool in primary care, where it demonstrated a high internal validity (Cronbach alpha correlations 0.86-0.92) and performed accurately with a higher sensitivity and specificity in OSA identification (86 and 77 %, respectively, for an apnea/hypopnea index [AHI] > 15),\(^{15}\) and 74.3 and 53.3 % for AHI > 15, 79.5 and 48.6 % for an AHI > 10, 67.2 and 52.8 % for an AHI > 5, and 65.6 and 60 % respectively for an AHI > 5, and 65 % for an AHI > 10).\(^{16}\)

More recently, BQ had been applied in tertiary settings with variable results.\(^{15-19}\)

In general cardiovascular patients, namely in a subgroup with atrial fibrillation, BQ showed a good performance (sensitivity of 86 % and specificity of 89 % for an AHI > 5),\(^{16}\) although in another subgroup with resistant hypertension, an adapted BQ version presented a lower specificity (sensitivity of 85.5 % and specificity of 65 % for an AHI > 10).\(^{16}\)

The same occurred in surgical patients, where BQ demonstrated a high moderate-sensitivity and specificity in OSA identification (86 and 77 %, respectively, for an AHI > 5, 74.3 and 53.3 % for AHI > 15, 79.5 and 48.6 % for AHI > 10).\(^{17}\)

Lastly, low sensitivity and specificity were found in patients undergoing pulmonary rehabilitation (62.5 and 53.8 %, respectively, for an AHI > 10, 67.2 and 52.8 % for an AHI > 15),\(^{18}\) as well as in a sleep clinic population (68 and 49 % respectively for an AHI > 5, 62 and 43 % for an AHI > 10, 57 and 41 % for an AHI > 15).\(^{18}\)

It should be noted that the last study, besides having a retrospective nature, took place in a general sleep clinic population, so BQ has never been applied as OSA screening tool in a specific sleep disordered breathing clinic (SDBC).

In addition, BQ was originally developed in English, being necessary to translate this questionnaire to other languages, so that it can be applied in other countries and contexts.

The main goals of this study were: 1) development of BQ Portuguese version; 2) assessing the usefulness of BQ in OSA identification, compared with AHI obtained by cardiorespiratory sleep study, in a SDBC. As a secondary objective, authors analyzed the association between the BQ items and AHI.

### Material and methods

#### Translation of BQ to Portuguese language

As suggested by back translation methodology,\(^{20}\) BQ was submitted to a process of translation, back translation and final review of obtained versions. Questionnaire was translated into Portuguese by two bilingual translators (two pulmonary specialists doctors) working independently, thus generating two Portuguese versions. These versions were translated into English by another two bilingual individuals and then compared with the original English and discussed in order to carry out necessary adjustments to obtain a single Portuguese BQ version (Annex 1), thus ensuring meaning equivalence.

#### Application of BQ in a SDBC

##### Population and study design

A prospective observational study was conducted during November and December 2008. Ninety consecutive patients referred to a SDBC of a university hospital pulmonology department, with suspicion of OSA, who underwent cardiorespiratory sleep study and completed the translated BQ Portuguese version, were included.

##### OSA diagnosis

A domiciliary six-channel cardiorespiratory sleep study (Alphascreen; Vyasis) was the test performed for OSA diagnosis. Oroonasal air flow and snoring (nasal cannula and microphone), pulse rate and arterial oxygen saturation (finger pulse oximetry), thoracoabdominal movements and body position (impedance belts) were the analyzed variables. The sleep data recorded by the device were manually scored by counting apnea events (air flow cessation lasting for at least 10 seconds) and hypopnea episodes (events of air flow reduction to 20 to 50 % of the previously observed lasting for at least 10 seconds, joined with a 4 % dip in oxygen saturation), dividing the total number of these episodes by the sleep time in hours, thus obtaining the manual AHI. According to established recommendations,\(^2\) the AHI was the gold standard criteria used for OSA diagnosis and severity defining (AHI > 5/h, 15/h and 30/h respectively considered mild, moderate and severe OSA).

##### BQ score

The BQ was scored (Annex 1) as previously reported by Netzer and colleagues.\(^{14}\) For items in categories 1 and 2, one point is assigned in the presence or occurrence of a symptom in a persistent or frequent way (3-4 times per week). Item 5, about witnessed apneas, is an exception, so for the same assumptions are awarded two points. Category 2 presents an additional item concerning frequency of drowsiness behind the wheel (item 9) for which no punctuation is consigned. Categories 1 and 2 are positive when the sum of all items punctuation is equal to or greater than two, and category 3 in the presence of hypertension (systolic or diastolic blood pressure > 140/90 mmHg) and/or obesity (Body Mass Index [BMI] > 30). Positivity in two or three categories defines a high risk score for OSA, while positivity in only one or none defines low risk.

