

PUTTING IT ALL TOGETHER

NU Pathology LearnR! Fall 2020 Session 6

The goal for this exercise is to replicate parts of Table 1 and Figure 1 from the following publication, using the associated, publicly available data set.

Source: LONG-TERM EFFECTS OF BUDESONIDE OR NEDOCROMIL IN CHILDREN WITH ASTHMA by THE CHILDHOOD ASTHMA MANAGEMENT PROGRAM RESEARCH GROUP published in N Engl J Med 2000;343:1054-63.

"The Childhood Asthma Management Program (CAMP) was a clinical trial carried out in children with asthma. The trial was designed to determine the long-term effects of 3 treatments (budesonide, nedocromil, or placebo) on pulmonary function as measured by normalized FEV1 over a 5-6.5 year period."

"Conclusions: In children with mild-to-moderate asthma, neither budesonide nor nedocromil is better than placebo in terms of lung function, but inhaled budesonide improves airway responsiveness and provides better control of asthma than placebo or nedocromil. The side effects of budesonide are limited to a small, transient reduction in growth velocity."

A copy of the article, a modified data set for teaching, and its associated documentation have been provided for this exercise.

STEPS:

1. Set up your environment for the analysis
 - 1.1. Create an RStudio Project called 'CAMP-analysis' in a new directory.
 - 1.2. Create a folder called 'data' within your new project directory and place the CAMP dataset file (camp_teach.csv) there.
 - 1.3. Create a folder called 'figures' within your new project directory. We will save our figure here later.
 - 1.4. Create a new Rmd notebook to hold your code and observation notes.

1.5. Modify your setup chunk as needed and install or load any libraries you will need for this analysis.

2. Read in and clean up the dataset

2.1 Read in the dataset

2.2 Explore the dataset to identify problems

Hint: The dataset is too big to inspect manually. Use some of the tricks you learned in the course to visualize the data and view summary statistics.

As you work, make sure you answer the following questions:

- Are categorical variables represented as factors?
- Are there any outliers? How do you want to handle them?
- Are there any missing values? How do you want to handle those?

2.3 Do the cleaning

- Use the scoped mutate() functions to convert between data types
- Handle NAs using functions like is.na(), replace_na(), fct_explicit_na()

3. Create a summary table of baseline characteristics like Table 1 in the article

CHARACTERISTIC†	BUDESONIDE (N=311)	NEDOCROMIL (N=312)	PLACEBO (N=418)
Age — yr	9.0±2.1	8.8±2.1	9.0±2.2
Race or ethnic group — no. (%)			
Non-Hispanic white	201 (64.6)	218 (69.9)	292 (69.9)
Non-Hispanic black	44 (14.1)	38 (12.2)	56 (13.4)
Hispanic	32 (10.3)	29 (9.3)	37 (8.9)
Other	34 (10.9)	27 (8.7)	33 (7.9)
Sex — no. (%)‡			
Female	130 (41.8)	106 (34.0)	184 (44.0)
Male	181 (58.2)	206 (66.0)	234 (56.0)
Age at onset of asthma — yr	3.1±2.3	3.1±2.4	3.0±2.6
Time since diagnosis of asthma — yr	5.2±2.6	5.0±2.7	4.9±2.7
Treatments in 6 mo before enrollment — no. of patients (%)			
Cromolyn or nedocromil	133 (42.8)	148 (47.4)	160 (38.3)
Inhaled corticosteroid	126 (40.5)	114 (36.5)	150 (35.9)
Oral corticosteroid	107 (34.4)	94 (30.1)	162 (38.8)
Severity of asthma — no. (%)			
Moderate	166 (53.4)	161 (51.6)	216 (51.7)
Mild	145 (46.6)	151 (48.4)	202 (48.3)
Hospitalizations for asthma in year before enrollment — no./100 person-yr	31	29	31
Recordings on daily diary card			
Episode-free days — no./mo§	9.7±7.8	9.9±8.1	9.6±7.6
Use of albuterol for symptoms — puffs/wk	10.4±9.8	10.5±9.8	10.2±9.6
Night awakenings — no./mo	0.9±1.7	1.0±1.7	0.8±1.5
FEV ₁ before bronchodilator use — % of predicted	93.6±14.4	93.4±14.5	94.2±14.0
FEV ₁ after bronchodilator use — % of predicted	103.2±13.2	102.3±12.7	103.3±12.2
Airway responsiveness to methacholine (FEV ₁ PC ₂₀) — mg/ml¶	1.1±3.3	1.2±3.3	1.1±3.3
Height — percentile	56.8±28.0	56.0±28.7	55.3±28.8

