Published in final edited form as:

J Crit Care. 2013 August ; 28(4): 490–497. doi:10.1016/j.jcrc.2012.11.009.

Pediatric upper airway obstruction: Inter-observer variability is the road to perdition. Short title: Inconsistent assessments of stridor amongst pediatric practitioners

Robinder G. Khemani, MD, MsCI^{1,2}, James B. Schneider, MD³, Rica Morzov, RN¹, Barry Markovitz, MD, MPH^{1,2}, and Christopher JL Newth, MD, FRCPC^{1,2}

¹Children's Hospital Los Angeles, Los Angeles CA

²University of Southern California Keck School of Medicine, Los Angeles, CA

³Cohen Children's Medical Center of New York; North Shore Long Island Jewish Health System; New Hyde Park, NY

Abstract

Purpose—To determine the inter-observer variability in the clinical assessment of pediatric upper airway obstruction (UAO). To explore how variability in assessment of UAO may contribute to risk factors and incidence of post-extubation UAO.

Materials—Prospective trial in two tertiary care Pediatric intensive care units. Bedside practitioners performed simultaneous, blinded UAO assessments on 112 children after endotracheal extubation.

Results—Agreement amongst RTs, RNs and MDs was poor for cyanosis ($\kappa = 0.01$), and hypoxemia at rest ($\kappa = 0.14$); and fair for consciousness ($\kappa = 0.27$), air entry ($\kappa = 0.32$), hypoxemia with agitation ($\kappa = 0.27$), and pulsus paradoxus ($\kappa = 0.23$). When looking at “stridor” and “retractions,” defined using more than 2 grades of severity from the Westley Croup Score, the inter-rater reliability was moderate ($\kappa = 0.43$ and $\kappa = 0.47$, respectively). This could be improved marginally by dichotomizing the presence or absence of stridor ($\kappa = 0.54$) or retractions ($\kappa = 0.53$). The overall incidence of UAO after extubation (stridor plus retractions) could range from 7–22%, depending on how many providers were required to agree.

Conclusions—Physical findings routinely used for UAO have poor inter-observer reliability amongst bedside providers. This variability may contribute to inconsistent findings regarding incidence, risk factors, and therapies for post-extubation UAO.

Keywords

Airway Obstruction; Pediatrics; Intensive Care; Intubation; Endotracheal

© 2012 Elsevier Inc. All rights reserved.

Corresponding Author and reprints: Robinder G. Khemani, MD, MsCI, 4650 Sunset Blvd Mailstop 12, Children's Hospital Los Angeles, Los Angeles, CA 90027, rkhemani@chla.usc.edu, Phone (323) 361-2376, Fax (323) 361-1001.

There are no prior publications or submissions with any overlapping information.

This study has been presented as an abstract at the American Thoracic Society Meeting in May of 2012 in San Francisco, CA.

No specific financial support was used for this study. There are no conflicts of interests for any of the authors

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Upper airway obstruction (UAO) is common to a variety of pediatric diseases. It is imperative that clinicians accurately diagnose UAO, assess its severity and likely course in order to provide timely and correct treatment.

Assessments of UAO are largely based on clinical measures of stridor plus work of breathing. Such assessments are observer dependent, with potentially high variability. Several scoring systems exist for croup, most derived from the Westley Croup Scale¹. For better or worse, these scores have been applied to other causes of pediatric UAO, particularly that following extubation. They have not been validated against objective measures of airway obstruction, and the literature evaluating the inter-observer reliability is sparse. In clinical practice, treatment decisions for UAO are often based on a single provider's assessment of stridor and retraction severity. An equally well-trained provider may have a different assessment and deem the child not to have UAO, resulting in different treatments.

This assessment practice is frequent in pediatric intensive care units (PICUs), specifically in relation to post-extubation UAO. Incidence rates for post-extubation UAO range from 6–69 %, and result in reintubation 2–22% of the time^{2–5}. Based on clinical grounds alone, it is estimated that up to 1/3 of all failed extubations result from UAO^{6,7}. Unfortunately, characterizing UAO after extubation is not standardized, and the subjective nature of this assessment may in part contribute to the variability in reported incidence, and may explain the inconsistent response to therapies. Given the uncertainty of evaluations of post-extubation UAO, definitive risk factors have not been identified in children, limiting the development of evidence-based standards for the prevention or treatment of post-extubation UAO.

We sought to determine the variability in the usual care clinical assessment of post extubation UAO by intensive care providers (respiratory therapists (RT), pediatric intensive care nurses (RN), and pediatric intensive care physicians (MD)). Providers performed simultaneous assessments of UAO severity on children extubated in the PICU. We hypothesized that providers who routinely assess for UAO severity would disagree on the presence of UAO, and this disagreement would contribute to differences in incidence rates for UAO and make risk factors for post-extubation UAO difficult to identify.

