Evaluation of Berlin Questionnaire Validity for Sleep Apnea Risk in Sleep Clinic Populations

Behnam Khaledi-Paveh 1, Habibolah Khazaie 1, Marzie Nasouri 1, Mohammad Rasoul Ghadami 1, Masoud Tahmasian 1,2*

1. Sleep Disorders Research Center, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran. 2. Department of Neurology, University Hospital of Cologne, Cologne, Germany.

ABSTRACT

Introduction: The Berlin questionnaire (BQ) is a common tool to screen for Obstructive Sleep Apnea (OSA) in the general population, but its application in the clinical sleep setting is still challenging. The aim of this study was to determine the specificity and sensitivity of the BQ compared to the apnea-hypopnea index obtained from polysomnography recordings obtained from a sleep clinic in Iran.

Methods: We recruited 100 patients who were referred to the Sleep Disorders Research Center of Kermanshah University of Medical Sciences for the evaluation of suspected sleep-disorder breathing difficulties. Patients completed a Persian version of BQ and underwent one night of PSG. For each patient, Apnea-Hypopnea Index (AHI) was calculated to assess the diagnosis and severity of OSA. Severity of OSA was categorized as mild when AHI was between 5 and 15, moderate when it was between 15 and 30, and severe when it was more than 30.

Results: BQ results categorized 65% of our patients as high risk and 35% as low risk for OSA. The sensitivity and the specificity of BQ for OSA diagnosis with AHI>5 were 65% and 36%, respectively. Positive predictive value was 78% and negative predictive value was 23%. Moreover, the area under curve was 0.53 (95% CI: 0.49 – 0.67, P=0.38).

Discussion: Our findings suggested that BQ, despite its advantages in the general population, is not a precise tool to determine the risk of sleep apnea in the clinical setting, particularly in the sleep clinic population.

Key Words: Sleep apnea, Obstructive, Berlin questionnaire, Sleep clinic, Validation

1. Introduction

Obstructive Sleep Apnea (OSA) is a common sleep-breathing disorder in primary health care with adverse outcomes in health and quality of life in human population (Ahmadi, Chung, Gibbs, & Shapiro, 2008; Khazaie et al., 2011). OSA is associated with severe medical diseases, including coronary artery disease, cardiopulmonary hypertension, cardiac arrhythmia, congestive heart failure, stroke, type 2 diabetes and it is also correlated with a higher risk for motor vehicle accidents (Ahmadi et al., 2008; Bentivoglio et al., 2008; Khazaie et al., 2011; Shaw et al., 2008). During the last decades, several studies evaluated the validity and reliability of sleep questionnaires as a cost-benefit screening tool to estimate the rate/prevalence of OSA in developed and developing countries (Crocker et al., 1990; Dealberto, Ferber, Garma, Lemoine, & Alperovitch, 1994; Deegan & McNicholas, 1996). The Berlin questionnaire (BQ), developed in 1999 by Netzer and colleagues in United States, is one of the most popular questionnaires for screening OSA, (Netzer, Strohs, Netzer, Clark, & Strohl, 1999). However, there is an obvious need for translate and validate the BQ in different languages.
Several studies translated and evaluated the validity of the BQ in different languages and demonstrated that it is a valid and useful questionnaire to screen patients with OSA in different populations (Amra, Nouranian, Golshan, Fietze, & Penzel, 2013; Asghari, Mohammadi, Kamrava, Jalessi, & Farhadi, 2013; Asghari, Mohammadi, Kamrava, Tavakoli, & Farhadi, 2012; Bouloukaki et al., 2013; Jazi et al., 2014; Kang et al., 2013; Khassawneh et al., 2009; Saleh, Ahmad, & Awadalla, 2011). Although BQ seems to be a valid tool to detect OSA in the general population, in the clinical sleep setting the literature is inconsistent. Supporting the use of the BQ, some studies reported that BQ can serve as a valid and useful screening method to detect OSA in the clinical sleep setting (Amra et al., 2013; Dealberto et al., 1994) and in general populations (Kang et al., 2013; Saleh et al., 2011).

