SMR

BRIEF COMMUNICATION
Sleep Med Res 2023;14(4):240-244
https://doi.org/10.17241/smr.2023.01935

pISSN 2093-9175
eISSN 2233-8853

Interobserver Agreement When Diagnosing Hypoventilation in Children With Neuromuscular Disorders

Adelaide Withers, MD[1,2,3], Andrew Wilson, MD[1,2,3], Peter Jacoby, MS[2], Leanne Gauld, MD[4,5], Andre Schultz, PhD[1,6,7], Anne O'Donnell, PhD[1], Archana Chacko, MD[4,5], Jenny Downs, PhD[2,3]

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

Received: October 17, 2023
Revised: November 15, 2023
Accepted: November 25, 2023

Corresponding Author
Adelaide Withers, MD
Department of Respiratory and Sleep Medicine,
Perth Children's Hospital,
Hospital Ave, Nedlands, WA,
Australia 6009
Tel +61 08 6456 5414
E-mail Adelaide.Withers@health.wa.gov.au

ORCID iDs
Adelaide Withers
https://orcid.org/0000-0002-1640-8743
Andrew Wilson
https://orcid.org/0000-0003-0875-2736
Peter Jacoby
https://orcid.org/0000-0001-5092-7697
Leanne Gauld
https://orcid.org/0000-0003-0819-894X
Andre Scultz
https://orcid.org/0000-0002-2000-8910
Anne O'Donnell
https://orcid.org/0009-0001-2619-2229
Archana Chacko
https://orcid.org/0000-0001-7265-6963
Jenny Downs
https://orcid.org/0000-0001-7358-9037

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 $\kappa = 0.574$, $p < 0.001$). When hypoventilation was present, the objective definition used most frequently was an average increase in partial pressure of carbon dioxide (pCO_2) ≥ 3 mm Hg from NREM to REM. The appearance of the transcutaneous CO_2 (TCO_2) trend graph and an increase in pCO_2 ≥ 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 TCO_2 data.

Sleep Med Res 2023;14(4):240-244

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 (pCO_2) > 45 mm Hg [1], a clinically relevant change in pCO_2 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 pCO_2 during polysomnography (PSG), most commonly by transcutaneous measurement (TCO_2) and less often by end-tidal measurement ($ETCO_2$). The American Academy of Sleep Medicine (AASM) scoring man-

 Copyright © 2023 The Korean Society of Sleep Medicine

ual defines hypoventilation in adults as an increase in arterial pCO_2 or appropriate surrogate to > 55 mm Hg for ≥ 10 minutes during sleep [1] and/or a rise in arterial pCO_2 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 pCO_2 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 hypoventilation and included two additional definitions; an increase in pCO_2 ≥ 10 mm Hg from awake to sleep [6] and an average rise in pCO_2 ≥ 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 pCO_2 ≥ 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 TCO_2 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 TCO_2 trend graph, TCO_2 and/or $ETCO_2$ 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 κ 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 TCO_2 trend graph (n = 23). An average increase in pCO_2 ≥ 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 TCO_2 trend graph (n = 6) and a pCO_2 increase of ≥ 10 mm Hg from awake to asleep (n = 4) were the definitions used most often in this circumstance.

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

PSG	Observer						Hypoventilation, n (%)	
	1	2	3	4	5	6	Present	Absent
1	Y	Y	Y	Y	Y	Y	6/6 (100)	
2	Y	Y	Y	Y	Y	N	5/6 (83)	
3	Y	Y	Y	N	Y	Y	5/6 (83)	
4	N	N	N	N	N	N		6/6 (100)
5	N	Y	N	N	Y	Y	3/6 (50)*	3/6 (50)*
6	Y	N	N	N	Y	N		4/6 (66)
7	N	N	N	N	N	N		6/6 (100)
8	Y	Y	Y	Y	Y	N	5/6 (83)	
9	Y	Y	Y	Y	Y	Y	6/6 (100)	
10	Y	N	N	N	N	N		5/6 (83)
11	N	N	N	N	Y	N		5/6 (83)
12	N	N	N	N	N	N		6/6 (100)
13	N	N	N	N	N	N		6/6 (100)
14	N	N	N	N	N	N		6/6 (100)
15	N	N	N	N	N	N		6/6 (100)
16	N	N	N	N	N	N		6/6 (100)
17	Y	N	N	N	Y	N		4/6 (66)
18	Y	N	Y	Y	Y	N	4/6 (66)	
19	N	N	N	N	N	N		6/6 (100)
20	N	N	N	N	Y	N		5/6 (83)