Methods

Subject Recruitment

We conducted a 2 site prospective observational trial in the pediatric and cardiothoracic intensive care units at Children's Hospital Los Angeles, and the PICU at Cohen Children's Medical Center of New York. The study was performed at 2 sites because of perceived different rates of post-extubation UAO between the two hospitals. The study was approved by the institutional review boards at both institutions (CHLA: 10-00088, CCH: 10-149A) with waiver of consent for the patients, and waiver of written consent for the providers (an information sheet). Patients were included in the study if they were endotracheally intubated (oral or nasal, cuffed or uncuffed) and had a planned extubation from mechanical ventilation during daytime hours, Monday through Friday. There were no specific exclusion criteria. The bedside providers assigned to the patient (nurse (RN), respiratory therapist (RT), pediatric critical care or cardiology fellow or pediatric critical care attending (MD)) completed the UAO assessments. If those providers were not available at the pre-defined study assessment time points (5 minutes and 60 minutes after extubation), then we recruited

another provider with the same job title working in the unit on that day. We required that each time assessment have a provider in each job category for inclusion in the analysis.

Study Measurements

Prior to extubation, research personnel prospectively gathered demographics, reason for intubation, the size of endotracheal tube, the presence of a leak at intubation (when documented), the use of steroids prior to extubation, standardized clinical assessment of the patient's respiratory examination and vital signs on minimal support prior to extubation, ventilator support prior to extubation, and Glasgow Coma Scale. Bedside respiratory therapists evaluated Negative Inspiratory Force just prior to extubation. To evaluate the appropriateness of endotracheal tube size, we computed the difference between the actual endotracheal tube size and that which would have been predicted by the modified Cole formula, based on age⁸, with a modification for of a half size smaller for cuffed endotracheal tubes, as suggested by Deakers et al³. Prior to extubation, we systematically assessed for the presence of an air leak around the endotracheal tube audible to the ear at a pressure less than 30 cm H₂O with the head in midline position, as is routinely done in most PICUs prior to extubation⁹. For 24 hours after extubation, we monitored patients for clinical outcomes including reintubation, non-invasive mechanical ventilation (Bi-Level Positive Pressure Ventilation (BIPAP), Continuous Positive Airway Pressure (CPAP), high flow nasal cannula), and the use of inhaled alpha agonists, heliox, or steroids.

UAO scoring by RTs, RNs, and MDs occurred independently and simultaneously 5 minutes and 60 minutes after extubation. Research personnel ensured that providers remained blinded to other provider's assessments. The scoring tool is detailed in Table 1, which contains elements of the original score by Westley, in addition to other physical examination findings which clinicians use for decision making.¹ Research personnel briefly instructed all providers on definitions for each of the elements of the scoring tool. Pulsus paradoxus was assessed by palpation of the brachial or radial artery, as well as examining pulse oximetry waveforms, and arterial blood pressure waveforms, when available. Hypoxemia was defined as oxygen saturation below the patient's normal baseline level (to account for patients with cyanotic heart disease).

Study Outcomes

Our primary outcome measure was the inter-observer variability between intensive care providers for each of the elements of the scoring tool. Secondary outcomes included descriptive analysis of potential risk factors for post-extubation UAO, the frequency of use of medications to treat or prevent post-extubation UAO, and the frequency of extubation failure and non-invasive mechanical ventilation after extubation. Finally, we describe how reported incidence and potential risk factors for UAO may vary based upon the provider's agreement of UAO severity.

Analysis

We assessed inter-observer variability between all three providers for each of the elements listed in Table 1 by computing non-weighted kappa statistics with 95% confidence intervals (CI). We did not evaluate the total or individual points of the Westley score, but rather the descriptive categories of each variable. For variables that had more than two possible choices (i.e. none, mild, moderate, severe), we also dichotomized the variables into "present" (mild, moderate or severe) or "absent" (none). We explored the source of the variability by further breaking down each assessment with two by two comparisons between providers (RN to RT, RN to MD, RT to MD), again computing kappa statistics. For 95% CI for kappa values with more than two possible outcomes or more than 2 raters, we employed

a bootstrapping method with 1000 replications¹⁰. Interpretation of the point estimates for the kappa statistic are based upon the scale initially proposed by Cohen^{11,12}.

We examined descriptive statistics for each of the potential risk factors and outcome variables, and summarize them as median and interquartile (IQR) range or count and frequency. To describe the potential variability in risk factors based upon agreement in UAO severity, we defined four categories for probability of UAO: none, possible, probable, and definite. “None” required that none of the three providers thought the patient had the combination of dichotomized stridor plus retractions. “Possible UAO” was defined as the presence of stridor plus retractions by one provider. “Probable UAO” required two providers to agree, and “Definite UAO” required all three providers to agree. Both time points were considered (5 minutes and 60 minutes) for categorization of UAO probability, but only simultaneous assessments were evaluated for agreement. We compared those who had no UAO to three composite groups (Definite UAO, Probable + Definite UAO, Possible + Probable + Definite UAO). This analysis was meant to be descriptive and was not powered for comparisons between groups, so p-values are not reported. All analysis was performed in Stata version 10 (StatCorp, College Station, TX).