On the other hand, Ahmadi and colleagues compared the BQ to the Respiratory Disturbance Index (RDI) values measured by polysomnography (PSG). Their findings provided evidence that BQ is not a valid tool for identifying patients with OSA in sleep clinic populations due to its low sensitivity and specificity with large numbers of false negatives and positives (Ahmadi et al., 2008). Although Ahmadi and colleagues used RDI to validate the BQ in a sleep clinic population, many clinical and epidemiological studies use Apnea-Hypopnea Index (AHI) to define OSA (Lurie, 2011). In addition, there is a divergent risk of OSA in different regions of Iran. For example, there is a relatively low prevalence (5%) of sleep apnea-related symptoms in Isfahan Province, Iran (Amra, Farajzadegan, Golshan, Fietze, & Penzel, 2011) but in our previous epidemiologic study, we found a noticeable high prevalence (27.3%) risk of OSA in Kermanshah Province in western of Iran (Khazaie et al., 2011).

Thus, it seems that the evaluation of OSA in different samples may provide different results due to genetic or other biological or ecological factors (Kaparianos, Sampsonas, Karkoulias, & Spiropoulos, 2006). The question of this study was whether the Persian version of the BQ is a valid tool to assess Iranian clinical sleep patients who have high risk of OSA. Hence, our aim was to determine the specificity and sensitivity of BQ compared to the AHI values obtained from PSG by recording in the clinical sleep setting in Kermanshah, Iran.

2. Methods

2.1. Subjects

The study population consisted of the patients referred to the Sleep Disorders Research Center, Farabi Hospital, affiliated with the Kermanshah University of Medical Sciences (KUMS) for evaluation of sleep problems from September 2012 until August 2014. Exclusion criteria were as follows: having current or a history of any neurological or psychiatric disorders; having an upper respiratory infection, or cardiovascular, hepatic, or renal diseases during the last one month; having a history of alcohol or drug abuse; and being pregnant. After receiving the approval from the Medical Ethics Board of KUMS, a written informed consent form was obtained from all patients. Additionally, demographic data, including age, gender, weight, and height were collected from all patients.

2.2. Berlin questionnaire

BQ is a simple and useful tool for screening OSA risk in the general population. Netzer and colleagues assessed efficacy of BQ in the primary care setting on 744 subjects with portable sleep monitoring (Netzer et al., 1999). BQ has 3 sections. In the first section, the participants are asked to score their snoring. In the second section, daytime fatigue and sleepiness during daily activities are investigated, and in the last section, medical history, demographic, and anthropometric measures such as height and weight, can be evaluated. The first two sections are assumed to be positive, if the total score is 2 or more. If a patient has hypertension or a BMI higher than 30kg/m², section 3 can be considered positive. In general, if there are two or more sections with positive scores, this subject is categorized as “high risk” for OSA. Previously, Amra and colleagues translated the BQ into Persian (Farsi) and back into English and validated it in general population of Iran. The reliability of BQ categories showed a Cronbach α of 0.70 for section 1 and 0.50 for section 2.

<table>
<thead>
<tr>
<th>Table 1. Demographic characteristics of the study population (n=100).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low risk (n=35)</strong></td>
</tr>
<tr>
<td>Age, year</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
</tr>
<tr>
<td>Obesity (BMI&gt;30) (n, %)</td>
</tr>
</tbody>
</table>

Data are presented as n(%) or Mean±SD.
for section 2 (Amra et al., 2011; Amra et al., 2013). In the present study, we have used the Persian version of the BQ, which was suggested by them.

2.3. Polysomnography

All patients who filled out the BQ underwent one night of PSG (Somnomedics GmbH, Germany). Measurement of PSG was based on the American Academy of Sleep Medicine guidelines and standard techniques, as well as monitoring of the electroencephalogram using frontal, central, and occipital leads; electrooculogram; electromyogram, flow (by oronasal thermistor and nasal air pressure transducer), thoracic and abdominal respiratory effort (induction plethysmography), oximetry, and body position (Grigg-Damberger, 2012; Morgenthaler et al., 2007). Sleep parameters, including AHI per hour, were extracted for diagnosis and severity of OSA. Severity of OSA was categorized as “none” when AHI was less than 5 per hour, “mild” when AHI was 5-15/h, “moderate” when AHI was 15-30/h, and “severe” when AHI was more than 30/h.