3.1 Create a relevant data subset, filter data for base line

3.2 Do the group_by treatment group, create summaries for:

- age - yr (mean ± SD)
- Race or ethnic group — no. (%)
- Sex — no. (%)

- FEV1 before bronchodilator use — % of predicted \pm SD
- FEV1 after bronchodilator use — % of predicted \pm SD

Hint: To format the data as in the paper, you will need to unite information from multiple columns and use the `str_c()` function

3.3 Fix labels to human readable vs column names and reorder, using `recode_factor()`

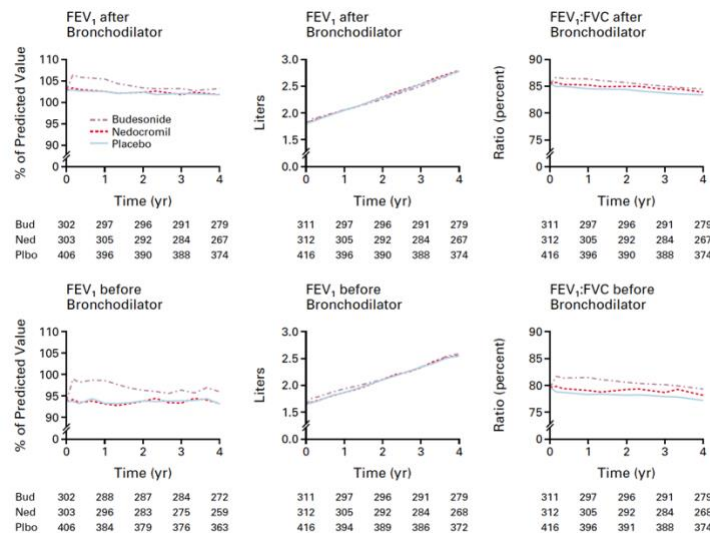
Hint: Use the following snippet to recode the treatment and placebo groups into useful factor names

```
mutate(TG = recode(TG,
  'A' = 'Budesonide',
  'B' = 'Nedocromil',
  'C' = 'Placebo'))
```

3.4 Combine the tables into a single table

Hint: Your tables will need additional columns to help things align nicely. Explore use of the `kableExtra` package to style the table.

4. Create plots like the first 6 plots in Figure 1 from the article (without the values underneath)



4.1 Create a relevant data subset using both filter and select

4.2 Find the mean values for each variable at each year timepoint and treatment group

Hint: remember the scoped variant of summarize, summarize_at() - it is useful here

4.3 Generate individual line plots for each of the variables where the linetype is mapped to the treatment group. Make these plots publication ready, meaning they have appropriate labels and aesthetics, etc.

Hint: The x-axis has to be numeric for a line plot. You will need to convert the visitc values from factor to character and then to numeric and divide by 12 to get years (they are currently in months).

4.4 Compile the individual plots into a single graphic that consists of 3 columns and 2 rows of plots (as shown in the paper). Explore using the ``ggpubr`` package to do this. If you are feeling ambitious, go ahead and figure out how to do it using `ggarrange()`.

4.5 Save your final figure to the figures folder

5. Knit your final report to an HTML document and inspect it

5.1 Modify your code chunk settings so you get the desired HTML output to show only the title of your analysis, your combined table and final figure.