Results

A total of 112 patients were included, with 224 possible time points for UAO assessment. Four patients did not have all providers available for the 60 minute post-extubation assessment, leaving 220 time points for analysis. Seven patients were re-intubated, and 30 patients received non-invasive ventilation within 24 hours of extubation.

Stridor Score Assessments

Table 2 summarizes the agreement of individual elements of the UAO tool. Agreement amongst RTs, RNs and MDs was poor for assessments of cyanosis ($\kappa = 0.01$), and hypoxemia at rest ($\kappa = 0.14$); and fair for consciousness ($\kappa = 0.27$), air entry ($\kappa = 0.32$), hypoxemia with agitation ($\kappa = 0.27$), and pulsus paradoxus ($\kappa = 0.23$). When looking at “stridor” and “retractions,” defined using the original granularity of the Westley score (Table 1), the inter-rater reliability was moderate ($\kappa = 0.43$ and $\kappa = 0.47$, respectively). This could be improved marginally by dichotomizing the presence or absence of stridor ($\kappa = 0.54$) or retractions ($\kappa = 0.53$). In general, patterns between types of providers (i.e. RN to RT, Table 3) were similar to the patterns for the entire group (RN, RT, MD), although some differences are noted in Table 2 (i.e. higher RN to RT agreement on retractions ($\kappa = 0.65$)).

Clinical risk factors and Outcomes

Given the high inter-observer variability of various elements of the scoring tool for UAO severity, we sought to describe the potential differences in incidence, risk factors, and outcomes that may be attributable to post-extubation UAO, based on different definitions. The only two elements with reasonable inter-observer reliability were dichotomized stridor and dichotomized retractions. Because retractions are not unique to upper airway obstruction, we required that individual providers assess patients as having both stridor and retractions to be labeled as UAO. Breakdown of potential risk factors and outcomes are analyzed in table 4, categorized by probability of UAO, as defined above.

The overall incidence of UAO could range from 7–22% of all extubations, depending on the definition (8, 17, or 25 patients). Overall, 7 patients (6%) were re-intubated: 4 of these patients had “definite UAO”, and 1 had “possible UAO.” The two other patients that were re-intubated were not thought to have UAO, but residual pulmonary or cardiac disease. When all 3 providers agreed the patient had UAO, 50% of patients were re-intubated; when

at least 1 provider thought the patient had UAO, 20% were re-intubated. Patients deemed not to have UAO received racemic epinephrine 2% of the time, as compared to 40–75% of the time for those deemed to have UAO (Table 4). Patients with Possible/Probable/Definite UAO had a median weight of 7 kg and median age of 8 months, compared to 10 kg and 21 months for those without UAO. This may have resulted in smaller absolute endotracheal tube sizes for patients with Definite UAO or Possible/Probable/Definite UAO. However, when correcting for age by normalizing to the difference between predicted endotracheal tube size from the modified Cole formula and the actual tube size, the groups were very similar, even after adjusting a half size down for cuffed endotracheal tubes using the Deakers modification^{3,4}. Of note, for all groups of patients the actual endotracheal tubes were on average smaller than those which would have been predicted (median 0.35 size smaller). The presence of an endotracheal tube air leak below 30 cm H₂O at extubation was similar amongst groups. Of the 11 patients intubated for UAO, 7 were deemed to have no UAO upon extubation. The presence of a leak at intubation was documented in only 21% of cases.

Discussion

We have demonstrated that amongst providers who routinely perform such evaluations, there is only fair inter-observer reliability in clinical assessments of upper airway obstruction severity, using elements from the Westley croup scale after extubation. The most “reliable” elements, stridor and retractions, have only moderate absolute agreement, even when dichotomized into simply being present or absent. This variability results in different estimates for the incidence of post-extubation UAO and its associated risk for re-intubation. This also highlights a potential reason for the inconsistency in identifying risk factors for post extubation UAO.

