Corresponding author:
Hyung-Chul Lee, M.D., Ph.D.
Department of Anesthesiology and Pain
Medicine, Seoul National University Hospital,
Seoul National University College of Medicine,
101 Daehak-ro, Jongno-gu, Seoul 03080,
Korea
Tel: +82-2-2072-0723
Fax: +82-2-747-8363
Email: vital@snu.ac.kr
ORCID: https://orcid.org/0000-0003-0048-7958

Predicting optimal endotracheal tube size and depth in pediatric patients using demographic data and machine learning techniques

Hyeonsik Kim[1], Hyun-Kyu Yoon[1], Hyeonhoon Lee[1,2], Chul-Woo Jung[1], Hyung-Chul Lee[1,3]

[1]Department of Anesthesiology and Pain Medicine, Seoul National University College of Medicine, Seoul National University Hospital, [2]Biomedical Research Institute, Seoul National University Hospital, [3]Department of Data Science Research, Innovative Medical Technology Research Institute, Seoul National University Hospital, Seoul, Korea

Background: Use of endotracheal tubes (ETTs) with appropriate size and depth can help minimize intubation-related complications in pediatric patients. Existing age-based formulae for selecting the optimal ETT size present several inaccuracies. We developed a machine learning model that predicts the optimal size and depth of ETTs in pediatric patients using demographic data, enabling clinical applications.

Methods: Data from 37,057 patients younger than 12 years who underwent general anesthesia with endotracheal intubation were retrospectively analyzed. Gradient boosted regression tree (GBRT) model was developed and compared with traditional age-based formulae.

Results: The GBRT model demonstrated the highest macro-averaged F1 scores of 0.502 (95% CI [0.486, 0.568]) and 0.669 (95% CI [0.640, 0.694]) for predicting the uncuffed and cuffed ETT size (internal diameter), outperforming the age-based formulae that yielded 0.163 (95% CI [0.140, 0.196], P < 0.001) and 0.392 (95% CI [0.378, 0.406], P < 0.001), respectively. In predicting the ETT depth (distance from tip to lip corner), the GBRT model showed the lowest mean absolute error of 0.71 cm (95% CI [0.69, 0.72]) and 0.72 cm (95% CI [0.70, 0.74]) compared to the age-based formulae that showed an error of 1.18 cm (95% CI [1.16, 1.20], P < 0.001) and 1.34 cm (95% CI [1.31, 1.38], P < 0.001) for uncuffed and cuffed ETT, respectively.

Conclusions: The GBRT model using only demographic data accurately predicted the ETT size and depth. If these results are validated, the model may be practical for predicting optimal ETT size and depth for pediatric patients.

Keywords: Airway management; Demography; General anesthesia; Intratracheal intubation; Machine learning; Pediatrics.

Introduction

Selecting an appropriate size and depth of the endotracheal tube (ETT) is essential to minimize intubation-related complications in pediatric patients. An improper ETT size may require reintubation, increasing the risk of airway injury and prolonged apnea [1–3]. Moreover, inaccurate estimation of tube depth can cause bronchial intubation that can result in pneumothorax or atelectasis. By contrast, shallow insertion of an ETT can lead to an unsecured airway or inadequate ventilation [4].

Several methods have been proposed to select the optimal ETT size. Among those, Cole's age-based formula is typically used in clinical practice to estimate the internal diameter (ID) of uncuffed ETTs [5]. Other age-based formulae, such as those proposed by Khine et al. [6] and Duracher et al. [7], have been suggested for cuffed ETTs. The age-based formulae have also been used to estimate the optimal depth of ETT insertion [8]. However, several inaccuracies have been reported in these age-based formulae [9–11]. These inaccuracies might be because of the nonlinearity of tracheal growth with age. Another possible reason is inter-individual discrepancies in ETT size among individuals of the same age [12–14].

Machine learning algorithms handling complex nonlinear relationships have shown excellent performance in various medical fields [15]. However, few studies have integrated machine learning models to suggest the optimal ETT size and depth for pediatric patients [16]. Zhou et al. [16] implemented machine learning techniques with image-based features such as tracheal diameter at the C6, C7, and T2 levels or the distance from C6 to the tracheal carina. However, their model requires manual measurements by clinicians using X-ray images that are not usually available for pediatric surgical patients. By contrast, basic demographic data, such as age, sex, weight, and height, can be easily acquired from the most recent electronic health record system.

In this study, we aimed to develop and validate an explainable machine learning model to predict the optimal ETT size and depth for pediatric patients using only demographic data. Our hypothesis was that the machine learning model would outperform traditional age-based formulae in predicting the optimal ETT size and depth. A favorable model developed through this approach may be beneficial in routine anesthesia practice.

Materials and Methods

The Institutional Review Board of Seoul National University Hospital (Approval number: 2304-012-1418) approved this study and waived the requirement for informed consent owing to the retrospective nature of the study design. We followed the recommendations of the 'Strengthening the Reporting of Observational Studies in Epidemiology' guidelines [17].

Study population

Data were collected from 151,651 pediatric surgical patients who underwent general anesthesia with endotracheal intubation at Seoul National University Hospital from October 2004 to November 2022. Cases with the following characteristics were ex-

cluded: (1) age > 12 years; (2) specialized ETT type, such as right angle endotracheal, double lumen, and electromyogram tubes; (3) missing values for ETT type and size in the anesthesia note; and (4) surgical cases of second or subsequent surgeries for a single patient.

Data collection

Nursing and anesthesia notes were extracted from the hospital's clinical data warehouse. The most recent values of sex, height, and weight before surgery were extracted from the nursing notes. The ETTs utilized throughout the study period were Shiley™ Oral/Nasal Endotracheal Tube Cuffless Murphy Eye (Medtronic, Ireland) or Shiley™ Hi-Lo Oral/Nasal Tracheal Tube (Medtronic, Ireland). The type, size (ID), and fixed depth (distance from tip to lip corner) of the ETT were identified from the anesthesia notes.

