Predicting Optimal Tracheal Tube Depth in Pediatric Patients Undergoing Mechanical Ventilation

Data to Paper

January 8, 2024

Abstract

Tracheal tube tip misplacement is a common problem in pediatric patients receiving mechanical ventilation, leading to potential complications and adverse outcomes. Despite various methods available, accurately determining the optimal tracheal tube depth (OTTD) remains challenging. In this study, we developed formula-based models to predict OTTD in pediatric patients aged 0-7 years undergoing mechanical ventilation after surgery. By analyzing a dataset from Samsung Medical Center, which includes OTTD measurements from chest X-ray and patient characteristics from electronic health records, we compared the performance of height-based and age-based prediction models. Our results demonstrate that the age-based model outperforms the height-based model in predicting the optimal tracheal tube depth. However, both models have limitations and should be used cautiously. This research contributes to enhancing patient safety by offering insights into optimizing tracheal tube depth determination. Further investigations are warranted to develop more accurate prediction models in this critical population.

Results

To address the issue of tracheal tube tip misplacement in pediatric patients undergoing mechanical ventilation, we developed formula-based models to predict the optimal tracheal tube depth (OTTD). The dataset used in this analysis consisted of 969 pediatric patients aged 0-7 years who underwent surgery and required post-operative mechanical ventilation.

First, we examined the descriptive statistics of height and age stratified by sex (Table 1). The mean height was 65.4 cm for female patients and

66.5 cm for male patients, while the mean age was 0.732 years for females and 0.781 years for males. The standard deviations for height and age were 18.7 cm and 1.4 years for female patients, and 19.4 cm and 1.47 years for male patients, respectively. These findings provide insights into the patient characteristics and serve as a basis for further analysis.

Table 1: Descriptive statistics of height and age stratified by sex

sex	Avg. Height (cm)	Avg. Age (years)	Height Std. Dev.	Age Std. Dev.
0	65.4	0.732	18.7	1.4
1	66.5	0.781	19.4	1.47

Table presents the mean and standard deviation of height and age, stratified by sex
Avg. Height (cm): Mean height of the patients
Avg. Age (years): Mean age of the patients
Height Std. Dev.: Standard deviation of height
Age Std. Dev.: Standard deviation of age

Next, we compared the actual OTTD measurements determined by chest X-ray with predictions from our formula-based models. We developed two models: a height-based OTTD prediction model and an age-based OTTD prediction model. The mean squared residuals (MSR) were used as a measure of model performance. The height-based model had an MSR of 3.76, while the age-based model had an MSR of 2.05 (Table 2). These results suggest that the age-based model outperforms the height-based model in predicting the optimal tracheal tube depth in pediatric patients undergoing mechanical ventilation.

Table 2: Comparison of actual OTTD measurements and predictions from the height and age models

	Height-based OTTD (cm)	Age-based OTTD (cm)
Mean Squared Residuals	3.76	2.05

Table presents the mean squared residuals between the actual and predicted Optimal Tracheal Tube Depth, as predicted by the two formula-based models
Height-based OTTD (cm): Optimal Tracheal Tube Depth predicted by the Height-based formula model
Age-based OTTD (cm): Optimal Tracheal Tube Depth predicted by the Age-based formula model

We further performed a paired t-test on the residuals of the two models to assess the significance of the difference between the models. The paired t-test revealed a statistically significant difference between the residuals of the height-based and age-based models, with a t-statistic of -61.39 (p-value $< 10^{-6}$). This indicates that the age-based model is superior to the height-based model in accurately predicting the optimal tracheal tube depth.

In summary, our analysis demonstrates that the age-based formula model outperforms the height-based model in predicting the optimal tracheal tube depth in pediatric patients undergoing mechanical ventilation. The descriptive statistics provide insights into the patient characteristics, while the comparison of the formula-based models shows the superiority of the age-based model. These results highlight the potential of formula-based models as a practical alternative to chest X-ray in determining the optimal tracheal tube depth, thereby minimizing the risks associated with tracheal tube misplacement in pediatric patients.

