Cloud algorithm-driven oximetry-based diagnosis of obstructive sleep apnoea in symptomatic habitually snoring children

Zhifei Xu¹, Gonzalo C. Gutiérrez-Tobal², Yunxiao Wu³, Leila Kheirandish-Gozal⁴, Xin Ni³,⁵, Roberto Hornero² and David Gozal⁴,⁵

Affiliations: ¹Respiratory Dept, Beijing Children’s Hospital, Capital Medical University, National Center for Children’s Health, Beijing, People’s Republic of China. ²Biomedical Engineering Group, Universidad de Valladolid, Valladolid, Spain. ³Otolaryngology, Head and Neck Surgery Dept, Beijing Children’s Hospital, Capital Medical University, National Center for Children’s Health, Beijing, People’s Republic of China. ⁴Dept of Child Health, University of Missouri School of Medicine, Columbia, MO, USA. ⁵David Gozal and Xin Ni contributed equally to this article as lead authors and supervised the work.

Correspondence: David Gozal, Dept of Child Health, University of Missouri School of Medicine, 400 N. Keene Street, Suite 010, Columbia, MO 65201, USA. E-mail: gozald@health.missouri.edu

ABSTRACT The ability of a cloud-driven Bluetooth oximetry-based algorithm to diagnose obstructive sleep apnoea syndrome (OSAS) was examined in habitually snoring children concurrently undergoing overnight polysomnography.

Children clinically referred for overnight in-laboratory polysomnographic evaluation for suspected OSAS were simultaneously hooked to a Bluetooth oximeter linked to a smartphone. Polysomnography findings were scored and the apnoea/hypopnoea index (AHIPSG) was tabulated, while oximetry data yielded an estimated AHIOXI using a validated algorithm.

The accuracy of the oximeter in identifying correctly patients with OSAS in general, or with mild (AHI 1–5 events·h⁻¹), moderate (5–10 events·h⁻¹) or severe (>10 events·h⁻¹) OSAS was examined in 432 subjects (6.5±3.2 years), with 343 having AHIPSG >1 event·h⁻¹. The accuracies of AHIOXI were consistently >79% for all levels of OSAS severity, and specificity was particularly favourable for AHI >10 events·h⁻¹ (92.7%). Using the criterion of AHIPSG >1 event·h⁻¹, only 4.7% of false-negative cases emerged, from which only 0.6% of cases showed moderate or severe OSAS.

Overnight oximetry processed via Bluetooth technology by a cloud-based machine learning-derived algorithm can reliably diagnose OSAS in children with clinical symptoms suggestive of the disease. This approach provides virtually limitless scalability and should alleviate the substantial difficulties in accessing paediatric sleep laboratories while markedly reducing the costs of OSAS diagnosis.

Published in volume 53, issue 2 of the European Respiratory Journal on 21 Feb 2019; republished 7 May 2021 with amendments to the author list footnote indicating equal contribution of the lead authors.

Received: Sept 20 2018 | Accepted after revision: Nov 13 2018

Copyright ©ERS 2019
Introduction

Obstructive sleep apnoea/hypopnoea syndrome (OSAS) has emerged in recent decades as a highly prevalent disease in children all over the world, and is estimated to affect 2–5% of all children. However, the cardinal symptom of OSAS in children is habitual snoring, which affects a much higher proportion of children (range 6–25%). Furthermore, paediatric OSAS has been consistently associated with increased risk of major end-organ adverse consequences affecting neurocognitive, behavioural, cardiovascular and metabolic systems, ultimately resulting in overall declines in health and quality of life, as well as increased healthcare costs [1–4]. Based on current guidelines, nocturnal polysomnography (PSG) in an accredited sleep laboratory is considered the gold-standard approach to diagnose OSAS in children [1, 5, 6]. However, the scarcity of paediatric sleep laboratories around the world, the elevated costs of PSG and their labour-intensive nonscalable characteristics, and the obvious inconvenience to parents and children have led to the unfortunate reality that only a minute proportion of symptomatic habitually snoring children are evaluated objectively before undergoing adenotonsillectomy, the first line of therapy [7, 8].