A routine practice during the study at our hospital was selecting the ETT size based on Cole's formula, as decided by the attending anesthetists. If ventilation was inadequate owing to a leak, the patient was reintubated with a larger ETT. By contrast, if the tube size was large and did not advance within the trachea, a smaller size was retried. The optimal tube depth was determined by auscultation. After tracheal intubation, the ETT was introduced until the right upper lobe breath sounds disappeared. Subsequently, the tube was withdrawn until the upper lobe breath sounds reappeared. An additional length (1–2 cm) was retracted to prevent bronchial intubation by position change. Once fixed, the presence of breath sounds from both lung fields was reconfirmed, and the depth marker at the lip corner was recorded in the anesthesia note. The ID and depth of the ETT were recorded as 0.5 mm and 0.5 cm, respectively.

Model development

We developed regression models using gradient boosted regression tree (GBRT) and linear regression (LR) to predict the size and depth of the ETT separately. Due to the distinct rationale behind tube selection, we trained separate models to predict the size and depth of uncuffed and cuffed ETTs. Statistical outliers (± 2SD [standard deviation]) for height, weight, tube size, and depth within one-year intervals were considered as missing values. We performed multiple imputations to substitute the missing height and weight values.

The most recent 20% of the data was designated as the test dataset. The remaining data were assigned as the training dataset, separately for uncuffed and cuffed ETT types, to train the models. The test dataset was used to evaluate and compare the perfor-

mances with that of the traditional formulae. Subsequently, we used the BorutaSHAP method to select the necessary input variables from demographic data (age, sex, height, and weight) in the GBRT model. This method combines the Boruta feature selection algorithm with the Shapley value calculations [18]. After selecting the most relevant variables, they were incorporated into the final input of the machine learning models to predict the ID and fixed depth of the ETT. The hyperparameters for the GBRT model were determined using ten-fold cross-validation, and a grid search was performed for each combination of the hyperparameters. Supplementary Table 1 lists the hyperparameter combinations.

Outcome variables

The ETT size predicted by the models was rounded to the nearest 0.5 mm. The primary outcome for the size model was the macro-averaged F1 score that comprehensively evaluates the model's performance across all classes by calculating the unweighted mean value of the F1 score for each class. Additionally, we computed the accuracy of predicting the exact size and the size within 0.5 mm of the tube, given that clinicians typically prepare three sizes of ETTs in case of failure.

To compare the performance of our model in predicting the size of an ETT, we selected Cole's formula [5] for an uncuffed ETT (ID [mm] = age in years / 4 + 4.0) and Duracher's formula [7] for a cuffed ETT (ID [mm] = age in years / 4 + 3.5) as traditional age-based formulae. For below one year of age, an ID of 3.5 mm was used, and for between one and two years of age, 4.0 mm was used for the uncuffed ETT, as Cole's formula applies over the age of two. A size smaller by one was used for cuffed ETTs for ages less than two years. The Penlington's formula (ID [mm] = age in years / 4 + 4.5) was also used to estimate the uncuffed ETT size [19].

The primary outcome of the depth model was measured in terms of the mean absolute error (MAE). Additionally, root mean squared error (RMSE) and R-squared were calculated to evaluate the performance of the depth model. To calculate the depth of the ETT, we selected traditional age-based formulae based on the Pediatric Advanced Life Support (PALS) guidelines (recommended depth of insertion [cm] = age in years / 2 + 12) [8]. We compared the performance of the GBRT models with that of traditional age-based formulae and LR models.

The linearity assumptions in the relationships between ETT size and depth with age were tested by verifying the normality of the residual distributions at a significance level of 0.05. The scatter plots of these variables and those of the residuals and fitted values were depicted to verify the linear relationship.

We adopted the Shapley additive explanation (SHAP) method to enhance the interpretability of the machine learning model. This method calculates the contribution of the input variables to the prediction and quantifies how each variable affects the output of the machine learning model [20].

To enhance the limited intuitive understanding of machine learning outcomes, we constructed a table presenting predictions for tube size using the GBRT model. This table was created by referencing the pediatric growth chart offered by the Korea Disease Control and Prevention Agency [21]. We incorporated weight and height data corresponding to the 5th, 15th, 25th, 50th, 75th, 85th, and 95th percentiles for each age from the pediatric growth chart.

We have released our data, model parameters, and code in a public repository (https://github.com/Hyeonsik/ endotracheal_ tube.git) and developed a web-based calculator (https://tubesize. net) to validate and apply the results.

Subgroup analysis

We performed a subgroup analysis of our predictive model for ETT size according to age. The patient population was stratified into three distinct age groups: neonates (< 1 month), infants (< 1 year), and others (≥ 1 year). Subsequently, we assessed and compared the predictive performance with the trained GBRT model within these subgroups without retraining.

Statistical analysis

Continuous variables, such as age, weight, and height, are presented as means (standard deviation) or medians (Q1, Q3), depending on the results of the Shapiro–Wilk test. Categorical variables, such as sex and ETT type, are presented numerically (percentages). Model performances were computed with a 95% CI through bootstrapping methods, and ml-stat-util (https://github. com/mateuszbuda/ml-stat-util) was employed for conducting statistical tests. The Mood's median test was performed for model comparisons in the subgroup analysis. The Mann-Whitney U test or two-sample t-test was performed to compare continuous variables depending on the Shapiro–Wilk test results. For the comparison of categorical variables, the chi-square test was performed. Considering the two outcomes (size and depth) and two tube types (cuffed and uncuffed), a P value < 0.0125 was considered statistically significant after the Bonferroni correction.

A custom program was developed using Python® (Python Software Foundation, USA) with scikit-learn 1.0.2, XGBoost 1.7.3, Keras 2.7.0, SHAP 0.41.0, BorutaSHAP 1.1, and stat-util libraries,

to develop and validate the model.