#### Statistical analysis

Data were described as mean – standard deviation for quantitative variables and as counts for proportions. Statistical analysis was performed using the SPSS software (SPSS, Inc., Chicago, Illinois, USA).

BQ performance, for each cut-off of diagnosis gold standard, was evaluated by calculation of sensitivity, specificity, predictive values, LR, odds ratio and their 95 % confidence intervals, as also under the curve area (AUC).

The LR provides a direct estimation of how the BQ score changes the odds of disease: > 10 large change, 5-10 moderate change, 2-5 small change, 0.5-2 little or no change, 0.2-0.5 small change, 0.1-0.2 moderate change, < 0.1 large change.
BQ discrimination capability was considered insufficient if the area under the ROC curve was less than 0.6, acceptable when between 0.6 and 0.8 and excellent if above 0.8. Linear regression and respectively regression coefficients were used to estimate the association between each item and the gold standard. AHI was logarithm because it showed a skewed distribution. A separated model for each category of items was estimated, as also a global model using only the items that showed a statistically positive association with AHI (P < .1). This last model was used to calculate the modified Berlin score.

Results

General characteristics of population

Ninety-five individuals, more often male (67.4%; n = 64), with ages between 20 and 79 years (51 – 13) were evaluated. Studied population characteristics are presented in Table 1.

Cardiorespiratory study confirmed OSA diagnosis in the high risk group for OSA and the remaining 31.6% (n = 30), in the low risk (Table 2). In subjects with a high risk score, OSA diagnosis was confirmed by cardiorespiratory study in 87.7% (n = 57), while in those with a low-risk score, diagnosis was excluded in only 26.7% (n = 8). Most of the subjects in the high risk group, 35.4% (n = 23), had severe OSA, 29.2% (n = 19) presented mild OSA and 23.1% (n = 15) moderate. Of the subjects in the low risk group, the majority, 46.7% (n = 14), had mild OSA, followed by those 26.7% (n = 8), in which the syndrome was excluded. The global agreement between BQ score and AHI was 68.4% (p = 65).

The sensitivity and the specificity of BQ for OSA diagnosis were 72.1 and 50%, respectively, for an AHI > 5, 82.6 and 44.8% for an AHI > 15, 88.4 and 39.1% for an AHI > 30 (Table 3).

A small change in the probability of not having moderate or severe OSA (negative LR of 0.39 [AHI > 15] and 0.29 [AHI > 30]) was found in subjects with a BQ low risk score. In the remaining situations, BQ score demonstrated only a little or no change in the disease probability.

The obtained AUC of 0.611, 0.637 and 0.638, respectively for different AHI cut-offs, showed that BQ discrimination capability is within the limit of acceptable.

Discrimination capability of BQ items

Univariate linear regression demonstrated that in category 1, only the items 2 and 5 showed a statistically significant positive association with AHI that remained after multivariable adjustment (Table 4). In relation to items in category 2, none showed a significant crude or adjusted association with the gold standard. Category 3 showed a significant positive association with AHI in both, the univariate and multivariable analysis (Table 4).

A final model, considering only the items that showed a positive statistically significant association with AHI, was estimated, verifying an increase in BQ discrimination capability for all AHI cut-offs (AUC of 0.812, 0.73, 0.695, respectively) (Table 4).

Considering this last model, a modified Berlin score was calculated: two points for items 2 and 10, one point for item 5. The modified Berlin score demonstrated again a greater discrimination capability compared with the original BQ, especially for an AHI > 5, (sensitivity 65.2%; specificity 80%, positive LR 3.26, AUC 0.795) (Table 5).

Discussion

In face of most SDBC having long waiting lists, it seems attractive to use a screening tool to prioritize patients needing OSA diagnosis test according to probability that they might have a positive result.

When applied to a SDBC population, BQ demonstrated not to be a good test neither for OSA diagnosis nor for...
severity defining, presenting for different AHI cut-off points a moderate to high sensitivity (72.1 to 88.4%), a specific which in whole was low (39.1 to 50%) and a lower positive LR (1.44 to 1.49).

Our findings reflected, on the one hand, the high frequency of OSA in this population (83.2%), which influences the screening test operating characteristics, as also the large number of both false positives and false negatives demonstrated by BQ.

Most enrolled subjects (68.4%) had a BQ high risk score, as a result of population pre selection at the time of consultation referral, which is predominantly composed of snoring, hypertensive and obese subjects, as noted by the high prevalence of positivity in categories 1 (91.6%) and 3 (66.3%).

Besides the above, and although 87.7% of patients with BQ high risk score presented OSA, the same has not happened in individuals at low risk, where the sleep study ruled out OSA in only 26.7% of subjects, contributing to poor overall agreement between the two tests (68.4%).

