Interobserver Agreement When Diagnosing Hypoventilation in Children With Neuromuscular Disorders

Adelaide Withers, MD1,2,3, Andrew Wilson, MD1,2,3, Peter Jacoby, MS2, Leanne Gauld, MD4,5, Andre Schultz, PhD1,6,7, Anne O'Donnell, PhD1, Archana Chacko, MD4,5, Jenny Downs, PhD2,3

1Department of Respiratory and Sleep Medicine, Perth Children's Hospital, Perth, Australia
2Telethon Kids Institute, Perth, Australia
3Curtin School of Allied Health, Faculty of Health Sciences, Curtin University, Perth, Australia
4Queensland Children's Hospital, Brisbane, Australia
5University of Queensland, Brisbane, Australia
6Wal-Yan Respiratory Research Centre, Telethon Kids Institute, University of Western Australia, Perth, Australia
7Division of Paediatrics, Faculty of Medicine, University of Western Australia, Perth, Australia

Neuromuscular disorders can lead to nocturnal hypoventilation. Accurate diagnosis of hypoventilation is imperative to guide treatment decisions. This study determined interobserver agreement for a number of definitions of nocturnal hypoventilation in children and adolescents with neuromuscular disorders. Overall mean interobserver agreement was 89% (range 66–100%); however, reliability of agreement was moderate at best (Fleiss κ = 0.574, p < 0.001). When hypoventilation was present, the objective definition used most frequently was an average increase in partial pressure of carbon dioxide (pCO₂) ≥ 3 mm Hg from NREM to REM. The appearance of the transcutaneous CO₂ (TCO₂) trend graph and an increase in pCO₂ ≥ 10 mm Hg from awake to asleep were most often associated with a false positive diagnosis. The variation and at best moderate agreement between pediatric sleep physicians observed in this study when diagnosing hypoventilation in children with neuromuscular disorders may be partially explained by the existence of multiple definitions and failure to remove artifact and “drift” from the TCO₂ data.

Keywords Hypoventilation; Hypercapnia; Neuromuscular disorder; Polysomnography; Interobserver agreement.

INTRODUCTION

Neuromuscular disorders (NMD) that affect respiratory muscles can lead to nocturnal hypoventilation. Nocturnal hypoventilation is defined by the presence of hypercapnia during sleep. Although awake hypercapnia is defined as arterial partial pressure of carbon dioxide (pCO₂) > 45 mm Hg [1], a clinically relevant change in pCO₂ during sleep has not been determined [2]. Thresholds to define nocturnal hypoventilation vary throughout the world [3,4] as the level of hypercapnia associated with clinically relevant morbidity and mortality remains unknown [3]. Current definitions of hypoventilation are extrapolated from limited, normative data from healthy individuals [1] and decided by consensus and expert opinion rather than being evidence based.

Nocturnal hypoventilation is detected by measuring pCO₂ during polysomnography (PSG), most commonly by transcutaneous measurement (TCO₂) and less often by end-tidal measurement (ETCO₂). The American Academy of Sleep Medicine (AASM) scoring man-
ual defines hypoventilation in adults as an increase in arterial pCO₂ or appropriate surrogate to > 55 mm Hg for ≥ 10 minutes during sleep [1] and/or a rise in arterial pCO₂ or appropriate surrogate of ≥ 10 mm Hg (in comparison to an awake supine value) to > 50 mm Hg for ≥ 10 minutes during sleep [1]. For children, hypoventilation is defined as an increase in arterial pCO₂ or appropriate surrogate during sleep to > 50 mm Hg for > 25% of total sleep time (TST) [1]. The AASM manual states that adult rules may be used from the age of 13 [1].

The Australasian Sleep Association/Australasian Sleep Technologists Association (ASA/ASTA) published an addendum to the 2007 AASM scoring manual [5] to ensure that recommendations were relevant to the Australasian population. Specific pediatric recommendations [6] endorsed the AASM pediatric definition of hypventilation and included two additional definitions: an increase in pCO₂ ≥ 10 mm Hg from awake to sleep [6] and an average rise in pCO₂ ≥ 3 mm Hg from NREM to REM sleep (REM related hypoventilation) [6]. A recent survey demonstrated that some pediatric sleep physicians in Australasia use an average rise in pCO₂ ≥ 3 mm Hg during REM sleep to define REM related hypoventilation [7]; however, it is not known if this definition is used in other countries.

