Pediatric Coccidioidomycosis Analysis

James,Owen,Qin

Contents

Context

Pediatric coccidioidomycosis (Valley fever) is **a fungal infection acquired by inhaling soil-dwelling fungal spores**, often presenting in children with symptoms like fever, cough, and extreme fatigue, though many infections are asymptomatic. Complications, more common in children than adults, can include pleural effusion, empyema, and mediastinal involvement, with rare but serious cases potentially leading to disseminated disease affecting the brain, spine, or bones. Diagnosis involves serologic testing and imaging, while treatment for moderate to severe cases uses antifungal medications like fluconazole, sometimes requiring lifelong therapy for central nervous system involvement.

Symptoms

- **Common:** Fever, cough, fatigue, headache, muscle and joint aches, chills.
- **Chest-related:** Shortness of breath, chest pain.
- **Skin manifestations:** A rash and erythema nodosum (tender red bumps under the skin) can occur.
- **Severe complications:** Pleural effusions (fluid around the lung), empyema (pus in the chest cavity), and mediastinal involvement are more frequent in children.

Severe & Disseminated Disease

- **Spread to other areas:** In rare cases, the fungus can spread from the lungs to other parts of the body.
- **Central Nervous System (CNS) involvement:** Coccidioidomycosis can infect the brain or spinal cord, which is a serious and life-threatening condition.
- **Bone and joint infections:** The infection can also cause bone or joint disease.

Diagnosis

- **Imaging:**
  Chest X-rays and CT scans can show signs of the infection, such as lung nodules or inflammation.
- **Blood tests:**
  Serologic tests (antibody tests like IgM and IgG) are crucial for diagnosis.
- **Other tests:**
  In some cases, antigen tests and PCR (polymerase chain reaction) testing of blood, cerebrospinal fluid, or respiratory samples may be performed.

Treatment

- **Supportive care:**
  Mild cases may only require rest and over-the-counter pain and fever reducers.

- **Antifungal medications:**

  Moderate to severe infections or those with a high risk of complications are treated with antifungal drugs, most commonly fluconazole.

- **Serious infections:**

  Amphotericin B may be used for severe, diffuse, or disseminated infections.

- **Lifelong therapy:**

  Some very severe cases, such as those with meningitis (infections of the brain and spinal cord), may require lifelong antifungal treatment.

Prevention

- **Reduce dust exposure:**

  In endemic areas (like the Southwestern United States), efforts can be made to reduce dust during construction, and children may need to reduce outdoor play during windy conditions.

- **Wear masks:**

  Face masks can protect children and adults from inhaling fungal spores in dusty environments.

Data Exploration

Data Loading and formatting

Data Types:

- **Study ID**: Categorical. ID of the patient. (Not typical useful here)
- **Erythema nodosum:** Erythema nodosum is an inflammatory skin condition characterized by the development of painful, red, and tender nodules or lumps, typically on the shins (rash). Since it sounds like a more severe symptom, there should be an order on it, but we only have two categories (Yes or No) - Categorical without order.
- **Age at diagnosis:** Age - Continuous.
- **Ethnicity:** Without assume any superior ethnicity - Categorical without order.
- **Race:** Without assume any superior race - Categorical without order.
- **Gender**: Without assume any superior gender - Categorical without order.
- **Disseminated disease:** refers to a condition where an infection or other pathological process spreads throughout the body from its original site. It might be related to Erythema nodosum. - Categorical without order.
- **Associated hospitalization:** means the hospital within or in association with which a body corporate pursues its objects. (useful? maybe) - Categorical without order.
- **Antifungal treatment:** as it means. We do not know what happens after the treatment. - Categorical without order.
- **Comorbidity:**
  - **Pulmonary disease**: refers to a group of conditions that affect the lungs and respiratory system. These diseases can cause inflammation, damage, or obstruction of the airways, leading to various symptoms and complications. **This feature might be highly correlated to the rash.**
  - **DM:** diabetes.
  - **Primary or congenital immunodeficiency:** refers to a group of rare, genetic disorders where the immune system doesn't work correctly, leaving individuals vulnerable to recurrent, severe, or unusual infections. **This feature might be highly correlated to the rash.**
  - **Current malignancy:** refers to a pre-existing chronic condition or other disease (a "comorbidity") that coexists with cancer ("malignancy") at the same time.
  - **Prior malignancy:** Similar to the previous one. With cancer before.
  - **Immunosuppresant medication:** are drugs that weaken the immune system to prevent the body from rejecting transplanted organs or treating autoimmune disorders.
  - **Autoimmune disease:** the body's immune system mistakenly attacks its own healthy tissues and organs.
  - **HIV:** everyone knows.