In an effort to overcome these problems and expand the accessibility and objectivity of OSAS diagnosis, many alternative methodologies have been developed, ranging from questionnaires to simplified multichannel studies, and even exploration of diagnostic biomarker panels [9–13]. Nocturnal oximetry was proposed initially as a screening tool for OSAS in symptomatic children [11, 13], and this approach has gained increasing popularity despite exhibiting favourable specificity yet limited sensitivity, while also being marred by interscorer reliability issues, particularly at the low end of OSAS severity, as well as scalability concerns [11]. To overcome these issues, several investigative groups including ours have proposed a variety of automated procedures that circumvent the subjectivity of oximetry recording interpretation [14–28]. In this context, we have reported on the application of machine-learning procedures in the analysis of nocturnal oximetry recordings among children referred for clinical evaluation of suspected OSAS, and the derivation and validation of a diagnostic algorithm in a very large cohort of >4000 children [29]. Here, we furthered our quest for a scorer-independent scalable diagnostic approach of paediatric OSAS by exploring and comparing the diagnostic performance of a Bluetooth-enabled oximeter coupled to a smartphone for data transmission and derivation of the estimated apnoea/hypopnoea using a cloud-based algorithm when tested concurrent with a PSG study in the laboratory.

Patients and methods

Subjects

Consecutive, otherwise healthy, habitually snoring symptomatic children (≥3 nights per week) aged 2–15 years who were referred for suspected OSAS were recruited from Beijing Children’s Hospital, Capital Medical University (Beijing, China) between June 1, 2017 and June 1, 2018. All participants underwent an overnight PSG evaluation while concurrently wearing a commercially and readily available Bluetooth oximeter linked to an Android smartphone via a custom designed application (Serenium, Palo Alto, CA, USA). Written informed consent and assent were obtained from parents and children, respectively (for children aged >7 years). The study was approved by the ethics committee of Beijing Children’s Hospital affiliated to Capital Medical University (protocol #2017-151), and received approval for processing of the de-identified oximetry recordings from the University of Chicago human subject committee (protocol #IRB14-1241).

Exclusion criteria

Children who were known to be suffering from congenital heart disease, systemic or pulmonary hypertension, diabetes mellitus or dyslipidaemia, those with craniofacial anomalies, neuromuscular disease or defined genetic syndromes were excluded. In addition, children with any known acute or chronic illness, or who received previous treatment for OSAS were excluded. Children with recordings from either PSG or Bluetooth oximeter lasting <3 h were excluded.

Anthropometry

All children were weighed on a calibrated scale and their weights were recorded to the nearest 0.1 kg. Height (to 0.1 cm) was measured using a stadiometer. The body mass index (BMI) and BMI z-score were calculated using Chinese normative datasets [30]. The definition of obesity in our study was BMI z-score ≥1.65.

Polysomnography

Children were monitored during the PSG using a digital acquisition system (Compumedics E; Compumedics, Melbourne, Australia or ALICE 5; Philips Respironics, Amsterdam, the Netherlands). No coffee, tea, cola-containing products or sedative hypnotics were taken before sleep. Total sleep time was >7.5 h. PSG monitoring included the following parameters: electroencephalogram from four leads (C3/A2, C4/A1, O1/A2, O2/A3), bilateral electro-oculogram, electromyogram of mentalis activity and bilateral anterior tibialis, chest and abdominal movements, ECG, arterial oxyhaemoglobin saturation and plethysmographic signal by pulse.
oximetry, air flow thermistor and nasal pressure cannula, snoring sensor and body position. Sleep data were scored manually by experienced paediatric PSG technicians according to the scoring manual published by the American Academy of Sleep Medicine (AASM) [31]. Oxygen desaturation index (ODI3%) was defined as the number of ≥3% arterial oxygen desaturations per hour of sleep. The definition of arousal was based on the AASM guidelines. The diagnosis of children with OSAS was defined by the presence of an obstructive apnoea/hypopnoea index (AHIPSG) ≥1 event·h⁻¹ of total sleep time according to the most frequent clinical practice, as described in the 2012 American Academy of Pediatrics consensus guideline for the diagnosis and management of childhood OSAS [1]. Primary snoring was defined as AHIPSG <1 event·h⁻¹. Mild OSAS was defined as AHIPSG ≥1 event·h⁻¹ and <5 events·h⁻¹, moderate OSAS was defined as AHIPSG ≥5 events·h⁻¹ and severe OSAS was defined as AHIPSG ≥10 events·h⁻¹.