Although polysomnography remains the gold standard, cardiorespiratory sleep study, having demonstrating a high accuracy in OSA diagnosis; and like occurred in the original study of Netzer, was the test applied.

BQ performance in this SDBC population is in accordance with results obtained in other studies undertaken in tertiary health care setting, particularly in resistant HBP and surgical patients. In patients undergoing pulmonary rehabilitation and in the retrospective study that took place in a general sleep clinic, the obtained results, compared with those in our population, were even less satisfactory. Noteworthy, in the last case, the association of the sleep clinic with the psychiatry department of a tertiary hospital, which contributed to the high proportion of patients with positivity in category 2 (74.6%) and was postulated by the authors as may have influenced BQ performance. To confirm this hypothesis a shortened BQ version was developed, excluding category 2 and item 4 of category 1, in this case because it was poorly answered. Although BQ sensitivity has increased, specificity remained low (80 and 42% respectively, for an AHI>5), similar to the obtained in the original questionnaire, corroborating the importance of other factors, in addition to daytime sleepiness, in the clinical picture of this syndrome.

### Table 3 Predictive parameters of BQ

<table>
<thead>
<tr>
<th></th>
<th>AHI &gt; 5 (95% CI)</th>
<th>AHI &gt; 15 (95% CI)</th>
<th>AHI &gt; 30 (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity %</td>
<td>72.1 (61.7-81.1)</td>
<td>82.6 (70.2-91.5)</td>
<td>88.4 (72.9-96.9)</td>
</tr>
<tr>
<td>Specificity %</td>
<td>50 (27.8-72.2)</td>
<td>44.8 (31.7-58.6)</td>
<td>39.1 (28.3-50.7)</td>
</tr>
<tr>
<td>PPV, %</td>
<td>87.7 (78.3-94.1)</td>
<td>58.4 (46.4-69.8)</td>
<td>35.4 (24.6-47.3)</td>
</tr>
<tr>
<td>NPV, %</td>
<td>26.7 (13.4-43.5)</td>
<td>73.3 (56.4-86.6)</td>
<td>90 (76.1-97.4)</td>
</tr>
<tr>
<td>Positive LR</td>
<td>1.44 (0.87-2.40)</td>
<td>1.49 (1.13-1.99)</td>
<td>1.45 (1.15-1.84)</td>
</tr>
<tr>
<td>Negative LR</td>
<td>0.56 (0.3-1.02)</td>
<td>0.39 (0.19-0.78)</td>
<td>0.29 (0.1-0.89)</td>
</tr>
<tr>
<td>Odds ratio</td>
<td>2.59 (0.86-7.9)</td>
<td>3.87 (1.55-10.5)</td>
<td>4.93 (1.52-22.2)</td>
</tr>
<tr>
<td>AUC</td>
<td>0.611</td>
<td>0.637</td>
<td>0.638</td>
</tr>
</tbody>
</table>

AHI: apnea/hipopnea index; AUC: area under the curve; BQ: Berlin Questionnaire; LR: likelihood ratio; NPV: negative predictive value; PPV: positive predictive value.

### Table 4 Linear regression of BQ items with statistically significant positive association with AHI logarithm

<table>
<thead>
<tr>
<th></th>
<th>Crude coefficient (95% CI)</th>
<th>Adjusted coefficient (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Snoring loudness</td>
<td>0.67 (0.2-1.14)</td>
<td>0.79 (0.34-1.25)</td>
</tr>
<tr>
<td>5. Witnessed apnea</td>
<td>0.53 (0.07-0.99)</td>
<td>0.36 (0.04-0.77)</td>
</tr>
<tr>
<td>Category 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. HBP/obesity</td>
<td>0.91 (0.53-1.28)</td>
<td>0.77 (0.4-1.14)</td>
</tr>
<tr>
<td>AUC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHI &gt; 5</td>
<td>0.812</td>
<td></td>
</tr>
<tr>
<td>AHI &gt; 15</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>AHI &gt; 30</td>
<td>0.695</td>
<td></td>
</tr>
</tbody>
</table>

AHI: apnea/hipopnea index; AUC: area under the curve; BQ: Berlin Questionnaire; HBP: high blood pressure.