Results

After excluding 114,594 patients, the final analysis included 37,057 surgical procedures (Fig. 1). The general characteristics of the data are summarized in Table 1. There were differences in age, height, weight, and the distribution of tube depth between training and test sets for both cuffed and uncuffed ETT data. The BorutaSHAP method was employed to identify significant input variables for the size and depth models, and the variable 'sex' was removed, except for the model predicting the depth of uncuffed ETTs, as they did not significantly affect the output (P < 0.05, Fig. 2). The results showed that age, weight, and height are critical factors in predicting ETT size and cuffed ETT depth. By contrast, age, sex, weight, and height are critical factors in predicting uncuffed ETT depth. Scatter plots depicting ETT size and depth by age and scatter plots depicting the residuals and fitted values are
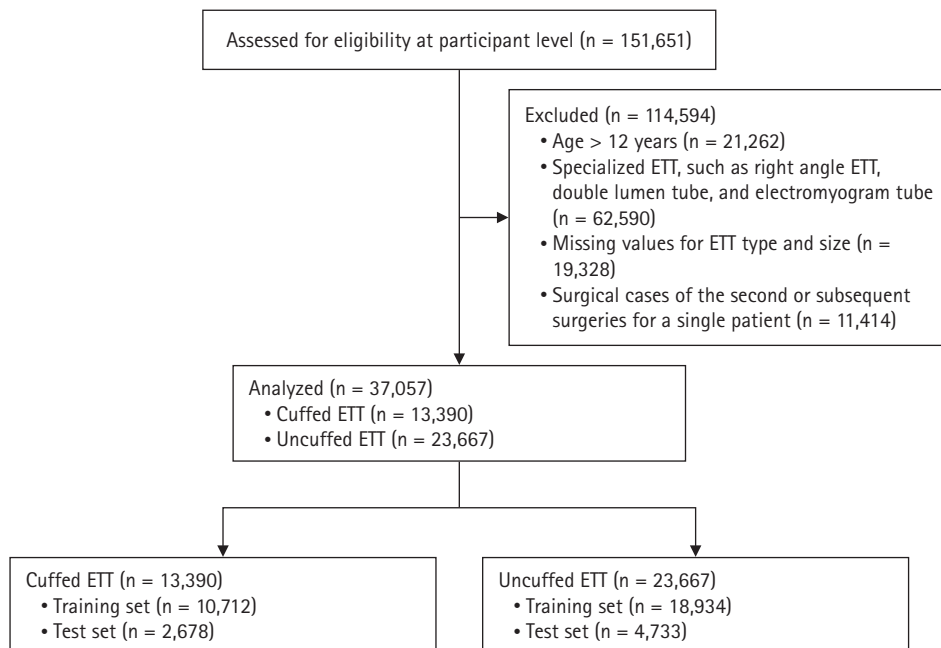
```
┌──────────────────────────────────────────────────────────┐
│ Assessed for eligibility at participant level (n = 151,651) │
└──────────────────────────────────────────────────────────┘
                    │
                    │        ┌────────────────────────────────────────┐
                    │───────▶│ Excluded (n = 114,594)                 │
                    │        │ • Age > 12 years (n = 21,262)          │
                    │        │ • Specialized ETT, such as right angle ETT, │
                    │        │   double lumen tube, and electromyogram tube │
                    │        │   (n = 62,590)                         │
                    │        │ • Missing values for ETT type and size (n = │
                    │        │   19,328)                              │
                    │        │ • Surgical cases of the second or subsequent │
                    │        │   surgeries for a single patient (n = 11,414) │
                    │        └────────────────────────────────────────┘
                    │
       ┌────────────────────────────────┐
       │ Analyzed (n = 37,057)          │
       │ • Cuffed ETT (n = 13,390)      │
       │ • Uncuffed ETT (n = 23,667)    │
       └────────────────────────────────┘
          │                        │
┌──────────────────────────┐  ┌──────────────────────────┐
│ Cuffed ETT (n = 13,390)  │  │ Uncuffed ETT (n = 23,667)│
│ • Training set (n = 10,712) │  │ • Training set (n = 18,934) │
│ • Test set (n = 2,678)   │  │ • Test set (n = 4,733)   │
└──────────────────────────┘  └──────────────────────────┘
```

Fig. 1. Study flowchart. ETT: endotracheal tube.

Table 1. Comparison of Demographic and Tube Data between Training and Test Datasets for Cuffed and Uncuffed ETTs in This Study

Variable	Missing (%)	Training dataset	Test dataset	P value
Uncuffed ETT		18,934 (80.0)	4,733 (20.0)	
Age (yr)	0	3.32 (0.88, 5.41)	2.94 (0.91, 4.75)	< 0.001
Sex (M)	0	11,122 (58.7)	2,735 (57.8)	0.239
Height (cm)	6.6	92.2 (74.2, 111.0)	90.1 (75.0, 107.0)	< 0.001
Weight (kg)	2.7	14.6 (9.0, 19.0)	13.7 (9.2, 17.7)	< 0.001
ID of ETT (mm)	0	4.8 (4.0, 5.5)	4.8 (4.0, 5.5)	0.867
Fixed depth (cm)	10.9	13.6 (12.0, 15.5)	13.2 (11.5, 15.0)	< 0.001
Cuffed ETT		10,712 (80.0)	2,678 (20.0)	
Age (yr)	0	7.28 (3.22, 10.9)	4.36 (0.539, 7.56)	< 0.001
Sex (M)	0	6,403 (59.8)	1,528 (57.1)	0.011
Height (cm)	3.9	120.0 (96.0, 144.1)	97.6 (67.0, 125.4)	< 0.001
Weight (kg)	2.9	28.4 (14.5, 39.5)	18.6 (7.5, 25.4)	< 0.001
ID of ETT (mm)	0	5.3 (4.5, 6.0)	4.5 (3.5, 5.5)	< 0.001
Fixed depth (cm)	9.6	16.2 (14.0, 19.0)	14.1 (11.0, 17.0)	< 0.001