Inconsistency in the assessment of a clinical outcome such as post extubation UAO may in part explain why few therapeutic studies on post-extubation stridor have demonstrated a benefit^{2,5,13,14}. Directly meaningful outcomes for therapeutic studies, such as re-intubation secondary to UAO, are relatively rare when applied to the global PICU population (4 or 5 of the 112 extubations in our study would have met these criteria (3.6–4.5%)). For this reason, “high risk” populations need to be identified and targeted for therapeutic interventions. However, without objective and reproducible outcomes, risk factors also can not be identified¹⁵. We sought to describe some potential risk factors in our study. From our data, patients with UAO had lower median age and weight than those without UAO, which is only marginally consistent with other studies^{3,4,6,16,17}. However, this was not true when looking at only cases of “Definite” UAO. It may be that our clinical assessment of UAO is less reliable for younger patients, and age may not actually be a risk factor for the development of post-extubation UAO. In our study, the presence of an air leak at extubation was similar regardless of the definition of post extubation UAO, echoing previous investigations^{18,19}. Nevertheless, this factor alone is used in clinical decision making for extubation readiness or prophylactic corticosteroids by close to 70% of US pediatric critical care training programs⁹. It may be that leak at intubation is a more appropriate predictor of successful extubation; although it was too infrequently documented in our study to make any conclusions. Previous studies have demonstrated that neurologic impairment²⁰ may have an impact on post extubation UOA, although we did not see a difference in pre-extubation Glasgow Coma Score amongst groups. We acknowledge that GCS was not intended to be used in this fashion, but is a commonly used clinical measure of level of responsiveness. We did not evaluate cuff-leak volume, which has been used in numerous adult studies to define high risk populations^{21–25} for which multiple doses of intravenous corticosteroids prior to extubation are beneficial. Because cuff-leak volume is not routinely used in most pediatric ICUs, and has not specifically been validated in children, we elected not to include it in this

analysis. However, as this analysis highlights, we must first have reproducible objective metrics to evaluate UAO before we can evaluate the true effects of potential risk factors such as cuff-leak volume.

From our data, it appears as if many of the elements of the Westley score for croup are too unreliable to consider for inclusion in UAO scoring, without rigorous common training. Cyanosis and hypoxemia were particularly problematic. It may be that this information is better captured with a combination of degree of oxygen support and pulse oximetry, although this needs to be studied. Interestingly, even air entry and level of consciousness, elements which clinicians assess multiple times per day as part of routine care for ICU patients, have significant variability. The simple combination of stridor and retractions appears to have the most internal validity. However, it is unclear if this simple combination will have enough external validity to measure the severity of UAO. This needs to be tested.

This analysis has examined the use case of post-extubation UAO. However, inter-observer variability plays a role in many other areas of critical care. In truth, there are few completely objective outcomes in critical care: namely direct physiologic changes that can be measured (such as a change in cardiac output, blood pressure etc, which still have some variability), or indisputable outcomes such as death. However, most clinical trials use considerably more subjective outcomes. For example, Ventilator Free Days, the primary outcome of most trials on ventilated patients in pediatric critical care, depend on highly variable ventilator, sedation, and fluid management practices (both at the intra- and inter intuition level), as well as the subjective and variable decision about when to actually remove the endotracheal tube, or how or when to use non-invasive ventilation after extubation. The same holds for studies which use ICU length of stay as an outcome, which is subject to a variety of institutional and individual practitioner variation. This is an important issue that should be considered when interpreting clinical research studies, and evaluating their generalizability to other populations.

There are several limitations to our analysis. Our primary objective was to determine the inter-observer variability amongst clinical assessments of UAO severity. It was not powered to evaluate risk factors for post-extubation UAO. Therefore, inferences regarding risk factors and outcomes are meant to be descriptive and highlight the variability based upon the uncertainty of the outcome. The sample size would need to be significantly larger to make meaningful inferences regarding these potential risk factors. In addition, several study-specific stridor scores have been developed, with reports of high internal validity through rigorous common training^{2,26}. We provided only brief training to the bedside providers before stridor assessments. It may be possible to attain higher levels of reliability with more rigorous common training. However, these are the providers who are routinely making assessments for clinical decision-making regarding therapies such as re-intubation or inhaled alpha agonists. Therefore, we sought to evaluate the “real-life” scenario to speak to the potential generalizability of applying such a scoring tool to routine practice. In addition, we did not analyze or record the specific level of training of providers, nor did we have physicians with different levels of training perform simultaneous assessments on patients. Therefore, we can make no conclusions regarding how the level of training affects the reliability of assessments between providers. Finally, several data elements (particularly air leak immediately following intubation and previous use of corticosteroids) were incomplete, despite the prospective nature of this study. Unfortunately, this is a limitation of medical documentation, since study personnel were not present at intubation, and providers inconsistently document this finding, or these medications, particularly for patients intubated at outside institutions.

Even if we can improve the internal validity of UAO scoring tools through rigorous common training, or simplification of the elements in the score, it is still unclear if these clinical assessments actually capture UAO severity. For this reason, we need to perform clinical investigations evaluating the most internally consistent clinical stridor scoring system against objective, physiology-based assessments of upper airway obstruction. These include combination measurements of flow (through spirometry or respiratory inductance plethysmography), and intrathoracic pressure (through esophageal manometry) to demonstrate inspiratory flow-limitation^{27–29}.