### Table 5 Performance of modified BQ

<table>
<thead>
<tr>
<th></th>
<th>AHI &gt; 5 (95% CI)</th>
<th>AHI &gt; 15 (95% CI)</th>
<th>AHI &gt; 30 (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity %</td>
<td>65.2 (54-75.5)</td>
<td>78.5 (64.8-88.9)</td>
<td>81.8 (63.1-93.7)</td>
</tr>
<tr>
<td>Specificity %</td>
<td>80</td>
<td>62.2</td>
<td>50.7</td>
</tr>
<tr>
<td>PPV, %</td>
<td>94 (85.1-98.4)</td>
<td>66 (52.4-77.9)</td>
<td>36 (23.7-49.6)</td>
</tr>
<tr>
<td>NPV, %</td>
<td>32.4</td>
<td>75.7</td>
<td>89.2</td>
</tr>
<tr>
<td>Positive LR</td>
<td>3.26</td>
<td>2.07</td>
<td>1.66</td>
</tr>
<tr>
<td>Negative LR</td>
<td>0.43</td>
<td>0.34</td>
<td>0.35</td>
</tr>
<tr>
<td>Odds ratio</td>
<td>7.52</td>
<td>6.04</td>
<td>4.64</td>
</tr>
<tr>
<td>AUC</td>
<td>0.795</td>
<td>0.72</td>
<td>0.671</td>
</tr>
</tbody>
</table>

AHI: apnea/hipopnea index; AUC: area under the curve; BQ: Berlin Questionnaire; LR: likelihood ratio; NPV: negative predictive value; PPV: positive predictive value.
In fact, when analyzing the association of category 2 with AHI, the same authors found no significant correlations. As previously reported, in our population, none of the category 2 items was associated with the AHI and only items, 2 and 5, of category 1, referring to snoring loudness and witnessed apneas, and category 3, referring to HBP/obesity, were significantly associated to the diagnosis gold standard.

The model established by these three BQ items association had a greater ability to discriminate OSA for all AHI cut-off points, compared to the original BQ, especially for an AHI > 5. The abbreviated BQ version and the modified score confirmed these findings, although it didn’t present enough predictive accuracy, so that only these items could be recommended in OSA identification in a SDBC.

A recent meta-analysis of screening tests for OSA corroborated some of these data, having found that BMI, history of hypertension and nocturnal choking are significant test elements in the more accurate prediction models. Besides, it also demonstrated that the Epworth sleepiness scale was the least accurate predictive questionnaire used in this ambit, possibly because excessive daytime sleepiness occurs commonly in obese individuals without OSA, driven by mechanism other than nighttime sleep deprivation. 23

Although BQ remains the most accurate questionnaire for predicting OSA diagnosis, it is not an appropriate screening test for a high risk population in a SDBC. BQ was translated into Portuguese language and can thus be used by our country clinicians, in the contexts in which it is validated.

Conflict of interest

Authors declare that they don’t have any conflict of interest.

References

Annex 1 Berlin Questionnaire

| Altura _____m | Peso _____kg | Idade _____ | Sexo Masculino/Feminino |

Escolha a resposta correcta para cada questão:

**Categoria 1:**

1. Ressona?
   a. Sim
   b. Não
   c. Não sei

Se ressona:

2. O seu ressonar Ø:
   a. Ligeiramente mais alto do que a sua respiração
   b. Tão alto como quando fala
   c. Mais alto do que quando fala
   d. Tão alto que pode ser ouvido noutras divisões da casa

3. Com que frequência ressona?
   a. Quase todos os dias
   b. 3-4 vezes por semana
   c. 1-2 vezes por semana
   d. 1-2 vezes por mês
   e. Nunca ou quase nunca

4. O seu ressonar alguma vez incomodou outras pessoas?
   a. Sim
   b. Não
   c. Não sei

5. Alguma pessoa notou que parava de respirar durante o sono?
   a. Quase todos os dias
   b. 3-4 vezes por semana
   c. 1-2 vezes por semana
   d. 1-2 vezes por mês
   e. Nunca ou quase nunca

**Pontuação do Questionário de Berlim:**

**Categoria 1:** itens 1, 2, 3, 4 e 5
- Item 1 se a resposta foi sim: 1 ponto
- Item 2 se a resposta foi c ou d: 1 ponto
- Item 3 se a resposta foi a ou b: 1 ponto
- Item 4 se a resposta foi a: 1 ponto
- Item 5 se a resposta foi a ou b: 2 pontos

**Categoria 2:** itens 6, 7 e 8 (item 9 deve ser considerado separadamente)
- Item 6 se a resposta foi a ou b: 1 ponto
- Item 7 se a resposta foi a ou b: 1 ponto
- Item 8 se a resposta foi a: 1 ponto

**Pontuação:**
- Categoría 1 Ø positiva se a pontuação Ø maior ou igual a 2 pontos
- Categoría 2 Ø positiva se a pontuação Ø maior ou igual a 2 pontos
- Categoría 3 Ø positiva se a resposta ao item 10 Ø sim ou se o índice de massa corporal (IMC) do doente Ø superior a 30 kg/m²

Doente de alto risco para SAOS: duas ou mais categorias com pontuação positiva
Doente de baixo risco para SAOS: nenhuma ou apenas uma categoria com pontuação positiva