Values are presented as mean ± SD, median (Q1, Q3), or number (proportion). ETT: endotracheal tube, ID: internal diameter.
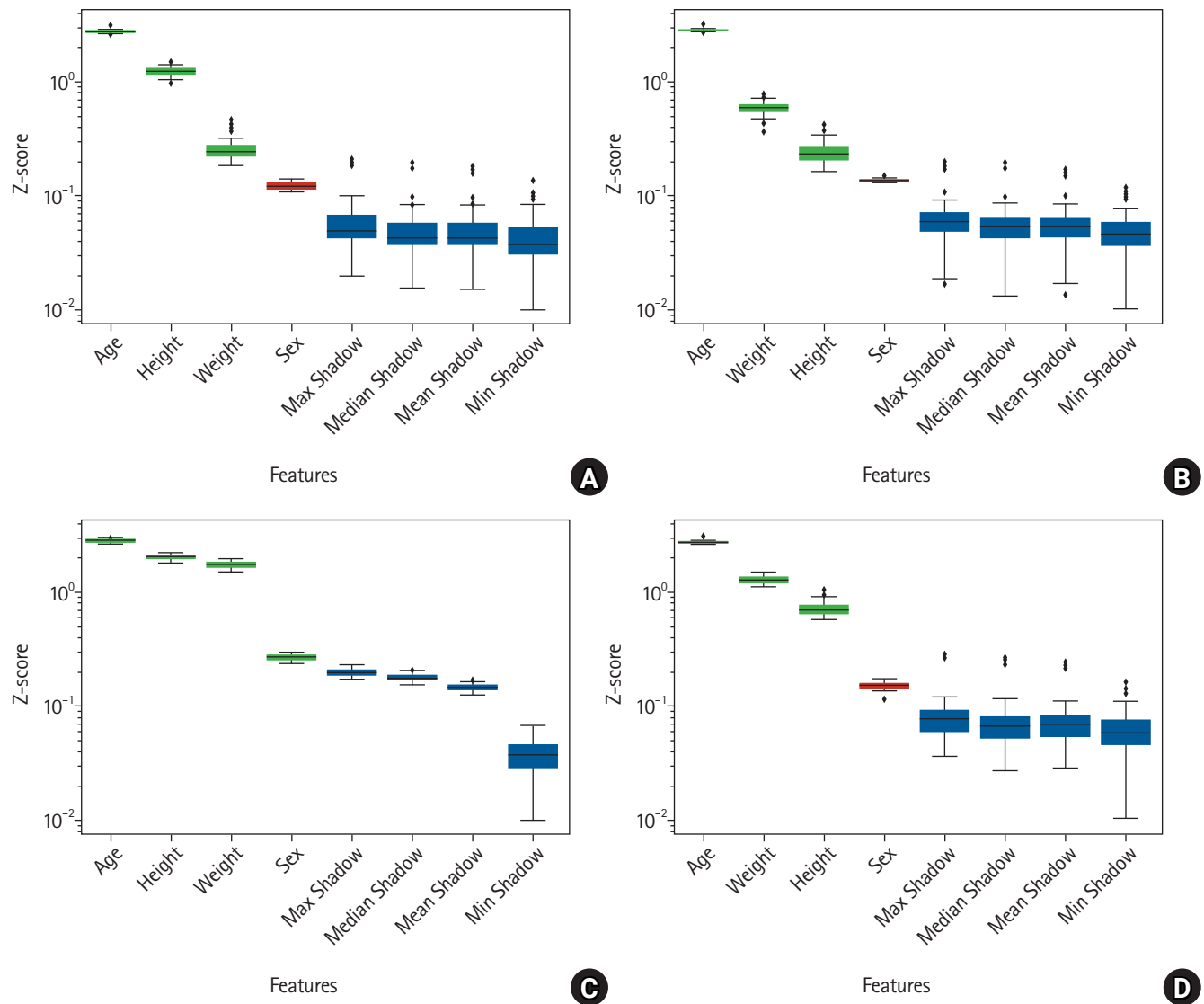
Fig. 2. Boxplot of the feature importance from input candidates using the BorutaSHAP method. (A) Boxplot of the feature importance from input candidates (age, sex, weight, height, and existence of cuff) for predicting uncuffed ETT size using the BorutaSHAP method. (B) Boxplot of the feature importance from input candidates (age, sex, weight, height, and existence of cuff) for predicting cuffed ETT size using the BorutaSHAP method. (C) Boxplot of the feature importance from input candidates (age, sex, weight, height, and existence of cuff) for predicting uncuffed ETT depth using the BorutaSHAP method. (D) Boxplot of the feature importance from input candidates (age, sex, weight, height, and existence of cuff) for predicting cuffed ETT depth using the BorutaSHAP method. X-axis presents the input features and Y-axis shows the Z-score of whether each feature has an importance significantly lower than the threshold. Features confirmed important are presented in green (P < 0.05) and blue colors, while red color represents unimportant features (P < 0.05). The term 'Shadow' on the X-axis refers to shadow features generated by randomly permuting the dataset of each original feature. Then, the feature importance are computed in the original and the generated shadow features. ETT: endotracheal tube.

shown in Fig. 3. The linearity assumption between ETT size and age was not achieved (P < 0.001).

The GBRT model showed the highest macro-averaged F1 score of 0.502 (95% CI [0.486, 0.568]) in predicting the size of uncuffed ETTs and 0.669 (95% CI [0.640, 0.694]) for cuffed ETTs. This performance was superior to that of traditional age-based formulae that achieved a macro-averaged F1 score of 0.163 (95% CI [0.140,

0.196], P < 0.001) for uncuffed ETTs and 0.392 (95% CI [0.378, 0.406], P < 0.001) for cuffed ETTs (Table 2).