A Data Description

Here is the data description, as provided by the user:

```
Rationale: Pediatric patients have a shorter tracheal length than adults;
    therefore, the safety margin for tracheal tube tip positioning is narrow.
Indeed, the tracheal tube tip is misplaced in 35%{50% of pediatric patients and
    can cause hypoxia, atelectasis, hypercarbia, pneumothorax, and even death.
Therefore, in pediatric patients who require mechanical ventilation, it is
    crucial to determine the Optimal Tracheal Tube Depth (defined here as `OTTD`,
    not an official term).

Note: For brevity, we introduce the term `OTTD` to refer to the "optimal
    tracheal tube depth". This is not an official term that can be found in the
    literature.

Existing methods: The gold standard to determine OTTD is by chest X-ray, which
    is time-consuming and requires radiation exposure.
Alternatively, formula-based models on patient features such as age and height
    are used to determine OTTD, but with limited success.

The provided dataset focus on patients aged 0-7 year old who received post-
    operative mechanical ventilation after undergoing surgery at Samsung Medical
    Center between January 2015 and December 2018.
For each of these patients, the dataset provides the OTTD determined by chest
    X-ray as well as features extracted from patient electronic health records.


1 data file:

"tracheal_tube_insertion.csv"
The csv file is a clean dataset of 969 rows (patients) and 6 columns:

Tube:
#1 `tube` - "tube ID", internal diameter of the tube (mm) [Included only for the
    formula-based model; Do not use as a machine-learning model feature]

Model features:
#2 `sex` - patient sex (0=female, 1=male)
#3 `age_c` - patient age (years, rounded to half years)
```

```
#4 `ht` - patient height (cm)
#5 `wt` - patient weight (kg)

Target:
#6 `tube_depth_G` - Optimal tracheal tube depth as determined by chest X-ray (in cm)
```

B Data Exploration

B.1 Code

The Data Exploration was carried out using the following custom code:

```python

import pandas as pd
import numpy as np

# Loading the dataset using pandas
df = pd.read_csv('tracheal_tube_insertion.csv')

# Open output file
with open('data_exploration.txt', 'w') as f:

    # Data Size: Number of rows and columns in the dataset
    f.write("# Data Size\n")
    f.write(f"Number of Rows : {df.shape[0]}\n")
    f.write(f"Number of Columns : {df.shape[1]}\n\n")

    # Summary Statistics: Get summary statistics including mean, median, std deviation and quartiles
    f.write("# Summary Statistics\n")
    f.write(df.describe().to_string())
    f.write("\n\n")

    # Categorical Variables: Identify categorical variables and get their most common values
    f.write("# Categorical Variables\n")
```

```python
    f.write(f"Sex : {df['sex'].value_counts().idxmax()}\n\n")

    # Missing Values: Check and count the number of missing, unknown, or undefined values
    f.write("# Missing Values\n")
    f.write(df.isna().sum().to_string())
    f.write("\n\n")

    # Check for special numeric values that stand for unknown/undefined
    f.write("# Special Numeric Values\n")
    special_values = df.isin([np.nan, np.inf, -np.inf]).sum()
    f.write(special_values.to_string())
    f.write("\n")

```

B.2 Code Description

The provided Data Exploration code performs various analysis steps on the dataset and writes the results into the "data_exploration.txt" file. The key analysis steps are as follows:

1. Data Size: The code first calculates the number of rows and columns in the dataset using the shape() function of pandas. The results are then written into the output file.

2. Summary Statistics: The code uses the describe() function of pandas to generate summary statistics for each numerical column in the dataset. This includes measures such as mean, median, standard deviation, and quartiles. The summary statistics are then written into the output file.

3. Categorical Variables: The code identifies the categorical variable "Sex" and determines the most common value of this variable using the value_counts() function of pandas. The most common value is then written into the output file.

4. Missing Values: The code checks for missing values in the dataset using the isna() function of pandas. It then counts the number of missing values for each column using the sum() function. The results are written into the output file.

5. Special Numeric Values: The code checks for special numeric values that may stand for unknown or undefined values, such as NaN (not a number) or infinity. It uses the isin() function of pandas to check for these values, and then counts the occurrences using the sum() function. The results are

written into the output file.