Statistical analysis: Statistical analyses were performed using SPSS (version 20.0) with a significance threshold of P<0.05. Data is presented as numbers and percentages as well as means and standard deviation (SD). Data was compared using a t-test or Chi-square analysis. Sensitivity, specificity, positive and negative predictive values and the AUC for each cut-off value of the gold standard diagnosis were calculated using standard methods by comparing the AHI results.

<table>
<thead>
<tr>
<th></th>
<th>Low risk (n=35)</th>
<th>High risk (n=65)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHI (events per hour %)</td>
<td>17.6±15.5</td>
<td>24±21.5</td>
<td>0.122</td>
</tr>
<tr>
<td>AHI&lt;5</td>
<td>8 (22)</td>
<td>14 (21)</td>
<td>0.553</td>
</tr>
<tr>
<td>5&lt;AHI&lt;15</td>
<td>11 (31)</td>
<td>13 (20)</td>
<td></td>
</tr>
<tr>
<td>15&lt;AHI&lt;30</td>
<td>9 (25)</td>
<td>18 (27)</td>
<td></td>
</tr>
<tr>
<td>AHI&gt;30</td>
<td>7 (20)</td>
<td>20 (30)</td>
<td></td>
</tr>
</tbody>
</table>

Data are presented as n(%) or Mean±SD.

3. Results

3.1. Demographic information

One hundred participants (60% male) with the age range of 11-82 years were recruited for this study. Their average BMI was 29.5±6.1 kg/m² (Table 1).

3.2. Polysomnography

In our subjects, the prevalence of AHI>5 was 24%, AHI>10 was 27%, and AHI>15 was 27%. Details of AHI data obtained from the overnight polysomnographic studies in subjects with low and high risk OSA are shown in Table 2.

3.3. The association between AHI and Berlin questionnaire

Seventy-two percent of participants had positive scores in the first category, 58% had positive scores in the second category, and 47% showed positive scores in the third category (Table 3). The BQ identified 65% of our patients as having a high risk of OSA and 35% as low risk. Our findings showed that BQ has moderate sensitivity and low specificity at detecting the risk of OSA. Table 2 shows the number of individuals with positive (high risk of OSA) and negative (low risk of OSA) BQ scores in each AHI range. The sensitivity and the specificity of the BQ for OSA diagnosis with AHI>5 were 65% and 36%, respectively, with a positive predictive value (PPV) of 78% and a negative predictive value (NPV) of 23%. The area under curve (AUC) was only 0.53 (95% CI: 0.49–0.67, P=0.38)

Table 3. Berlin Questionnaire values (numbers of patients who were positive in each category).

<table>
<thead>
<tr>
<th></th>
<th>Low risk (n=35)</th>
<th>High risk (n=65)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1 (%)</td>
<td>11 (31)</td>
<td>61 (94)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Category 2 (%)</td>
<td>12 (34)</td>
<td>46 (70)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Category 3 (%)</td>
<td>5 (14)</td>
<td>42 (64)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Data are presented as n(%).
(Figure 1). With subjects scoring AHI>15, the sensitivity and the specificity of the BQ were 70% and 41%, respectively. The BQ sensitivity and specificity were 74% and 38% for AHI>30.