**Oximetry data processing**

Oximetry signals from the Bluetooth oximeter were transferred via the smartphone to the cloud and were all rounded to the second decimal place. Artefacts were then automatically removed according to the method proposed by MAGALANG et al. [32]. Signals were automatically processed using the algorithm previously developed and validated [29], which consisted of a multilayer perceptron (MLP) model with the ability to estimate AHI automatically. MLP constitutes an artificial neural network that is typically arranged in three layers of mathematical units called neurons: input, hidden and output [29], and Matlab R2016b (MathWorks, Cambridge, UK) was used to implement feature extraction and classification stages. Accordingly, an estimate of the AHI was computed (AHIOXI), and compared with AHIPSG.

**Statistical analyses**

SPSS Statistics software (version 20; IBM, Chicago, IL, USA) was used, and data are presented as mean±SD. Intraclass correlation coefficient (ICC) was used to directly assess the agreement between the AHIPSG and AHIOXI, as well as Bland–Altman and Cohen’s κ [33]. In addition, the diagnostic performance for three cut-offs (AHIPSG <1 event·h⁻¹, 5 events·h⁻¹ and 10 events·h⁻¹) was assessed by means of sensitivity, specificity, positive likelihood ratio (LR⁺), negative likelihood ratio (LR⁻) and accuracy. Plotting of the LR values was performed using a freely available web-based calculator developed by Alan Schwartz (http://araw.medc.uic.edu/cgi-bin/testcalc.pl?DT=&Dt=&dT=&dt=&2×2=Compute). For comparisons of continuous variables across clinical groupings, Kruskal–Wallis (nonparametric) or Mann–Whitney tests were used as appropriate. For comparisons of discrete variables, a Chi-squared test was used. A p-value <0.05 was considered to be indicative of statistical significance.

**Results**

432 children completed the study out of 435 who were approached and agreed to participate. The reasons for inability to complete the study was related to intolerance of the PSG equipment by the three very young children (ages 2 years, 2.5 years and 3 years), which led to them spending most of the recording time awake or without an appropriate PSG recording montage. Table 1 provides the demographic and anthropometric characteristics of the cohort as well as their PSG findings.

### TABLE 1 Demographic, anthropometric and polysomnographic characteristics of 432 symptomatic habitually snoring Chinese children undergoing overnight polysomnography and concurrent Bluetooth oximetry for suspected obstructive sleep apnoea syndrome (OSAS)

<table>
<thead>
<tr>
<th></th>
<th>All participants</th>
<th>Primary snoring</th>
<th>OSAS AHIPSG ≥1 event·h⁻¹</th>
<th>OSAS AHIPSG &gt;5 events·h⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>432</td>
<td>89</td>
<td>343</td>
<td>171</td>
</tr>
<tr>
<td>Age years</td>
<td>6.3±2.5</td>
<td>6.5±2.4</td>
<td>6.3±2.6</td>
<td>6.4±2.5</td>
</tr>
<tr>
<td>Male %</td>
<td>65.3</td>
<td>62.8</td>
<td>69.4</td>
<td>66.5%</td>
</tr>
<tr>
<td>BMI (% obese)</td>
<td>17.8±4.5 (26.3)</td>
<td>16.6±3.9 (23.2)</td>
<td>18.3±4.7 (26.7)</td>
<td>19.2±5.4 (33.1)**</td>
</tr>
<tr>
<td>Total sleep time min</td>
<td>474±154.4</td>
<td>460.4±72.2</td>
<td>478.1±47.8</td>
<td>471.7±48.4</td>
</tr>
<tr>
<td>Sleep efficiency %</td>
<td>83.5±8.6</td>
<td>83.7±8.9</td>
<td>83.5±8.6</td>
<td>82.2±8.8</td>
</tr>
<tr>
<td>AHIPSG events·h⁻¹ (median; IQR)</td>
<td>10.0±21.3 (3; 8.1)</td>
<td>0.5±0.3 (0.5; 0.5)</td>
<td>11.4±23.3 (4.5; 9.6)</td>
<td>22.3±29.6 [12.2; 16.4]#</td>
</tr>
<tr>
<td>ODI3% events·h⁻¹</td>
<td>6.7±16.2</td>
<td>0.2±0.7</td>
<td>8.3±17.7</td>
<td>14.8±21.4*</td>
</tr>
<tr>
<td>SpO₂ nadir</td>
<td>89.8±7.2</td>
<td>94.3±2.0</td>
<td>88.6±7.6</td>
<td>85.2±9.0*</td>
</tr>
</tbody>
</table>