The GBRT model achieved the best performance in predicting the ETT depth, with an MAE of 0.71 cm (95% CI [0.69, 0.72]) for uncuffed ETTs and 0.72 cm (95% CI [0.70, 0.75]) for cuffed ETTs. The GBRT model outperformed the traditional age-based formula (MAE for uncuffed ETTs = 1.18 cm [95% CI 1.16, 1.20], MAE
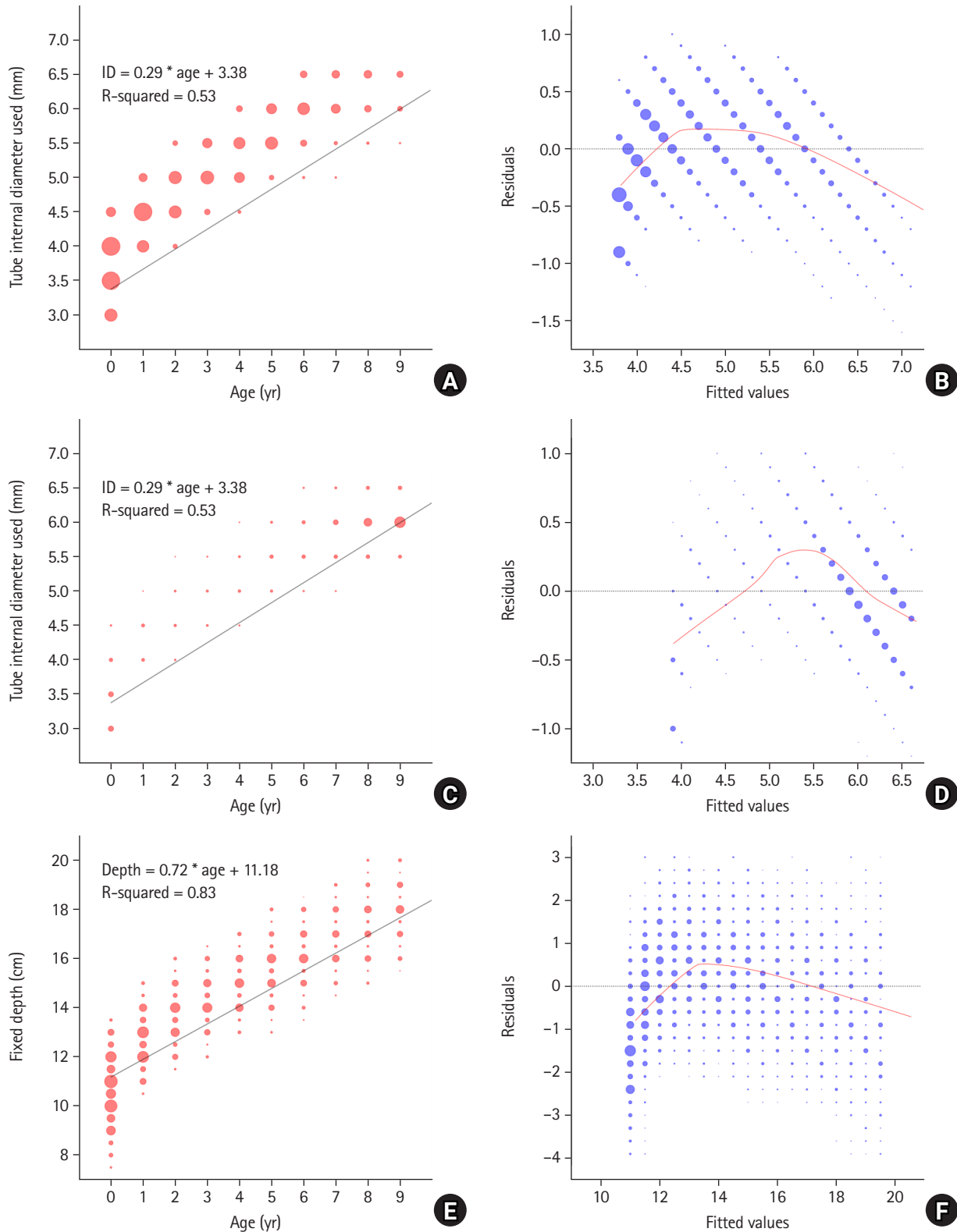
Fig. 3. Scatter plots and residuals analysis for ETT size and depth by age. (A) Scatter plot of uncuffed ETT size by age. (B) Scatter plot of residuals for LR analysis between uncuffed ETT size and age. X-axis presents residuals that indicate the difference between the observed and predicted ETT sizes. Y-axis presents the fitted values generated using a LR model. (C) Scatter plot of cuffed ETT size according to age. (D) Scatter plot of residuals and fitted values for uncuffed ETT size by age. (E) Scatter plot of ETT depth by age. (F) Scatter plot of residuals and fitted values for ETT depth according to age. The black line refers to the LR trend between two axes, and red line refers to a locally weighted scatterplot smoother fitted to the residual scatter plot. ETT: endotracheal tube, LR: linear regression.