4. Discussion

OSA is a common sleep disorder, which is characterized by repetitive complete (apnea) and/or partial collapses (hypopnea) of the upper airways. Clinically, OSA is defined as the occurrences of at least 5 episodes of apnea or hypopnea per hour (AHI) in association with symptoms attributed to sleep disordered breathing. OSA is classified as mild, moderate, and severe when assessed by PSG (Ruehland et al., 2009). Although PSG is a gold standard for OSA diagnosis, it is a time-consuming and expensive procedure. The increased awareness of sleep apnea, intensifies the long waiting list in many sleep clinics. Thus, there is a growing interest in developing screening tools for OSA (Bouloukaki et al., 2013; Flemons, Douglas, Kuna, Rodenstein, & Wheatley, 2004; Kang et al., 2013; Saleh et al., 2011). Our study demonstrated that BQ had moderate to low sensitivity, specificity, PPV, and NPV in the sleep clinic population. These findings support the results of Ahmadi and colleagues, who reported sensitivity and specificity of 68% and 49% for RDI>5, 62% and 43% for RDI>10, and 57% and 41% for RDI>15 (Ahmadi et al., 2008). Although they used RDI data from PSG, we also demonstrated low sensitivity and specificity of the BQ using AHI data. These findings confirmed that BQ is not a valid tool to identify patients with sleep apnea particularly in the sleep clinic population.

In our study, the sensitivity and specificity are significantly different between distinct AHI groups (65% and 36% for AHI>5, 70% and 41% for AHI>15, and 74% and 38% for AHI>30, respectively). Several studies have shown that BQ is not a promising screening tool to detect sleep apnea in clinical setting, showing a similar or lower sensitivity and specificity compared to our results (Vaz et al., 2011; Weinreich, Plein, Teschler, Resler, & Teschler, 2006). The validity of BQ in our population is in agreement with the results of other studies in a sleep laboratory setting in Portugal (72% and 50% for AHI≥5–<15, 82% and 45% for AHI≥15–≤30, and 88% and 39% for AHI>30) (Vaz et al., 2011). However, our findings support the previous studies. For instance, BQ has been shown to detect sleep apnea in patients referred to a sleep laboratory with a lower sensitivity and specificity in subjects undergoing pulmonary rehabilitation (sensitivity and specificity of 68% and 49% for RDI>5, 62% and 43% for RDI>10, and 57% and 41% for RDI>15 (Ahmadi et al., 2008). Although they used RDI data from PSG, we also demonstrated low sensitivity and specificity of the BQ using AHI data. These findings confirmed that BQ is not a valid tool to identify patients with sleep apnea particularly in the sleep clinic population.

In our study, the sensitivity and specificity are significantly different between distinct AHI groups (65% and 36% for AHI>5, 70% and 41% for AHI>15, and 74% and 38% for AHI>30, respectively). Several studies have shown that BQ is not a promising screening tool to detect sleep apnea in clinical setting, showing a similar or lower sensitivity and specificity compared to our results (Vaz et al., 2011; Weinreich, Plein, Teschler, Resler, & Teschler, 2006). The validity of BQ in our population is in agreement with the results of other studies in a sleep laboratory setting in Portugal (72% and 50% for AHI≥5–<15, 82% and 45% for AHI≥15–≤30, and 88% and 39% for AHI>30) (Vaz et al., 2011). However, our findings support the previous studies. For instance, BQ has been shown to detect sleep apnea in patients referred to a sleep laboratory with a lower sensitivity and specificity in subjects undergoing pulmonary rehabilitation (sensitivity and specificity of 68% and 49% for RDI>5, 62% and 43% for RDI>10, and 57% and 41% for RDI>15 (Ahmadi et al., 2008). Although they used RDI data from PSG, we also demonstrated low sensitivity and specificity of the BQ using AHI data. These findings confirmed that BQ is not a valid tool to identify patients with sleep apnea particularly in the sleep clinic population.

In our study, the sensitivity and specificity are significantly different between distinct AHI groups (65% and 36% for AHI>5, 70% and 41% for AHI>15, and 74% and 38% for AHI>30, respectively). Several studies have shown that BQ is not a promising screening tool to detect sleep apnea in clinical setting, showing a similar or lower sensitivity and specificity compared to our results (Vaz et al., 2011; Weinreich, Plein, Teschler, Resler, & Teschler, 2006). The validity of BQ in our population is in agreement with the results of other studies in a sleep laboratory setting in Portugal (72% and 50% for AHI≥5–<15, 82% and 45% for AHI≥15–≤30, and 88% and 39% for AHI>30) (Vaz et al., 2011). However, our findings support the previous studies. For instance, BQ has been shown to detect sleep apnea in patients referred to a sleep laboratory with a lower sensitivity and specificity in subjects undergoing pulmonary rehabilitation (sensitivity and specificity of 68% and 49% for RDI>5, 62% and 43% for RDI>10, and 57% and 41% for RDI>15 (Ahmadi et al., 2008). Although they used RDI data from PSG, we also demonstrated low sensitivity and specificity of the BQ using AHI data. These findings confirmed that BQ is not a valid tool to identify patients with sleep apnea particularly in the sleep clinic population.