Data are presented as n or mean±SD, unless otherwise stated. AHIPSG: apnoea/hypopnoea index measured using polysomnography; BMI: body mass index; AH: apnoea/hypopnoea index; IQR: interquartile range; ODI3%; oxygen desaturation index (the number of ≥3% arterial oxygen desaturations per hour of sleep); SpO₂: arterial oxygen saturation measured by pulse oximetry. **: p<0.01 versus all others; #: p<0.0001 versus primary snoring.
Figure 1 displays the Bland–Altman plot comparing the AHIPSG of the subjects with their corresponding AHIOXI estimation from the oximeter data concurrently acquired during their PSG testing. In addition, a low mean positive difference (slight AHI underestimation by the algorithm) is apparent, with 95% confidence intervals within (−40.5–35.3), which reflects the dispersion that occurs when AHIPSG values are very high, indicative of extremely severe OSAS. In addition, a high ICC is reached (0.317).

Table 2 and figure 2 show the confusion matrix comparing the classification derived from the AHIPSG with the classification achieved by the cloud-based algorithm based on oximetry alone, i.e. AHIOXI. Accordingly, Cohen’s κ was 0.339. In addition, table 2 displays sensitivity, specificity, LR+ and LR− for the AHI 1 event·h−1, 5 events·h−1 and 10 events·h−1 cut-offs, derived from the confusion matrix, and figure 3.

Table 2. Confusion matrix showing the classification agreement of Bluetooth oximeter cloud-based algorithm-calculated apnoea/hypopnoea index (AHI) estimate and the nocturnal polysomnography (PSG)-derived AHI (AHIPSG) in 432 symptomatic habitually snoring Chinese children undergoing overnight polysomnography and concurrent Bluetooth oximetry (OXI) for suspected obstructive sleep apnoea.

<table>
<thead>
<tr>
<th>AHIOXI</th>
<th>&lt;1 event·h−1</th>
<th>1–5 events·h−1</th>
<th>5–10 events·h−1</th>
<th>&gt;10 events·h−1</th>
<th>Nocturnal PSG</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>LR+</th>
<th>LR−</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1 event·h−1</td>
<td>17</td>
<td>66</td>
<td>4</td>
<td>2</td>
<td>89</td>
<td>95.3</td>
<td>19.1</td>
<td>79.6</td>
<td>1.18</td>
<td></td>
</tr>
<tr>
<td>1–5 events·h−1</td>
<td>14</td>
<td>113</td>
<td>41</td>
<td>4</td>
<td>172</td>
<td>77.8</td>
<td>80.5</td>
<td>79.4</td>
<td>3.99</td>
<td></td>
</tr>
<tr>
<td>5–10 events·h−1</td>
<td>2</td>
<td>22</td>
<td>27</td>
<td>18</td>
<td>69</td>
<td>73.5</td>
<td>92.7</td>
<td>88.2</td>
<td>10.07</td>
<td></td>
</tr>
<tr>
<td>&gt;10 events·h−1</td>
<td>0</td>
<td>14</td>
<td>13</td>
<td>75</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nocturnal OXI</td>
<td>33</td>
<td>215</td>
<td>85</td>
<td>99</td>
<td>432</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cohen’s κ