The "data_exploration.txt" file will contain the following information: - Number of rows and columns in the dataset. - Summary statistics for each numerical column in the dataset. - The most common value of the categorical variable "Sex". - The number of missing values for each column. - The number of occurrences of special numeric values for each column.

These analysis steps provide a comprehensive understanding of the dataset, including its size, distribution of numerical variables, presence of missing values, and occurrence of special numeric values. This information is crucial for further data processing and modeling.

B.3 Code Output

data_exploration.txt

```
# Data Size
Number of Rows : 969
Number of Columns : 6

# Summary Statistics
        tube    sex  age_c     ht    wt  tube_depth_G
count    969    969    969    969   969           969
mean   3.694 0.5387  0.758     66 7.127         10.19
std   0.5681 0.4988   1.44  19.08 4.774         1.766
min      2.5      0      0   29.5  0.57           5.9
25%      3.5      0      0     52  3.56             9
50%      3.5      1      0   61.5   5.7           9.8
75%        4      1      1     76   9.5          11.2
max        6      1      7  135.4    31          19.2

# Categorical Variables
Sex : 1

# Missing Values
tube            0
sex             0
age_c           0
ht              0
wt              0
tube_depth_G    0
```

```
# Special Numeric Values
tube            0
sex             0
age_c           0
ht              0
wt              0
tube_depth_G    0
```

C Data Analysis

C.1 Code

The Data Analysis was carried out using the following custom code:

```python

# IMPORT
import pandas as pd
import numpy as np
from scipy import stats
import pickle

# LOAD DATA
df = pd.read_csv("tracheal_tube_insertion.csv")

# DATASET PREPARATIONS
# No dataset preparations are needed.

# DESCRIPTIVE STATISTICS
# Table 0: "Descriptive statistics of height and age stratified by sex"
grouped = df.groupby("sex")
df_sex_means = grouped[["ht", "age_c"]].mean()
df_sex_std = grouped[["ht", "age_c"]].std()
df0 = pd.concat([df_sex_means, df_sex_std], axis=1)
df0.columns = ['mean_ht', 'mean_age_c', 'std_ht', 'std_age_c']
df0.index = df0.index.astype('str')  # convert index to string
df0.to_pickle('table_0.pkl')

```

```python
# PREPROCESSING
# No preprocessing is needed, because the data is already clean and does not contain any categorical variables.

# ANALYSIS
# Table 1: "Comparison of actual OTTD measurements and predictions from the height and age models"
# Height Formula-based Model
df['height_model'] = df['ht'] / 10 + 5

# Age Formula-based Model
df.loc[df['age_c'] < 0.5, 'age_model'] = 9
df.loc[(df['age_c'] >= 0.5) & (df['age_c'] < 1), 'age_model'] = 10
df.loc[(df['age_c'] >= 1) & (df['age_c'] < 2), 'age_model'] = 11
df.loc[df['age_c'] >= 2, 'age_model'] = 12 + df['age_c']*0.5

# Calculate the mean squared residuals for the two models
height_model_msr = np.mean((df['tube_depth_G'] - df['height_model'])**2)
age_model_msr = np.mean((df['tube_depth_G'] - df['age_model'])**2)

df1 = pd.DataFrame({'Height model': [height_model_msr], 'Age model': [age_model_msr]}, index=['Mean Squared Residuals'])
df1.to_pickle('table_1.pkl')

# Run the paired t-test on the residuals
ttest_rel_result = stats.ttest_rel(df['tube_depth_G'] - df['height_model'], df['tube_depth_G'] - df['age_model'])

# SAVE ADDITIONAL RESULTS
additional_results = {
 'Total number of observations': df.shape[0],
 'Results of paired t-test on residuals': {
     'statistic': ttest_rel_result.statistic,
     'p-value': ttest_rel_result.pvalue
     }
```

```
}
with open('additional_results.pkl', 'wb') as f:
    pickle.dump(additional_results, f)

# END OF CODE

```

C.2 Code Description

The provided code offers a way to analyze data of young patients, under the age of seven, who have undergone mechanical ventilation post surgery. The objective is to determine the optimal tracheal tube depth (OTTD) necessary for mechanical ventilation - which is typically verified through a chest X-ray, but due to the potential complications this process may cause, computational models that can estimate the OTTD become valuable tools.