Table 2. Performance of GBRT Model, MLR Model, and Age-based Formulae for Predicting the Size of ETT

Model	Macro-averaged F1	P value	Accuracy within 0.5 mm (%)	P value	Accuracy (%)	P value
Uncuffed ETT						
GBRT	0.502 (0.486, 0.568)	Reference	98.1 (97.8, 98.4)	Reference	58.2 (57.0, 59.4)	Reference
MLR	0.407 (0.395, 0.424)	< 0.001	97.2 (96.8, 97.6)	< 0.001	53.8 (52.5, 55.0)	< 0.001
Penlington's*	0.203 (0.196, 0.211)	< 0.001	82.6 (81.7, 83.5)	< 0.001	41.3 (40.2, 42.5)	< 0.001
Cole's†	0.163 (0.140, 0.196)	< 0.001	78.1 (77.1, 79.1)	< 0.001	20.3 (19.3, 21.2)	< 0.001
Cuffed ETT						
GBRT	0.669 (0.640, 0.694)	Reference	99.5 (99.3, 99.7)	Reference	70.1 (68.6, 71.5)	Reference
MLR	0.576 (0.551, 0.600)	< 0.001	99.4 (99.1, 99.6)	0.589	58.4 (56.8, 59.9)	< 0.001
Duracher's‡	0.392 (0.378, 0.406)	< 0.001	96.6 (96.0, 97.2)	< 0.001	46.9 (45.3, 48.5)	< 0.001

Values are presented as numbers (95% CI). GBRT: gradient boosted regression tree, MLR: multiple linear regression, ETT: endotracheal tube, ID: internal diameter. *Penlington's formula (ID of the uncuffed ETT [mm] = age in years / 4 + 4.5), †Cole's formula (ID of the uncuffed ETT [mm] = age in years / 4 + 4.0), ‡Duracher's formula (ID of the cuffed ETT [mm] = age in years / 4 + 3.5).

Table 3. Performance of GBRT, MLR Models, and Age-based Formula for Predicting the Depth of ETT

Model	MAE (cm)	P value	RMSE (cm)	P value	R-squared	P value
Uncuffed ETT						
GBRT	0.71 (0.69, 0.72)	Reference	0.88 (0.87, 0.90)	Reference	0.831 (0.823, 0.839)	Reference
MLR	0.74 (0.73, 0.76)	< 0.001	0.94 (0.92, 0.96)	< 0.001	0.803 (0.793, 0.812)	< 0.001
PALS*	1.18 (1.16, 1.20)	< 0.001	1.46 (1.44, 1.49)	< 0.001	0.572 (0.554, 0.589)	< 0.001
Cuffed ETT						
GBRT	0.72 (0.70, 0.74)	Reference	1.00 (0.91, 1.14)	Reference	0.904 (0.875, 0.921)	Reference
MLR	0.77 (0.75, 0.80)	< 0.001	1.05 (0.97, 1.20)	< 0.001	0.884 (0.852, 0.903)	< 0.001
PALS	1.34 (1.31, 1.38)	< 0.001	1.67 (1.61, 1.75)	< 0.001	0.720 (0.693, 0.740)	< 0.001

Values are presented as numbers (95% CI). GBRT: gradient boosted regression tree, MLR: multiple linear regression, ETT: endotracheal tube, MAE: mean absolute error, RMSE: root mean squared error, PALS: pediatric advanced life support. *PALS guideline (depth of insertion [cm] = age in years / 2 + 12).

for cuffed ETTs = 1.34 cm [95% CI 1.31, 1.38]). There was a significant performance difference between the GBRT model and the traditional age-based formula (P < 0.001) (Table 3).

In the subgroup analysis, the size model showed the highest macro-averaged F1 score in the infant group for uncuffed ETTs and the other groups for cuffed ETTs, while the other groups showed the lowest accuracy for both uncuffed and cuffed ETT sizing (Table 4).

The tube sizes and depths predicted by the GBRT model for the representative demographic values are presented in Supplementary Table 2.

The SHAP summary plot in Supplementary Fig. 1 illustrates the contribution of each input variable to the output of the GBRT model. Older age, uncuffed ETT, heavier weight, and taller height contributed to larger ETT size. Older age, heavier weight, taller height, and male sex were associated with deeper ETT depth. The SHAP dependence plots presented in Supplementary Fig. 2 and Supplementary Fig. 3 illustrate the effect of each input variable on

the prediction.

Discussion

In this study, we developed and validated machine learning models to predict the optimal ETT size and depth in pediatric patients. Our models used only demographic variables and considered the GBRT algorithm. The developed models outperformed the traditional age-based formulae.

Previous studies on optimal ETT size using age-based formulae have reported an accuracy in the range of 15%–50% in predicting the exact uncuffed or cuffed ETT size [9,10,14,16]. However, our model exhibited an accuracy of 58.2% and 70.1% for exact matching and 98.1% and 99.5% for an accuracy within 0.5 mm for uncuffed and cuffed ETTs, respectively. The differences in performance might be attributable to the use of machine learning algorithms that can model nonlinear relationships. The linearity test results and SHAP dependency plot in our study

Table 4. Subgroup Analyses based on Age for Predicting ETT Size using GBRT Model

	Macro-averaged F1	P value	Accuracy within 0.5 mm (%)	P value	Accuracy (%)	P value
Uncuffed ETT						
Neonate	0.371 (0.300, 0.467)	< 0.001	98.2 (96.3, 99.4)	< 0.001	65.0 (58.9, 71.2)	< 0.001
Infant	0.521 (0.500, 0.543)	< 0.001	98.6 (98.0, 99.2)	< 0.001	65.7 (63.3, 67.9)	< 0.001
Others	0.426 (0.389, 0.480)	Reference	97.9 (97.5, 98.3)	Reference	55.5 (54.1, 56.9)	Reference
Cuffed ETT						
Neonate	0.541 (0.459, 0.674)	< 0.001	100.0 (100.0, 100.0)	< 0.001	88.3 (83.0, 93.6)	< 0.001
Infant	0.510 (0.429, 0.632)	< 0.001	99.6 (99.3, 99.9)	< 0.001	82.4 (80.2, 84.6)	< 0.001
Others	0.626 (0.591, 0.657)	Reference	99.4 (99.0, 99.7)	Reference	63.1 (61.2, 65.0)	Reference

Values are presented as numbers (95% CI). ETT: endotracheal tube, GBRT: gradient boosted regression tree.

confirmed the nonlinear relationship between the size or depth of the ETT and age.