Incorrect application or interpretation of scoring rules can lead to inaccuracy when examining PSG data [5,8]. It is imperative that rules are applied consistently [9] to ensure good interobserver agreement and concordance [5]. Clarification of scoring rules is a key component to achieving consistency and accuracy [5,9-11].

It is crucial that the onset of nocturnal hypoventilation is accurately identified in individuals with NMD to allow timely treatment with non-invasive ventilation (NIV). As the existence of numerous definitions of hypoventilation could have a negative impact upon concordance, interobserver agreement and diagnostic accuracy, this study examined interobserver agreement when diagnosing nocturnal hypoventilation in children and adolescents with NMD.

METHODS

This study was conducted in accordance with The National Statement of Ethical Conduct in Human Research 2007 (updated 2018) and registered with approval to publish by the Child and Adolescent Health Services GEKO platform (approval number 6245), a low and negligible risk pathway for audits of clinical practice and quality improvement.

Reports from 20 PSGs performed between 2003 and 2016 were selected from children with a range of NMD diagnoses including Duchenne muscular dystrophy, spinal muscular atrophy, and congenital myopathy. Ages ranged from 10 months to 15 years (median 4 years) with 11 females. None of the PSGs were performed with NIV.

Each PSG report (with clinical and identifying details removed) and overnight sleep technologist’s notes of relevance to TCO₂ measurements (for example, times of probe recalibration) was presented in a randomized order to 6 pediatric sleep physicians (4 from Perth Children’s Hospital, 2 from Queensland Children’s Hospital). Each physician documented whether hypoventilation was present for each PSG, and if so, which definitions they used to determine this. The definitions of hypoventilation that could be used included the pediatric AASM definitions [1] and the pediatric ASA/ASTA definitions [6]. Adult AASM [1] definitions were included as some PSGs were from adolescents. The appearance of the TCO₂ trend graph, TCO₂ and/or ETCO₂ values from overnight observation sheets were included as per a survey of Australasian pediatric sleep physicians [7]. Physicians could use more than one definition of hypoventilation for each PSG. Raw PSG data was not available for review to avoid other parameters that were not part of the definitions of hypoventilation that were being examined to influence the physician’s decision about the presence or absence of hypoventilation.

The number of times each definition was used to diagnose hypoventilation was tallied. Interobserver agreement was calculated as a percentage and reliability of agreement was assessed with Fleiss kappa (κ). Interobserver agreement of > 60% was deemed substantial [12] and used to determine the presence or absence of hypoventilation for each PSG.

Statistical analysis was completed in IBM SPSS (version 26; IBM Corp., Armonk, NY, USA) and GraphPad Prism (version 9, GraphPad Software, Inc., Boston, MA, USA). A p-value of 0.05 was chosen to indicate statistical significance.

RESULTS

Interobserver agreement of > 60% was reached for 19/20 PSGs, with hypoventilation present in 6/20 and absent in 13/20 (Table 1). Mean interobserver agreement was 89% (range 66%-100%) with Fleiss k of 0.574 (95% CI 0.570–0.577, p < 0.001). Mean interobserver agreement was slightly higher for the absence (91%) than the presence (86%) of hypoventilation.

The frequency with which observers used each definition of hypoventilation is shown in Fig. 1. When there was > 60% agreement that hypoventilation was present (open bars), the definition of hypoventilation used most frequently was the appearance of the TCO₂ trend graph (n = 23). An average increase in pCO₂ ≥ 3 mm Hg from NREM to REM was the most frequently used objective definition of hypoventilation (n = 15).