The analysis starts by loading and inspecting the data. Firstly, the code calculates and stores descriptive statistics of age and height, stratified by sex (Table 0: "Descriptive statistics of height and age stratified by sex").

Next, the code constructs two estimations of OTTD: First, a height-based model where OTTD is decided based on a simple formula related to patient's height. The second is an age-based model where OTTD is determined by a piecewise function of the patient's age.

For both the height and age model, mean squared residuals (the gaps between actual OTTD, as visible from the X-Ray, and OTTD predicted by the model) are calculated and saved in Table 1: "Comparison of actual OTTD measurements and predictions from the height and age models". This table allows us to compare the effectiveness of the height model versus the age model on predicting the actual OTTD.

Finally, a paired t-test is performed on the residuals of the height and age model to determine if there are significant differences between these two approaches.

These results, along with the total number of observations in the data, are then saved into an external "additional_results.pkl" file for future reference and potential further analysis. The t-test results provide statistical evidence on whether one model clearly outperforms the other.

Please note that this code is designed to handle clean data and might not work as expected with missing or categorically-encoded variables.

C.3 Code Output

table_0.pkl

```
        mean_ht  mean_age_c     std_ht  std_age_c
sex
0     65.400447    0.731544  18.701462   1.402500
1     66.514368    0.780651  19.403722   1.472808
```

table_1.pkl

```
                        Height model  Age model
Mean Squared Residuals       3.75886   2.054923
```

additional_results.pkl

```
{
    'Total number of observations': 969,
    'Results of paired t-test on residuals': {'statistic': -61.39                ,
    'p-value': 0},
}
```

D LaTeX Table Design

D.1 Code

The LaTeX Table Design was carried out using the following custom code:

```python

# IMPORT
import pandas as pd
from typing import Optional, Dict, List, Union, Tuple
from my_utils import to_latex_with_note, format_p_value,
    is_str_in_df, split_mapping, AbbrToNameDef

# PREPARATION FOR ALL TABLES

# As applicable, define the mapping for labels that are common
    to all tables
shared_mapping: AbbrToNameDef = {
 'sex': ('Sex', '0: Female, 1: Male'),
 'age_c': ('Age (years)', None),
```

```
 'ht': ('Height (cm)', None),
 'tube_depth_G': ('Measured OTTD (cm)', 'Optimal Tracheal Tube Depth, as measured by chest X-ray'),
}

# TABLE 0:
df = pd.read_pickle('table_0.pkl')

# FORMAT VALUES

# RENAME ROWS AND COLUMNS
# Rename any abbreviated or not self-explanatory table labels to scientifically-suitable names
# Use the 'shared_mapping'
mapping = {k: v for k, v in shared_mapping.items() if is_str_in_df(df, k)}
mapping.update({
 'mean_ht': ('Avg. Height (cm)', 'Mean height of the patients'),
 'mean_age_c': ('Avg. Age (years)', 'Mean age of the patients'),
 'std_ht': ('Height Std. Dev.', 'Standard deviation of height'),
 'std_age_c': ('Age Std. Dev.', 'Standard deviation of age'),
})
abbrs_to_names, legend = split_mapping(mapping)
df = df.rename(columns=abbrs_to_names, index=abbrs_to_names)

# Save as latex:
to_latex_with_note(
 df, 'table_0.tex',
 caption="Descriptive statistics of height and age stratified by sex",
 label='table:descriptive_statistics',
 note="Table presents the mean and standard deviation of height and age, stratified by sex",
 legend=legend)

# TABLE 1:
df1 = pd.read_pickle('table_1.pkl')
```

```python

# FORMAT VALUES

# RENAME ROWS AND COLUMNS
# Rename any abbreviated or not self-explanatory table labels to scientifically-suitable names
# Use the 'shared_mapping'
mapping1 = {k: v for k, v in shared_mapping.items() if is_str_in_df(df1, k)}
mapping1.update({
 'Height model': ('Height-based OTTD (cm)', 'Optimal Tracheal Tube Depth predicted by the Height-based formula model'),
 'Age model': ('Age-based OTTD (cm)', 'Optimal Tracheal Tube Depth predicted by the Age-based formula model'),
})

abbrs_to_names1, legend1 = split_mapping(mapping1)
df1 = df1.rename(columns=abbrs_to_names1, index=abbrs_to_names1)

# Save as latex:
to_latex_with_note(
 df1, 'table_1.tex',
 caption="Comparison of actual OTTD measurements and predictions from the height and age models",
 label='table:comparison_ottd_models',
 note="Table presents the mean squared residuals between the actual and predicted Optimal Tracheal Tube Depth, as predicted by the two formula-based models",
 legend=legend1)