Other demographic variables, such as height and weight, also contributed significantly to improving the prediction of ETT size and depth. In the analysis based on the BorutaSHAP method, all variables, except for sex, were included in the GBRT model for predicting the ETT size. Therefore, adding these variables significantly improves model performance. These results are consistent with previous findings stating that there was no difference in terms of sex in developing the trachea throughout childhood [22]. Moreover, sex was only included in the GBRT model for predicting the depth of uncuffed ETTs. The uncuffed tube depth may be affected by sex owing to the difference in tongue size, as the ETT depth was measured at the lip corner.

In a previous study, Zhou et al. [16] developed machine learning models using demographic data and extracted features from the chest X-ray images of 990 patients to estimate the ETT size. The accuracies of their models were 57.5% and 52.3% for cuffed and uncuffed ETTs, respectively, whereas our model using only demographic data yielded accuracies of 70.1% and 58.2%, respectively. This difference can be attributed to the massive volume of data we used that was 25 times more than that used by Zhou et al.

Although Cole's formula has been used in clinical practice for several decades, several studies have reported that Penlington's formula is more accurate for predicting uncuffed ETT size [10,16]. Our study also found that Penlington's formula that suggests a larger ETT size was more accurate than Cole's formula in predicting uncuffed ETTs in pediatric patients. This difference in accuracy may be attributed to variations in the growth curve in pediatric populations over time and race since Cole's formula was first introduced in a North American pediatric population in 1957 [5]. Nevertheless, all age-based formulae investigated in this study were highly inaccurate compared to the machine learning models.

In our subgroup analysis, the accuracy of the 'others' group,

consisting of individuals aged one year or older, in predicting the ETT size was the lowest among the three age groups. This may be because the trachea size in the neonate and infant groups was relatively uniform compared to those in the other age groups. The difference in performance among the age groups also indicated a nonlinear relationship between age and tube size.

The strength of our model is its readiness in clinical situations because it is available as a web calculator, and its code is available online. In most electronic medical record systems, height and weight information is obtained before surgery. Additionally, according to the BorutaSHAP results obtained in this study, this additional information is significant. Therefore, a system implemented with the proposed model to provide automated suggestions could be practical for determining a more accurate ETT size and fixation depth in pediatric patients.

Our study has a few limitations. First, because our study was retrospective, there may be inevitable biases, and the excluded or missing data could have affected the results. Therefore, future prospective validation is needed to address these issues with minimal data loss. Second, the generalizability of our study may be limited because it was conducted for an Asian population at a single institute. The different patterns in clinical practices may influence the machine learning model's performance and limit its real-world applicability. Therefore, conducting external validation studies across multiple centers, encompassing diverse patient populations and clinical practices, is crucial to assess the robustness and reliability of the model's performance before the application. Third, we might have missed some important input variables, such as congenital diseases that may further affect airway anatomy and result in size depth variations of the ETT [23,24]. Fourth, different cuff designs, such as Hi-Contour or TaperGuard™ (Medtronic, Ireland), could result in variations in the optimal tube size and depth. Therefore, the models may require retraining before applying them to different tube types using the corresponding data for

each specific tube type. Fifth, although we utilized the minimal set of readily collectible demographic variables, additional input parameters, such as Mallampati classification or imaging data like X-rays and ultrasound images, can improve model performance. Sixth, the labeled ETT size and depth may not be optimal because there could be some tolerance for improper tube size and depth by the attending anesthetist based on auscultation. Additionally, there may be inaccuracies in the recorded tube depth because the fixed depth difference may be changed by the patient's position, especially in neonates and infants.

In conclusion, we developed and validated an explainable machine learning model to precisely estimate the size and depth of an ETT in pediatric patients using only basic demographic data. Prospective validation is warranted to validate our results before integration into clinical practice.

Funding

Conflicts of Interest

Hyung-Chul Lee was an Editor for the Korean Journal of Anesthesiology from 2020 to 2022. However, he was not involved in any process of review for this article, including peer reviewer selection, evaluation, or decision-making. There were no other potential conflicts of interest relevant to this article.

Data Availability

The datasets generated during and/or analyzed during the current study are available in the github repository (https://github.com/Hyeonsik/endotracheal_tube).

Author Contributions

Hyeonsik Kim (Data curation; Formal analysis; Software; Writing – original draft)
Hyun-Kyu Yoon (Conceptualization; Supervision)
Hyeonhoon Lee (Formal analysis; Software)
Chul-Woo Jung (Conceptualization; Writing – review & editing)
Hyung-Chul Lee (Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Supervision)

Supplementary Materials

Supplementary Table 1. Combinations of hyperparameters for the gradient boosted regression tree (GBRT) model predicting size and depth of endotracheal tube (ETT).

Supplementary Table 2. The predicted internal diameter (ID) and depth of the endotracheal tubes (ETTs) using the gradient boosted regression tree (GBRT) models based on the representative demographic values.

Supplementary Fig. 1. Shapley additive explanation summary plot for endotracheal tube (ETT) size and depth prediction by gradient boosted regression tree (GBRT) models. (A) Shapley additive explanation summary plot for input variables in the GBRT model for predicting the size of uncuffed ETTs. (B) Shapley additive explanation summary plot for input variables in the GBRT model for predicting the size of cuffed ETTs. (C) Shapley additive explanation summary plot for input variables in the GBRT model for predicting the depth of uncuffed ETTs. (D) Shapley additive explanation summary plot for input variables in the GBRT model for predicting the depth of cuffed ETTs. The red and blue dots represent the higher and lower values of the variables, respectively. Large Shapley values indicate a high contribution to output regardless of positive or negative. Older age, heavier weight, and taller height contribute to a larger size of the ETT. Older age, heavier weight, taller height, and male sex were associated with deeper ETT depth.