Conclusions

We have demonstrated that physical findings routinely used to assess upper airway obstruction have only fair to moderate inter-observer reliability amongst bedside providers. This variability may contribute to inconsistent findings regarding incidence, risk factors, and therapies for post-extubation upper airway obstruction. We need more objective tools to measure pediatric upper airway obstruction before making inferences on risk factors, or potentially beneficial therapies or preventative measures.

References

- Westley CR, Cotton EK, Brooks JG. Nebulized racemic epinephrine by IPPB for the treatment of croup: a double-blind study. *American Journal of Diseases of Children*. 1978; 132:484–487. [PubMed: 347921]
- Tellez DW, Galvis AG, Storgion SA, et al. Dexamethasone in the prevention of postextubation stridor in children. *Journal of Pediatrics*. 1991 Feb; 118(2):289–94. [PubMed: 1993963]
- Deakers TW, Reynolds G, Stretton M, et al. Cuffed endotracheal tubes in pediatric intensive care. *Journal of Pediatrics*. 1994; 125:57–62. [PubMed: 8021785]
- Newth, CJ; Rachman, B.; Patel, N., et al. The use of cuffed versus uncuffed endotracheal tubes in pediatric intensive care. *Journal of Pediatrics*. 2004; 144:333–337. [PubMed: 15001938]
- Anene O, Meert KL, Uy H, et al. Dexamethasone for the prevention of postextubation airway obstruction: a prospective, randomized, double-blind, placebo-controlled trial. *Critical Care Medicine*. 1996; 24:1666–1669. [PubMed: 8874303]
- Kurachek SC, Newth CJ, Quasney MW, et al. Extubation failure in pediatric intensive care: a multiple-center study of risk factors and outcomes. *Critical Care Medicine*. 2003; 31:2657–2664. [PubMed: 14605539]
- Edmunds S, Weiss I, Harrison R. Extubation failure in a large pediatric ICU population. *Chest*. 2001; 119:897–900. [PubMed: 11243974]
- Cole F. Pediatric formulas for the anesthesiologist. *A M A Journal of Diseases of Children*. 1957; 94:672–673.
- Foland JA, Super DM, Dahdah NS, et al. The use of the air leak test and corticosteroids in intubated children: a survey of pediatric critical care fellowship directors. *Respiratory Care*. 2002; 47:662–666. [PubMed: 12036435]
- Reichenheim M. Confidence intervals for the kappa statistic. *The Stata Journal*. 2004; 4:421–428.
- Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.
- Viera AJ, Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*. 2005; 37:360–363. [PubMed: 15883903]
- Khemani RG, Randolph A, Markovitz B. Corticosteroids for the prevention and treatment of post-extubation stridor in neonates, children and adults. *Cochrane Database of Systematic Reviews*. 2009:CD001000. [PubMed: 19588321]
- Khemani RG, Randolph A, Markovitz B. Steroids for post extubation stridor: pediatric evidence is still inconclusive. *Intensive Care Medicine*. 2010; 36:1276–1277. [PubMed: 20237767]

15. Khemani RG. Post extubation stridor the call for objectivity. *Indian Pediatrics*. 2010; 47:307–308. [PubMed: 20431158]
16. Koka BV, Jeon IS, Andre JM, et al. Postintubation croup in children. *Anesthesia & Analgesia*. 1977; 56:501–505. [PubMed: 560135]
17. Arens R, McDonough JM, Corbin AM, et al. Linear dimensions of the upper airway structure during development: assessment by magnetic resonance imaging. *American Journal of Respiratory & Critical Care Medicine*. 2002; 165:117–122. [PubMed: 11779740]
18. Mhanna MJ, Zamel YB, Tichy CM, et al. The “air leak” test around the endotracheal tube, as a predictor of postextubation stridor, is age dependent in children. *Critical Care Medicine*. 2002; 30:2639–2643. [PubMed: 12483052]
19. Wratney AT, Benjamin DK Jr, Slonim AD, et al. The endotracheal tube air leak test does not predict extubation outcome in critically ill pediatric patients. *Pediatric Critical Care Medicine*. 2008; 9:490–496. [PubMed: 18679147]
20. Harel Y, Vardi A, Quigley R, et al. Extubation failure due to post-extubation stridor is better correlated with neurologic impairment than with upper airway lesions in critically ill pediatric patients. *International Journal of Pediatric Otorhinolaryngology*. 1997; 39:147–158. [PubMed: 9104623]
21. Chung Y-H, Chao T-Y, Chiu C-T, et al. The cuff-leak test is a simple tool to verify severe laryngeal edema in patients undergoing long-term mechanical ventilation. *Critical Care Medicine*. 2006; 34:409–414. [PubMed: 16424722]
22. Cheng KC, Hou CC, Huang HC, et al. Intravenous injection of methylprednisolone reduces the incidence of postextubation stridor in intensive care unit patients. *Critical Care Medicine*. 2006 May; 34(5):1345–50. [PubMed: 16540947]
23. De Bast Y, De Backer D, Moraine J-J, et al. The cuff leak test to predict failure of tracheal extubation for laryngeal edema. *Intensive Care Medicine*. 2002; 28:1267–1272. [PubMed: 12209275]
24. Miller RL, Cole RP. Association between reduced cuff leak volume and postextubation stridor. *Chest*. 1996; 110:1035–1040. [PubMed: 8874265]
25. Jaber S, Chanques G, Matecki S, et al. Post-extubation stridor in intensive care unit patients. Risk factors evaluation and importance of the cuff-leak test. *Intensive Care Medicine*. 2003; 29:69–74. [PubMed: 12528025]
26. Johnson DW, Jacobson S, Edney PC, et al. A comparison of nebulized budesonide, intramuscular dexamethasone, and placebo for moderately severe croup. *New England Journal of Medicine*. 1998; 339:498–503. [PubMed: 9709042]
27. Argent AC, Hatherill M, Newth CJ, et al. The effect of epinephrine by nebulization on measures of airway obstruction in patients with acute severe croup. *Intensive Care Medicine*. 2008; 34:138–147. [PubMed: 17912497]
28. Argent AC, Newth CJ, Klein M. The mechanics of breathing in children with acute severe croup. *Intensive Care Medicine*. 2008; 34:324–332. [PubMed: 18008061]
29. Kaplan V, Zhang JN, Russi EW, et al. Detection of inspiratory flow limitation during sleep by computer assisted respiratory inductive plethysmography. *European Respiratory Journal*. 2000; 15:570–578. [PubMed: 10759455]