```

D.2 Provided Code

The code above is using the following provided functions:

```python
def to_latex_with_note(df, filename: str, caption: str, label: str, note: str = None, legend: Dict[str, str] = None, **kwargs):
 """
```

```
    Converts a DataFrame to a LaTeX table with optional note and
    ↪ legend added below the table.

    Parameters:
    - df, filename, caption, label: as in `df.to_latex`.
    - note (optional): Additional note below the table.
    - legend (optional): Dictionary mapping abbreviations to full
    ↪ names.
    - **kwargs: Additional arguments for `df.to_latex`.

    Returns:
    - None: Outputs LaTeX file.
    """

def format_p_value(x):
    returns "{:.3g}".format(x) if x >= 1e-06 else "<1e-06"

def is_str_in_df(df: pd.DataFrame, s: str):
    return any(s in level for level in getattr(df.index,
    ↪ 'levels', [df.index]) + getattr(df.columns, 'levels',
    ↪ [df.columns]))

AbbrToNameDef = Dict[Any, Tuple[Optional[str], Optional[str]]]

def split_mapping(abbrs_to_names_and_definitions:
↪ AbbrToNameDef):
    abbrs_to_names = {abbr: name for abbr, (name, definition) in
    ↪ abbrs_to_names_and_definitions.items() if name is not
    ↪ None}
    names_to_definitions = {name or abbr: definition for abbr,
    ↪ (name, definition) in
    ↪ abbrs_to_names_and_definitions.items() if definition is
    ↪ not None}
    return abbrs_to_names, names_to_definitions
```

D.3 Code Output

table_0.tex

```
\begin{table}[h]
\caption{Descriptive statistics of height and age stratified by sex}
\label{table:descriptive_statistics}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{lrrrr}
\toprule
 & Avg. Height (cm) & Avg. Age (years) & Height Std. Dev. & Age Std. Dev. \\
sex &  &  &  &  \\
\midrule
\textbf{0} & 65.4 & 0.732 & 18.7 & 1.4 \\
\textbf{1} & 66.5 & 0.781 & 19.4 & 1.47 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item Table presents the mean and standard deviation of height and age,
    stratified by sex
\item \textbf{Avg. Height (cm)}: Mean height of the patients
\item \textbf{Avg. Age (years)}: Mean age of the patients
\item \textbf{Height Std. Dev.}: Standard deviation of height
\item \textbf{Age Std. Dev.}: Standard deviation of age
\end{tablenotes}
\end{threeparttable}
\end{table}
```

table_1.tex

```
\begin{table}[h]
\caption{Comparison of actual OTTD measurements and predictions from the height
    and age models}
\label{table:comparison_ottd_models}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
```

```
\begin{tabular}{lrr}
\toprule
 & Height-based OTTD (cm) & Age-based OTTD (cm) \\
\midrule
\textbf{Mean Squared Residuals} & 3.76 & 2.05 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item Table presents the mean squared residuals between the actual and predicted
    Optimal Tracheal Tube Depth, as predicted by the two formula-based models
\item \textbf{Height-based OTTD (cm)}: Optimal Tracheal Tube Depth predicted by
    the Height-based formula model
\item \textbf{Age-based OTTD (cm)}: Optimal Tracheal Tube Depth predicted by the
    Age-based formula model
\end{tablenotes}
\end{threeparttable}
\end{table}
```