Supplementary Fig. 2. Shapley additive explanation dependence plot for each input variable in the gradient boosted regression tree (GBRT) model for predicting the size of uncuffed endotracheal tubes (ETTs): (A) age, (B) weight, and (C) height. Shapley additive explanation dependence plot for each input variable in the GBRT model for predicting the size of cuffed ETTs: (D) age, (E) weight, and (F) height. Effect of a feature on the model's output and the distribution of the feature's value is visualized as a scatter plot in the Shapley dependence plot. Horizontal axis represents the value of each feature, and the vertical axis represents the Shapley values of a feature. The light grey area at the base of the plot represents a histogram displaying the distribution of data values.

Supplementary Fig. 3. Shapley additive explanation dependence plot for each input variable in the gradient boosted regression tree (GBRT) model for predicting the depth of uncuffed ETTs: (A) age, (B) sex, (C) weight, and (D) height. Shapley additive explanation dependence plot for each input variable in the GBRT model for predicting the depth of cuffed ETTs: (E) age, (F) weight, and (G) height. Effect of a feature on the model's output and the distribution of the feature's value is visualized as a scatter plot in the Shapley dependence plot. The horizontal axis represents the value

of each feature, and the vertical axis represents the Shapley values of a feature. The light grey area at the base of the plot represents a histogram displaying the distribution of data values.

ORCID

Hyeonsik Kim, https://orcid.org/0009-0007-4657-5175
Hyun-Kyu Yoon, https://orcid.org/0000-0001-5424-3559
Hyeonhoon Lee, https://orcid.org/0000-0002-9426-823X
Chul-Woo Jung, https://orcid.org/0000-0001-7876-8659
Hyung-Chul Lee, https://orcid.org/0000-0003-0048-7958

References

1. Schweiger C, Manica D. Acute laryngeal lesions following endotracheal intubation: risk factors, classification and treatment. Semin Pediatr Surg 2021; 30: 151052.

2. Gálvez JA, Acquah S, Ahumada L, Cai L, Polanski M, Wu L, et al. Hypoxemia, bradycardia, and multiple laryngoscopy attempts during anesthetic induction in infants: a single-center, retrospective study. Anesthesiology 2019; 131: 830-9.

3. Patel R, Lenczyk M, Hannallah RS, McGill WA. Age and the onset of desaturation in apnoeic children. Can J Anaesth 1994; 41: 771-4.

4. Miller KA, Kimia A, Monuteaux MC, Nagler J. Factors associated with misplaced endotracheal tubes during intubation in pediatric patients. J Emerg Med 2016; 51: 9-18.

5. Cole F. Pediatric formulas for the anesthesiologist. AMA J Dis Child 1957; 94: 672-3.

6. Khine HH, Corddry DH, Kettrick RG, Martin TM, McCloskey JJ, Rose JB, et al. Comparison of cuffed and uncuffed endotracheal tubes in young children during general anesthesia. Anesthesiology 1997; 86: 627-31.

7. Duracher C, Schmautz E, Martinon C, Faivre J, Carli P, Orliaguet G. Evaluation of cuffed tracheal tube size predicted using the Khine formula in children. Paediatr Anaesth 2008; 18: 113-8.

8. Guidelines 2000 for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Part 10: pediatric advanced life support. The American Heart Association in collaboration with the International Liaison Committee on Resuscitation. Circulation 2000; 102(8 Suppl): I291-342.

9. Bae JY, Byon HJ, Han SS, Kim HS, Kim JT. Usefulness of ultrasound for selecting a correctly sized uncuffed tracheal tube for paediatric patients. Anaesthesia 2011; 66: 994-8.

10. Park HP, Hwang JW, Lee JH, Nahm FS, Park SH, Oh AY, et al. Predicting the appropriate uncuffed endotracheal tube size for children: a radiograph-based formula versus two age-based formulas. J Clin Anesth 2013; 25: 384-7.

11. Volsko TA, McNinch NL, Prough DS, Bigham MT. Adherence to Endotracheal tube depth guidelines and incidence of malposition in infants and children. Respir Care 2018; 63: 1111-7.

12. Wailoo MP, Emery JL. Normal growth and development of the trachea. Thorax 1982; 37: 584-7.

13. Fisk GC. Variation in sizes of endotracheal tubes for infants and young children. Anaesth Intensive Care 1973; 1: 418-22.

14. Ritchie-McLean S, Ferrier V, Clevenger B, Thomas M. Using middle finger length to determine the internal diameter of uncuffed tracheal tubes in paediatrics. Anaesthesia 2018; 73: 1207-13.

15. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019; 25: 44-56.

16. Zhou M, Xu WY, Xu S, Zang QL, Li Q, Tan L, et al. Prediction of endotracheal tube size in pediatric patients: Development and validation of machine learning models. Front Pediatr 2022; 10: 970646.

17. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008; 61: 344-9.

18. Keany E. A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values [Internet]. Geneva: Zenodo; 2020 Nov 5 [cited 2023 Jun 27]. Available from https://zenodo.org/record/4247618.

19. Penlington GN. Letter: Endotracheal tube sizes for children. Anaesthesia 1974; 29: 494-5.

20. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. California, Curran Associates. 2017, pp 4768-77.

21. Kim JH, Yun S, Hwang SS, Shim JO, Chae HW, Lee YJ, et al. The 2017 Korean National Growth Charts for children and adolescents: development, improvement, and prospects. Korean J Pediatr 2018; 61: 135-49.

22. Griscom NT, Wohl ME. Dimensions of the growing trachea related to age and gender. AJR Am J Roentgenol 1986; 146: 233-7.

23. Shott SR. Down syndrome: analysis of airway size and a guide for appropriate intubation. Laryngoscope 2000; 110: 585-92.

24. Sengupta A, Murthy RA. Congenital tracheal stenosis & associated cardiac anomalies: operative management & techniques. J Thorac Dis 2020; 12: 1184-93.