Table 1

Elements of the scoring system for UAO, based on the original Westley score breakdown.

Stridor
None
With Agitation
At Rest
Retractions
None
Mild
Moderate
Severe
Cyanosis
None
With Agitation
At Rest
Level of Consciousness
Normal (including sleep)
Disoriented
Air Entry
Normal
Decreased
Severely Decreased
Hypoxemia with Agitation *
None
Present
Hypoxemia at Rest *
None
Present
Pulsus Paradoxus (Any) *
None
Present

* Additional Items, not part of the original Westley Score

Table 2

Total number of cases and percent agreement for abnormal findings and inter-observer variability (kappa statistic) and 95% confidence interval for each of the elements for UAO scoring. Data are presented as agreement between all 3 providers (MD, RN, RT), as well as individual breakdowns between classes of providers. For elements with more than 2 possible grades, additional analysis is presented with the values dichotomized (Absent (score =0), Present (score >0)).

Element of Score	Cases (%) agreement on highest grade* N=220	Kappa Value (95% CI)	Cases (%) agreement on abnormal (present) N=220	Kappa Value (95% CI)
Stridor	3 grades		2 grades	
All	6 (2.7%)	0.43 (0.34,0.53)	18 (8.2%)	0.54 (0.42,0.66)
MD to RN	7 (3.2%)	0.44 (0.31,0.57)	24 (10.9%)	0.54 (0.40,0.69)
MD to RT	11 (5%)	0.44 (0.31,0.57)	24 (10.9%)	0.56 (0.41,0.70)
RN to RT	8 (3.6%)	0.42 (0.30,0.55)	23 (10.5%)	0.53 (0.39,0.68)
Retractions	4 grades		2 grades	
All	9 (4.1%)	0.47 (0.39,0.55)	48 (21.8%)	0.53 (0.44,0.61)
MD to RN	11 (5%)	0.44 (0.33,0.54)	55 (25%)	0.49 (0.38,0.61)
MD to RT	11 (5%)	0.38 (0.27,0.49)	55 (25%)	0.43 (0.31,0.56)
RN to RT	12 (5.5%)	0.59 (0.49,0.69)	60 (27.3%)	0.65 (0.55,0.76)
Cyanosis	3 grades		2 grades	
All	0 (0%)	0.01 (-0.04,0.11)	0 (0%)	0.00 (-0.05, 0.09)
MD to RN	0 (0%)	-0.02 (-0.04, -0.01)	0 (0%)	-0.04 (-0.07, -0.01)
MD to RT	0 (0%)	-0.03 (-0.04, -0.01)	0 (0%)	-0.04 (-0.07, -0.01)
RN to RT	1 (0.5%)	0.07 (-0.04,0.30)	1 (0.5%)	0.06 (-0.13,0.24)
Consciousness	2 grades			
All	2 (0.9%)	0.27 (0.03,0.49)		
MD to RN	2 (0.9%)	0.18 (-0.07,0.43)		
MD to RT	5 (2.3%)	0.54 (0.26,0.82)		
RN to RT	2 (0.9%)	0.13 (-0.09,0.35)		
Air Entry	3 grades		2 grades	
All	13 (5.9%)	0.32 (0.21,0.46)	14 (6.4%)	0.33 (0.20,0.46)
MD to RN	17 (7.7%)	0.26 (0.11,0.41)	18 (8.2%)	0.28 (0.13,0.42)
MD to RT	23 (10.5%)	0.40 (0.26,0.55)	24 (10.9%)	0.42 (0.27,0.56)