Where agreement was > 60% that hypoventilation was not present (shaded bars), a diagnosis of hypoventilation by an individual physician was considered a false positive. The appearance of the TCO₂ trend graph (n = 6) and a pCO₂ increase of ≥ 10 mm Hg from awake to asleep (n = 4) were the definitions used most often in this circumstance.
When determining whether nocturnal hypoventilation was present or absent in children and adolescents with NMD in this study, interobserver agreement was excellent, but reliability of agreement was moderate at best. This discrepancy is likely explained by the high probability of chance agreement when the outcome variable is binary and the very small number of PSGs with hypoventilation. Therefore, it is highly likely that interobserver agreement is overestimated, and reliability of agreement is more accurate. Even when all observers agreed that hypoventilation was present for an individual PSG, there was still significant variation between observers when defining hypoventilation. These results clearly demonstrate heterogeneity in clinical practice and suboptimal agreement between pediatric sleep physicians when defining hypoventilation in children with NMD, which may reduce accuracy of diagnosis. This is problematic because the clinical consequences of missing the onset of nocturnal hypoventilation could be profound; losing the opportunity to institute NIV and improve survival, quality of life [13], and mental health [14]. On the contrary, a false positive diagnosis of nocturnal hypoventilation could lead to unnecessary use of NIV, which could have a negative impact upon quality of life and care burden.

Accurate diagnosis of nocturnal hypoventilation requires both precise measurement of pCO₂ during sleep and consistent application of well-defined scoring rules. Data from TCO₂ measurements can be impacted by artifact or “drift” of the signal. Interpretation of TCO₂ data requires careful assessment to determine validity of the results, a highly subjective process which varies with experience and understanding of methods and limitations of TCO₂ measurement. In this study, a false positive di-

---

**DISCUSSION**

Table 1. Individual observer results and interobserver agreement for each polysomnogram

<table>
<thead>
<tr>
<th>PSG</th>
<th>Observer 1</th>
<th>Observer 2</th>
<th>Observer 3</th>
<th>Observer 4</th>
<th>Observer 5</th>
<th>Observer 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>8</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>11</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>14</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>15</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>16</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>17</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>18</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>19</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>20</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

*Study 5 excluded from further analysis as interobserver agreement was < 60%.

PSG, polysomnography.

---

Fig. 1. Frequencies of use by definitions of hypoventilation. Definition 1, increase in arterial pCO₂ to >50 mm Hg for >25% TST; Definition 2, increase in arterial pCO₂ of ≥10 mm Hg to >50 mm Hg for ≥10 min; Definition 3, increase in arterial pCO₂ to >55 mm Hg for ≥10 min; Definition 4, increase in arterial pCO₂ to >50 mm Hg for >25% TST and a clear difference between awake and asleep pCO₂; Definition 5, increase in pCO₂ ≥10 mm Hg from awake to asleep; Definition 6, average rise in pCO₂ ≥3 mm Hg from NREM to REM sleep; Definition 7, appearance of the TCO₂ trend graph; Definition 8, TCO₂ and/or ETCO₂ values from overnight observation. *Shaded bars are likely a false positive diagnosis of hypoventilation. pCO₂, partial pressure of carbon dioxide; TST, total sleep time; REM, rapid eye movement; NREM, non-REM; TCO₂, transcutaneous CO₂; ETCO₂, end-tidal CO₂.
agnosis of hypoventilation most often occurred when relying upon the appearance of the TCO₂ trend graph or a pCO₂ increase of ≥ 10 mm Hg from awake to asleep, likely caused by misinterpretation of artifact or “drift.” These results highlight the need to critically evaluate TCO₂ data, remove artifact and the importance of using objective parameters to define hypoventilation.

The use of multiple definitions of hypoventilation observed in this study likely explains the moderate at best reliability of agreement. Although consistent application of well-defined scoring rules is vital for accurate diagnosis of hypoventilation, these results are consistent with and reflect the real-world scenario where there are numerous scoring rules that differ for adults and children. Therefore, these results highlight the need for clarification of scoring rules and a unified approach to achieve consistency when defining nocturnal hypoventilation [5,9-11]. This is particularly the case for adolescents. In this study, adult definitions of hypoventilation were not often used. Although this could reflect pediatric sleep physicians being unfamiliar with adult scoring rules and information pertaining to adult definitions (for example time with TCO₂ > 55 mm Hg) not being included in the pediatric PSG report, the most likely explanation is that adult scoring rules have been shown to underestimate the presence of sleep related pathology in adolescents as they are not sensitive enough [15], hence pediatric sleep physicians are reluctant to use these definitions [7].