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

DISCUSSION

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 pCO_2 during sleep and consistent application of well-defined scoring rules. Data from TCO_2 measurements can be impacted by artifact or "drift" of the signal. Interpretation of TCO_2 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 TCO_2 measurement. In this study, a false positive di-

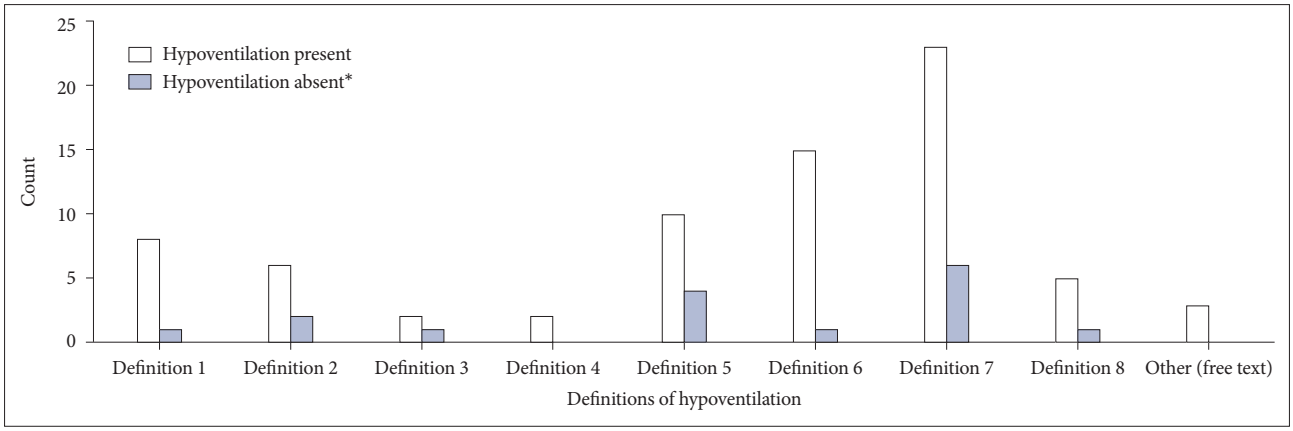

Fig. 1. Frequencies of use by definitions of hypoventilation. Definition 1, increase in arterial pCO_2 to >50 mm Hg for >25% TST; Definition 2, increase in arterial pCO_2 of ≥10 mm Hg to >50 mm Hg for ≥10 min; Definition 3, increase in arterial pCO_2 to >55 mm Hg for ≥10 min; Definition 4, increase in arterial pCO_2 to >50 mm Hg for >25% TST and a clear difference between awake and asleep pCO_2; Definition 5, increase in pCO_2 ≥10 mm Hg from awake to asleep; Definition 6, average rise in pCO_2 ≥3 mm Hg from NREM to REM sleep; Definition 7, appearance of the TCO_2 trend graph; Definition 8, TCO_2 and/or $ETCO_2$ values from overnight observation. *Shaded bars are likely a false positive diagnosis of hypoventilation. pCO_2, partial pressure of carbon dioxide; TST, total sleep time; REM, rapid eye movement; NREM, non-REM; TCO_2, transcutaneous CO_2; $ETCO_2$, end-tidal CO_2.

agnosis of hypoventilation most often occurred when relying upon the appearance of the TCO_2 trend graph or a pCO_2 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 TCO_2 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 TCO_2 > 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 pCO_2 ≥ 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 pCO_2 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 pCO_2 > 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 TCO_2 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 pCO_2 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