<table>
<thead>
<tr>
<th>AHIPSG</th>
<th>0.339</th>
</tr>
</thead>
</table>

LR+: positive likelihood ratio.
FIGURE 2 Schematic representation of apnoea/hypopnoea index (AHI) oximetry algorithm performance and corresponding likelihood ratios relative to polysomnography-derived AHI when diagnostic cut-off values are set at a) 1 event·h⁻¹, b) 5 events·h⁻¹ or c) 10 events·h⁻¹.

displays the receiver operator curves for each of these cut-off values. AHIOXI showed increasing degrees of diagnostic ability as the cut-off increased and became <90% specific at AHI cut-off of 10 events·h⁻¹. Accuracies remained >79% at all cut-off levels with LR⁺ displaying optimal performance for AHI of ≥5 events·h⁻¹.

Discussion

This study shows that a previously developed and validated neural network machine-learning algorithm based on overnight oximetry recordings [29] can be readily and accurately implemented as a readily scalable operator-independent diagnostic tool for the diagnosis of OSAS in symptomatic children referred for evaluation of OSAS.

Before we discuss the clinical implications of current findings, several methodological issues deserve mention. First, only a single commercially available oximeter model was employed for the present study, and displays industry standard accuracy. In this context, awareness of the potential imprecision of the oximeter being selected and employed during implementation of the approach used in the present study is obviously of great importance [34]. Similarly, the potential imprecisions introduced by the finger probe being selected should also be accounted for in materials and supplies selection [35]. However, we should also remark that the heterogeneity of the oximeters and their intrinsic performances was incorporated into the process of derivation and validation of the cloud-based algorithm, and included 13 different paediatric sleep centres around the world using vastly different oximeters and oximeter data-sampling frequencies [29]. Furthermore, the oximeters used in the PSG and the wearable Bluetooth oximeter were different in this study, but achieved concordance in the scored and automatically detected ODI3%, respectively (r²=0.35; data not shown). Second, while we attribute the designation of “gold standard” to the PSG, there can be considerable night-to-night variability, particularly in sleep architecture, embedded in the test [36, 38], which may lead to substantial imprecision in the diagnostic decision, particularly at the low end of OSAS severity [38, 39]. Although not immediately relevant to this study, since both PSG and Bluetooth
oximetry were implemented concurrently, we should point out that this issue has arisen, albeit inconsistently, in oximetry recordings in children, prompting the recommendation that if a single night yields a negative result in a patient with a high pretest probability, then the addition of two further nights of oximetry recordings should lead to a more reliable clinical decision. [40, 41]. Such an approach would be highly feasible with ambulatory oximetry, but impossible with PSG, such that if the accuracies of oximetry-based approaches were virtually indistinguishable from PSG, there would be inherent advantages at using oximetry in this context.

Overall, the accuracy of the portable oximeter cloud-based algorithm dyad was >79% for AHI estimates of 1–10 events·h⁻¹, and displayed the anticipated progressive declines in sensitivity with increasing AHI cut-off values to higher specificity at an AHI cut-off of 5 events·h⁻¹ (table 2). Thus, and as previously inferred during the process of developing the algorithm, optimal benefits of this automated methodological approach to diagnose OSAS in habitually snoring children [29] is achieved with AHI ≥5 events·h⁻¹. This cut-off value not only corresponds to a virtually universally agreed-upon equipoise criterion for surgical adenotonsillectomy, but is also associated with an upward inflexion in morbidity risks in children with OSAS [42–44]. However, note that the adoption of the AHI 1 event·h⁻¹ cut-off would lead to relatively high rate of false positives, which would then be treated even if their AHIPSG would have been <1 event·h⁻¹. Conversely, the false negative rate of our approach was small, as illustrated by the fact that using the criterion of AHIPSG >1 event·h⁻¹, only 4.7% of false negative cases emerged, from which only 0.6% of cases would be in the moderate or severe OSAS category. The relatively small proportion of children that would be missed using oximetry-based diagnostics might be further reduced by repeating the oximetry-based test for one or more additional consecutive nights [41], an issue that clearly deserves further exploration in future studies. Furthermore, repeating the test within weeks or months if the child’s symptoms persisted would be much more readily achievable than repeating PSG. Thus, the clinical management options and algorithms offered by the PSG, whereby the AHI serves as one of the major parameters guiding clinical intervention, would be indistinguishably afforded by the AHIOXI, albeit at a fraction of the cost and effort.