The significance of REM related hypercapnia and hypoventilation warrants further consideration, particularly as an average rise in pCO₂ ≥ 3 mm Hg in REM sleep is included in the pediatric ASA/ASTA guidelines [6]. Hypoventilation usually first appears in and is more pronounced in REM sleep due to atonia of accessory respiratory muscles [16]. However, REM hypercapnia is not always present in individuals with NMD and diaphragmatic weakness as some develop compensatory mechanisms such as reduced REM sleep time and preserved phasic activity of the sternomastoid and genioglossus muscles [17]. Despite these limitations, the results of this study highlight the importance of recognising REM hypoventilation in individuals with NMD, as an average rise in pCO₂ from NREM to REM of ≥ 3 mm Hg was the objective definition in this study that was used most frequently when agreement was substantial that hypoventilation was present. In individuals with NMD, the presence of REM hypoventilation without significant hypercapnia during NREM sleep is likely a sign of “early” [18] or impending nocturnal hypoventilation, which may warrant closer and more frequent follow-up.

Strengths and Limitations

The strengths of this study include all observers having expertise in interpreting PSG data from children with NMD (range 4 to over 15 years as a consultant sleep physician) and inclusion of two tertiary centers, which captures differences in clinical practice. The heterogeneity of the study population is reflective of real-world clinical scenarios.

The major limitations of this study include the retrospective nature and small sample size. Fleiss κ may have been underestimated due to the low prevalence of hypoventilation in this sample [19]. Due to small numbers, it was not possible to examine adolescents separately. There was a range of physician experience; however, the number of physicians was too small to determine the impact of differing levels of experience upon interobserver agreement. As this study only included Australian centers, the results cannot be generalized to other countries.

The authors chose to use physician agreement (> 60%) as the “gold standard”; however, this is not a perfect reference standard and can introduce “imperfect gold standard bias” [20]. The information included in the PSG report changed over time and did not always reflect the parameters that are used to define hypoventilation. For example, the percentage of TST with pCO₂ > 50 mm Hg was not stated for all reports. Therefore, the clinical utility of these definitions will have been underestimated.

Conclusion

The variation and at best moderate interobserver agreement demonstrated in this study when diagnosing nocturnal hypoventilation in children with NMD may be partially explained by the existence of multiple definitions of hypoventilation and failure to remove artifact and “drift” from the TCO₂ data. This is concerning as the suboptimal interobserver agreement may imply diagnostic inaccuracy. These results highlight the need to identify the most accurate definition of hypoventilation in this population and ensure definitions are applied consistently. A rise in pCO₂ of ≥ 3 mm Hg in REM sleep may be an early indicator of hypoventilation in individuals with NMD and warrants further investigation. A prospective study with a larger sample size and incorporation of additional sleep parameters (for example apnoea hypopnoea index) would strengthen the findings of this study. By including a larger number of observers from different countries with varying levels of experience, the conclusions would have context for other countries and allow exploration of the impact of physician experience upon the diagnosis of hypoventilation. Future research should focus on determining which definitions of hypoventilation correlate best with clinical outcome measures.

Availability of Data and Material
The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

Author Contributions
ny Downs, Peter Jacoby. Writing—original draft: Adelaide Withers, Jenny Downs. Writing—review & editing: Andrew Wilson, Jenny Downs, Peter Jacoby, Andre Schultz, Anne O’Donnell, Leanne Gauld, Archana Chacko.

Conflicts of Interest
The authors have no potential conflicts of interest to disclose.

Funding Statement
None

Acknowledgements
The authors would like to thank Professor Graham Hall for his contribution to conceptualisation of this study.

REFERENCES