Element of Score	Cases (%) agreement on highest grade* N=220	Kappa Value (95% CI)	Cases (%) agreement on abnormal (present) N=220	Kappa Value (95% CI)
RN to RT	16 (7.3%)	0.31 (0.15,0.47)	16 (7.3%)	0.31 (0.15,0.47)
Hypoxemia with Agitation	2 grades			
All	2 (0.9%)	0.27 (0.12,0.45)		
MD to RN	5 (2.3%)	0.27 (0.05,0.49)		
MD to RT	3 (1.4%)	0.14 (−0.06,0.34)		
RN to RT	6 (2.7%)	0.41 (0.16,0.65)		
Hypoxemia at Rest	2 grades			
All	0 (0%)	0.14 (−0.01,0.29)		
MD to RN	0 (0%)	−0.02 (−0.04,0.01)		
MD to RT	1 (0.5%)	0.10 (−0.12,0.32)		
RN to RT	4 (1.8%)	0.66 (0.22,1.00)		
Pulsus Paradoxus	2 grades			
All	2 (0.9%)	0.23 (0.10,0.39)		
MD to RN	9 (4.1%)	0.33 (0.14,0.53)		
MD to RT	3 (1.4%)	0.13 (−0.05,0.32)		
RN to RT	4 (1.8%)	0.19 (−0.01,0.39)		

* Highest category for each data element with more than 2 possible grades (stridor at rest, moderate retractions (moderate was used because there was only 1 case of severe retractions), cyanosis at rest, decreased air entry (severely decreased was not used because there was only 1 case). Interpretation of Kappa Statistic: <0 Less than chance agreement; 0.01–0.20 Poor/Slight Agreement; 0.21–0.40 Fair Agreement; 0.41–0.60 Moderate Agreement; 0.61–0.80 Substantial Agreement; 0.81–0.99 Almost Perfect Agreement.

Table 3

Classification of the individual element of “Stridor” for the 220 simultaneous assessments, stratified by agreement amongst classes of providers (MD to RN, RT to RN, MD to RT). The left hand side grades stridor with the three possible classes described in the Westley score, while the right hand side breaks stridor severity into present (score >0) or none (score =0). Note that although there are many cases when providers will agree that there is no stridor (i.e. MD none, RN none = 167), there were more cases when one provider thought there was stridor and the other did not (MD present, RN none = 15 and RN present, MD none =14), as there were cases when they both agreed there was stridor (MD present, RN present = 24 cases)

3 grades of stridor			2 grades of stridor		
MD			MD		
Stridor	0	1 2	Stridor	None	Present
RN	0	167	3 12	167	15
	1	12	8 6	14	24
	2	2 3 7			
RT			RT		
Stridor	0	1 2	Stridor	None	Present
RN	0	168	7 7	168	14
	1	12	6 8	15	23
	2	3 1 8			
MD			MD		
Stridor	0	1 2	Stridor	None	Present
RT	0	168	4 11	168	15
	1	7	4 3	13	24
	2	6 6 11			

Table 4

Demographics, pre-extubation factors and outcomes for all patients, and patients categorized based on the probability of upper airway obstruction.

	All Patients N=112	No UAO N=87	Def UAO N=8	Prob/Def UAO N=17	Poss/Prob/Def UAO N=25
	N (%) or Median (IQR)	N (%) or Median (IQR)	N (%) or Median (IQR)	N (%) or Median (IQR)	N (or Median)
Endotracheal Tube Factors					
ETT Size	4 (3,75,5)	4,5 (4,5,5)	3,75 (3,5,4,0)	4 (3,5,4,5)	4 (3,5,4,5)
Cole ETT Diff*	-0,35 (-0,70,-0,10)	-0,30 (-0,90,-0,10)	-0,35(-0,8,0)	-0,2 (-0,6,0)	-0,4 (-0,6,0)
Cuffed ETT (y)	50 (45)	46 (53)	2 (25)	3 (18)	4 (16)
Cuff Inflated (y)	36 (72)	33 (72)	1 (50)	2 (67)	3 (75)
Leak Intubation (y)	21 (88)	17 (89)	2 (100)	2 (100)	4 (80)
Missing	88 (79)	68 (78)	6 (75)	15 (88)	20 (80)
Leak Extubation (y)	76 (69)	60 (70)	7 (100)	10 (63)	16 (67)
Missing	2 (2)	1 (1)	1 (13)	1 (6)	1 (4)
Reason for intubation					
Neurologic	14 (13)	11 (13)	1 (13)	3 (18)	3 (12)
Cardiac Surgery	64 (57)	51 (59)	2 (25)	7 (41)	13 (52)
UAO	11 (10)	7 (8)	3 (38)	3 (18)	4 (16)
Extrathoracic	10 (9)	9 (10)	0 (0)	1 (6)	1 (4)
Parenchyma	11 (10)	7 (8)	2 (25)	3 (18)	1 (4)
Other	2 (2)	2 (2)	0 (0)	0 (0)	0 (0)
Demographics					
Age (mo)	13,5 (3,5,121)	21 (4, 140)	8,5 (3, 18,5)	12 (4,23)	8 (3,23)
Weight	8,5 (5,5,20,5)	10 (6,29)	7 (5,5, 9,5)	8 (6,11)	7 (4,11)