In our study, the sensitivity and specificity are significantly different between distinct AHI groups (65% and 36% for AHI>5, 70% and 41% for AHI>15, and 74% and 38% for AHI>30, respectively). Several studies have shown that BQ is not a promising screening tool to detect sleep apnea in clinical setting, showing a similar or lower sensitivity and specificity compared to our results (Vaz et al., 2011; Weinreich, Plein, Teschler, Resler, & Teschler, 2006). The validity of BQ in our population is in agreement with the results of other studies in a sleep laboratory setting in Portugal (72% and 50% for AHI≥5–<15, 82% and 45% for AHI≥15–≤30, and 88% and 39% for AHI>30) (Vaz et al., 2011). However, our findings support the previous studies. For instance, BQ has been shown to detect sleep apnea in patients referred to a sleep laboratory with a lower sensitivity and specificity in subjects undergoing pulmonary rehabilitation (sensitivity and specificity of 68% and 49% for RDI>5, 62% and 43% for RDI>10, and 57% and 41% for RDI>15 (Ahmadi et al., 2008). Although they used RDI data from PSG, we also demonstrated low sensitivity and specificity of the BQ using AHI data. These findings confirmed that BQ is not a valid tool to identify patients with sleep apnea particularly in the sleep clinic population.

In our study, the sensitivity and specificity are significantly different between distinct AHI groups (65% and 36% for AHI>5, 70% and 41% for AHI>15, and 74% and 38% for AHI>30, respectively). Several studies have shown that BQ is not a promising screening tool to detect sleep apnea in clinical setting, showing a similar or lower sensitivity and specificity compared to our results (Vaz et al., 2011; Weinreich, Plein, Teschler, Resler, & Teschler, 2006). The validity of BQ in our population is in agreement with the results of other studies in a sleep laboratory setting in Portugal (72% and 50% for AHI≥5–<15, 82% and 45% for AHI≥15–≤30, and 88% and 39% for AHI>30) (Vaz et al., 2011). However, our findings support the previous studies. For instance, BQ has been shown to detect sleep apnea in patients referred to a sleep laboratory with a lower sensitivity and specificity in subjects undergoing pulmonary rehabilitation (sensitivity and specificity of 68% and 49% for RDI>5, 62% and 43% for RDI>10, and 57% and 41% for RDI>15 (Ahmadi et al., 2008). Although they used RDI data from PSG, we also demonstrated low sensitivity and specificity of the BQ using AHI data. These findings confirmed that BQ is not a valid tool to identify patients with sleep apnea particularly in the sleep clinic population.

In our study, the sensitivity and specificity are significantly different between distinct AHI groups (65% and 36% for AHI>5, 70% and 41% for AHI>15, and 74% and 38% for AHI>30, respectively). Several studies have shown that BQ is not a promising screening tool to detect sleep apnea in clinical setting, showing a similar or lower sensitivity and specificity compared to our results (Vaz et al., 2011; Weinreich, Plein, Teschler, Resler, & Teschler, 2006). The validity of BQ in our population is in agreement with the results of other studies in a sleep laboratory setting in Portugal (72% and 50% for AHI≥5–<15, 82% and 45% for AHI≥15–≤30, and 88% and 39% for AHI>30) (Vaz et al., 2011). However, our findings support the previous studies. For instance, BQ has been shown to detect sleep apnea in patients referred to a sleep laboratory with a lower sensitivity and specificity in subjects undergoing pulmonary rehabilitation (sensitivity and specificity of 68% and 49% for RDI>5, 62% and 43% for RDI>10, and 57% and 41% for RDI>15 (Ahmadi et al., 2008). Although they used RDI data from PSG, we also demonstrated low sensitivity and specificity of the BQ using AHI data. These findings confirmed that BQ is not a valid tool to identify patients with sleep apnea particularly in the sleep clinic population.
and specificity 62.5% and 53.8% with a cut-off for AHI≤10) (Weinreich et al., 2006).