As alluded previously, the overall scarcity and labour-intensive and financially onerous nature of PSG has prompted the exploration of multiple other suitable diagnostic alternative approaches ranging from questionnaires to ambulatory PSG or to simplified multichannel recordings [45]. However, some such approaches, e.g. respiratory polygraphy, are becoming increasingly accepted as a surrogate diagnostic test in children, despite their reduced accuracy at the low end of the OSAS severity spectrum [10], a limitation that has prompted a lack of endorsement by the AASM [46]. In the present study, our findings clearly show that automated analysis of nocturnal oximetry provides a useful approach to the diagnosis of OSAS among high pretest symptomatic children being referred for evaluation of suspected OSAS. Indeed, and according to the protocol proposed in our previous study which as mentioned above proposed a cut-off of AHI ≥5 events·h⁻¹ [29], OSAS would have been discarded in 38 children, most of them with AHIPSG <5 events·h⁻¹. In addition, we should remark that several of these children would potentially require treatment anyway, due to concurrent SDB-related morbidity. In addition, 184 subjects would be referred for treatment, with 96.7% showing mild OSAS. Such an approach would clearly reduce the need for conventional PSG, a finding that is coherent with our previous results [29], whereby 77.8% children with moderate-to-severe OSAS based on PSG would be identified as such by oximetry coupled to a mobile phone app interfaced with a cloud-based algorithm. Our findings are in close agreement with the
approach and conclusions from a recent study by Papadakis et al. [47], in which oximetry approaches consisting of ODI3% >3.5 events·h⁻¹ successfully predicted post-surgical intervention outcomes.

In summary, this study provides initial confirmatory demonstration that diagnostic precision can be readily achieved via a portable oximeter linked via a smartphone to a cloud-based automated analytic algorithm. This approach offers not only a valid alternative to standard PSG in the context of childhood OSAS, but also that it is a highly scalable, i.e. incorporates the ability to seamlessly continue to deliver the desired service, in this case oximetry-based diagnostics, in the context of increasing demands in order to meet a user need without requiring additional expert personnel (current rate-limiting factor), and therefore should serve as a remarkably affordable option. Therefore, integrated collection of ambulatory nocturnal oximetry signals and their automated processing by well validated algorithms as the one employed herein, should lead to accurate and widely implementable diagnostic tools for childhood OSAS, thereby enabling timely objective evaluation and treatment with the attendant downstream benefits of reduced morbidity.

Inasmuch as the current findings are promising, expanded implementation of the current system to the domiciliary venues and its real life performance in the clinical setting will need to be critically investigated and confirmed.

Conflict of interest: Z. Xu has nothing to disclose. G.C. Gutiérrez-Tobal has nothing to disclose. Y. Wu has nothing to disclose. L. Kheirandish-Gozal acted as a scientific consultant to Serenium Inc., during the conduct of the study. Serenium Inc. graciously provided the oximeter devices, Bluetooth phones and links to algorithm, but was not involved in study design, data acquisition or analysis. X. Ni has nothing to disclose. R. Hornero has nothing to disclose. D. Gozal has nothing to disclose.

Support statement: This work was supported by the Capital Health Research and Development of Special Funding (2018-1-2091); Beijing Municipal Science and Technology Project (Z16110000116050); National Key Research and Development Plan (2017YFC0112502) to X. Ni and Y. Wu; US National Institutes of Health grant HL130984 (L. Kheirandish-Gozal, D. Gozal); and projects DPI2017-84280-R and RTC-2015-3446-1 from “Ministerio de Ciencia, Innovación y Universidades” (Spanish Government) and European Regional Development Fund (FEDER) to G.C. Gutiérrez-Tobal and R. Hornero. Funding information for this article has been deposited with the Crossref Funder Registry.

References