	All Patients N=112	No UAO N=87	Def UAO N=8	Prob/Def UAO N=17	Poss/Prob/Def UAO N=25
	N (%) or Median (IQR)	N (%) or Median (IQR)	N (%) or Median (IQR)	N (%) or Median (IQR)	N (or Median)
Negative Inspiratory Force	30 (25, 40)	31 (25, 40)	37.5 (30,42)	35 (30,60)	30 (25,42)
Glasgow Coma Scale	11 (11,13.5)	11 (10,11)	11 (11,11)	11 (11,14)	11 (11,14)
Steroids Pre Extubation (Y)	21 (30)	16 (30)	1 (25)	3 (25)	5 (31)
Missing	43 (38)	34 (39)	4 (50)	5 (29)	9 (36)
Findings on minimal Level of Support prior to Extubation					
Extubated from CPAP	69 (61)	57 (66)	4 (50)	9 (53)	12 (48)
Extubated from PS/CPAP	43 (39)	39 (45)	4 (50)	8 (47)	13 (52)
PS level for extubation (n=43)	10 (5,10)	10 (5,10)	9 (6.5,10)	6.5 (5,10)	8 (5,10)
Fraction Inspired Oxygen (FiO ₂)	0.35 (0.35,0.4)	0.35 (0.3,0.4)	0.3 (0.3,0.4)	0.3 (0.3,0.4)	0.35 (0.3,0.4)
PEEP (cmH ₂ O)	5 (4,5)	5 (4,5)	5 (4,5)	5 (4,5)	5 (4,5)
RR (breaths/minute)	28 (20,36)	28 (18,35)	33.5 (27,42)	28 (22,35)	33 (23,42)
HR (beats/minute)	128 (114,149)	127 (113, 149)	141 (134,165)	135 (122, 142)	135 (122, 154)
SPO ₂ (%)	99 (95.5,100)	99 (96, 100)	99.5 (98.5,100)	99 (97, 100)	98 (92, 100)
VT (ml/kg)	8 (6,10)	8 (6,10)	7.5 (5.5,9.5)	8 (7, 9)	8 (7,9)
Nasal Flaring	3 (3)	1 (1)	0 (0)	2 (12)	2 (8)
Retractions	7 (6)	3 (3)	1 (13)	2 (12)	4 (16)
Outcomes within 24 hours of Extubation					
Re-Intubated	7 (6)	2 (2)	4 (50)	4 (24)	5 (20)
Non Invasive Ventilation	30 (27)	23 (26)	1 (13)	4 (24)	7 (28)
Racemic Epinephrine	12 (11)	2 (2)	6 (75)	8 (47)	10 (40)

	All Patients N=112	No UAO N=87	Def UAO N=8	Prob/Def UAO N=17	Poss/Prob/Def UAO N=25
	N (%) or Median (IQR)	N (%) or Median (IQR)	N (%) or Median (IQR)	N (%) or Median (IQR)	N (or Median)
Heliox	1 (1)	0 (0)	0 (0)	1 (6)	1 (4)
Nebulized Epinephrine **	1 (1)	0 (0)	0 (0)	0 (0)	1 (4)
Steroids After Extubation	14 (13)	10 (12)	2 (25)	2 (12)	4 (16)

[†] p values compare each category of UAO probability to those with no upper airway obstruction.

^{*} This value represents the actual endotracheal tube size minus that predicted by the Cole formula, with adjustment a half size down for cuffed endotracheal tubes. NA For cells with many missing data elements or zero values, p values were not calculated.

^{**} Nebulized Epinephrine refers to epinephrine administered continuously through a high flow humidified nasal cannula device.

Abbreviations: ETT = Endotracheal Tube, CPAP= Continuous Positive Airway Pressure, PS= Pressure Support, PEEP= Positive End Expiratory Pressure, RR= Respiratory Rate, HR= Heart Rate, VT= Spontaneous Tidal Volume in ml/kg. Def=Definite UAO, Prob=Probable UAO, Poss=Possible UAO