Moreover, Ulasli and colleagues tried to determine the predictive accuracy of BQ and the Epworth Sleepiness Scale in patients hospitalized in a sleep clinic. Their findings demonstrated that BQ predicted an AHI≥5 with a sensitivity of 0.73 and a specificity of 0.44 and a sensitivity of 0.80 with a cut-off value of AHI≥30. They concluded that BQ is a poor predictor of OSA in subjects admitted to the sleep clinic (Ulasli et al., 2014). Recently, Karakoc and colleagues assessed 217 subjects who were at risk for OSA using BQ. Their findings demonstrated that 82.03% of their patients were at high risk of sleep apnea and 17.9% were at low risk. The sensitivity and specificity were 83.4% and 22.2% and the PPV value and NPV were 76.4% and 30.8%, respectively, using a cut-off point of AHI>5. They also showed that the sensitivity and specificity were 89.3% and 22.6% and the PPV value and NPV were 42.1% and 76.9%, respectively, when a cut-off point of AHI>15 was used. The findings of the present study are in the same line with their study, suggesting that BQ is not an optimal predictor of OSA in patients at high risk of OSA (Karakoc et al., 2014).

The inconsistency in the studies on Iranian populations might be due to the several reasons. First of all, the prevalence of OSA in our sample is rather high. Previously, we reported that approximately 27.3% of individuals from the general population in Kermanshah Province in western Iran have symptoms and risk factors for OSA (Khazaie et al., 2011). However, the prevalence of sleep apnea symptoms is lower (5%) in Isfahan Province in the middle of Iran compared to our province (Amra et al., 2011). The prevalence of being high-risk for OSA in Kermanshah is comparable to the studies which have been performed in the USA (26%) (Hiestand, Britz, Goldman, & Phillips, 2006) and Saudi Arabia (33.3%) (Bahammam, Alrajeh, Al-Jahdali, & BinSaeed, 2008). Secondly, as there is currently only one sleep clinic in the western part of Iran, other regional medical specialists (e.g. pulmonologists, otolaryngologist, internists, neurologist, and psychiatrists) refer their patients to this center for the evaluation of sleep problems. Thus, our case selection might have a bias due to various sleep disorders that patients have.

Taken together, in the geographic region, which has a high prevalence of OSA in general population, it is important for primary care medical practitioners to have a practical and valid screening tool to identify patients at high risk of OSA. In spite of the advantages of BQ in the general population, utilization of the BQ as a screening tool for OSA should be considered with caution in the sleep clinic populations.

Our findings suggest that although BQ is a valid tool in general population, it is not an appropriate instrument for identifying patients with OSA in sleep clinic population in regions with high risk of OSA.

Acknowledgments

We would like to thank all patients and their relatives for their participation in the study. Moreover, we thank the staff of Sleep Disorders Research Center of Kermanshah University of Medical Sciences, especially Mr. Salimi for his kind cooperation.

Conflict of interest

The authors declared that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors’ Contribution

All authors are fully responsible for the study and declare to have no conflict-of-interest. Behnam Khaledi-Paveh, Masoud Tahmasian, and Habibolah Khazaie designed the study; Behnam Khaledi-Paveh and Marzie Nasouri recruited subjects and conducted the experiment; Behnam Khaledi-Paveh and Mohammad Rasoul Ghadami analyzed the data; Masoud Tahmasian and Habibolah Khazaie supervised the experiment; Behnam Khaledi-Paveh, Mohammad Rasoul Ghadami, Habibolah Khazaie and Masoud Tahmasian wrote the manuscript.

References