Conceptulization: Adelaide Withers, Andrew Wilson. Data curation: Adelaide Withers, Andrew Wilson, Andre Schultz, Anne O'Donnell, Leanne Gauld, Archana Chacko. Formal analysis: Adelaide Withers, Jenny Downs, Peter Jacoby. Investigation: Adelaide Withers, Andrew Wilson. Methodology: Adelaide Withers, Andrew Wilson, Jenny Downs, Peter Jacoby. Project Administration: Adelaide Withers. Supervision: Jenny Downs, Andrew Wilson, Peter Jacoby. Validation: Adelaide Withers, Jen-

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

1. Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. *J Clin Sleep Med* 2012;8:597-619.
2. Berlowitz DJ, Spong J, O'Donoghue FJ, Pierce RJ, Brown DJ, Campbell DA, et al. Transcutaneous measurement of carbon dioxide tension during extended monitoring: evaluation of accuracy and stability, and an algorithm for correcting calibration drift. *Respir Care* 2011;56:442-8.
3. Georges M, Nguyen-Baranoff D, Griffon L, Foignot C, Bonniaud P, Camus P, et al. Usefulness of transcutaneous PCO2 to assess nocturnal hypoventilation in restrictive lung disorders. *Respirology* 2016;21:1300-6.
4. Paiva R, Krivec U, Aubertin G, Cohen E, Clément A, Fauroux B. Carbon dioxide monitoring during long-term noninvasive respiratory support in children. *Intensive Care Med* 2009;35:1068-74.
5. Thornton AT, Ruehland WR, Duce B, Wheatley JR, Douglas J, Rochford PD, et al. *ASTA/ASA commentary on AASM manual for the scoring of sleep and associated events.* North Strathfield: Australasian Sleep Association 2010.
6. ASA/ASTA Paediatric Working Group. *ASTA/ASA addendum to AASM guidelines for recording and scoring of paediatric sleep.* Version 5.0. North Strathfield: Australasian Sleep Association 2011.
7. Withers AL, Downs J, Wilson AC, Hall G. Diagnosis of nocturnal hypoventilation in pediatric neuromuscular disorders: a survey of clinical practice in Australia and New Zealand. *J Sleep Med* 2023;20:35-40.
8. Stepnowsky CJ Jr, Berry C, Dimsdale JE. The effect of measurement unreliability on sleep and respiratory variables. *Sleep* 2004;27:990-5.
9. Redline S, Budhiraja R, Kapur V, Marcus CL, Mateika JH, Mehra R, et al. The scoring of respiratory events in sleep: reliability and validity. *J Clin Sleep Med* 2007;3:169-200.
10. Danker-Hopfe H, Kunz D, Gruber G, Klösch G, Lorenzo JL, Himanen SL, et al. Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *J Sleep Res* 2004;13:63-9.
11. Whitney CW, Gottlieb DJ, Redline S, Norman RG, Dodge RR, Shahar E, et al. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep* 1998;21:749-57.
12. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276-82.
13. Piper AJ. Nocturnal hypoventilation - identifying & treating syndromes. *Indian J Med Res* 2010;131:350-65.
14. Huston M, Withers AL, Lam J, Wilson A, Downs J. Respiratory health, sleep dysfunction, and mental health in children and adolescents with a neuromuscular disorder: a descriptive qualitative study. *J Sleep Med* 2023;20:11-8.
15. Evangelisti M, Forlani M, De Pozzo M, Liverani ME, Villa MP. Sleep clinical record and polysomnography scores in adolescents with sleep disordered breathing. *Eur Respir J* 2018;52:PA4586.
16. Becker HF, Piper AJ, Flynn WE, McNamara SG, Grunstein RR, Peter JH, et al. Breathing during sleep in patients with nocturnal desaturation. *Am J Respir Crit Care Med* 1999;159:112-8.
17. Aboussouan LS. Sleep-disordered breathing in neuromuscular disease. *Am J Respir Crit Care Med* 2015;191:979-89.
18. Schaefer J, Davey MJ, Nixon GM. Sleep-disordered breathing in school-aged children with Prader-Willi syndrome. *J Clin Sleep Med* 2022;18:1055-61.
19. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257-68.
20. Weinstein S, Obuchowski NA, Lieber ML. Clinical evaluation of diagnostic tests. *AJR Am J Roentgenol* 2005;184:14-9.