  - **Prematurity:** health issues that are more likely to occur in individuals born preterm, including respiratory problems (like bronchopulmonary dysplasia), infections due to immature immune systems. **This feature might be highly correlated to the rash.**
  - **Congenital heart disease:** refers to structural defects in the heart that are present at birth.
  - **comorbidities (choice=None):** Not sure what it is. Maybe it means comorbidity in general?
  - All the categories of comobodity are coded in yes or no. We can assign them categorical without order or take look at the correlations among them first.

- **EIA IgG:** Detection of rubella IgM by enzyme immunoassay (EIA) is used to confirm suspected cases of acute rubella infection and congenital rubella syndrome (CRS). (Categorical without order)
- **EIA IgM:** An EIA IgM test uses a form of enzyme immunoassay (EIA) to detect Immunoglobulin M (IgM) antibodies, which are the first antibodies the body produces in response to a new infection, indicating recent or active exposure to a specific virus, bacterium, or other pathogen. (Categorical without order)
- **Titer 1:** A blood test that measures the level of specific antibodies in the blood (This might be highly correlated to the rash, IgG, IgM) (Categorical with order by taking log_2)
- **CXR: Normal:** X - ray
- **CXR: Lymphadenopathy:** chest X-ray
- **CXR: Pleural effusion:** Pleural effusion" is commonly used as a catch-all term to describe any abnormal accumulation of fluid in the pleural cavity
- CXR: Cavitation
- CXR: Consolidation/Opacity
- CXR: Nodules/Micronodules
- CXR: Pneumothorax
- No CXR performed
- Was tissue/fluid cultured?
- Was specimen sent for pathology?

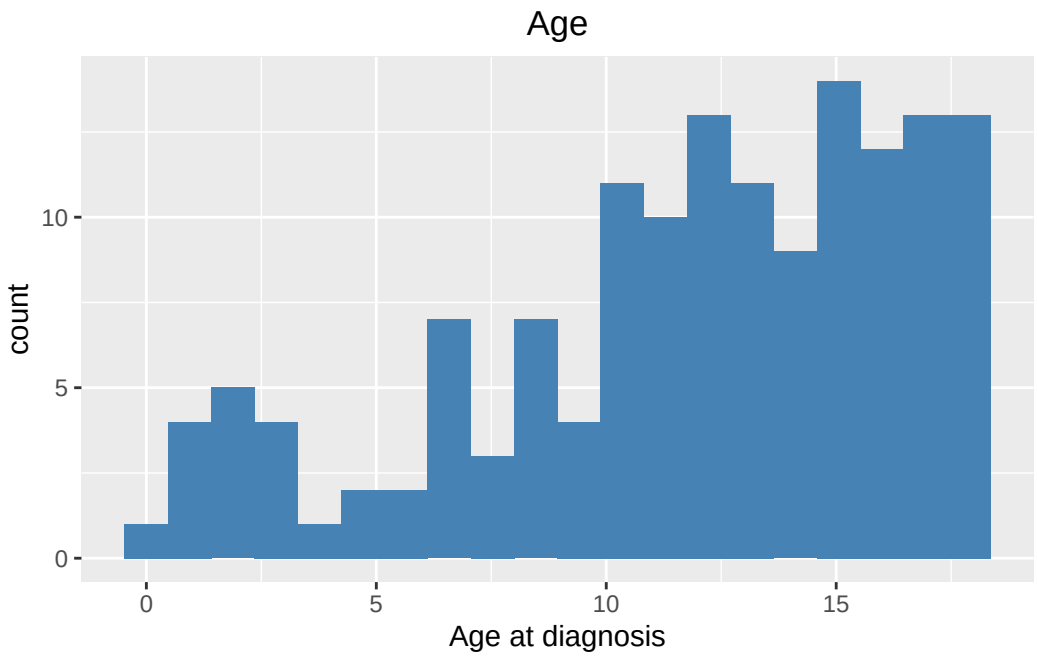
```
# data cleaning
# drop the na

rash_df[rash_df == ""] <- NA
rash_df_cleaned <- na.omit(rash_df)
```

The original dataset has 157 instances. After dropped the empty rows, it still has 157 .

Graph

First question, what does the distribution of age look like?

```
library(ggplot2)
# transform the string to number in age
rash_df_cleaned$`Age at diagnosis` <- as.numeric(rash_df_cleaned$`Age at diagnosis`)
ggplot(rash_df_cleaned, aes(x = `Age at diagnosis`)) +
  geom_histogram(bins = 20,fill="steelblue") +
  labs(title = "Age") +
  theme(plot.title = element_text(hjust = 0.5))
```




```
library(dplyr)
```

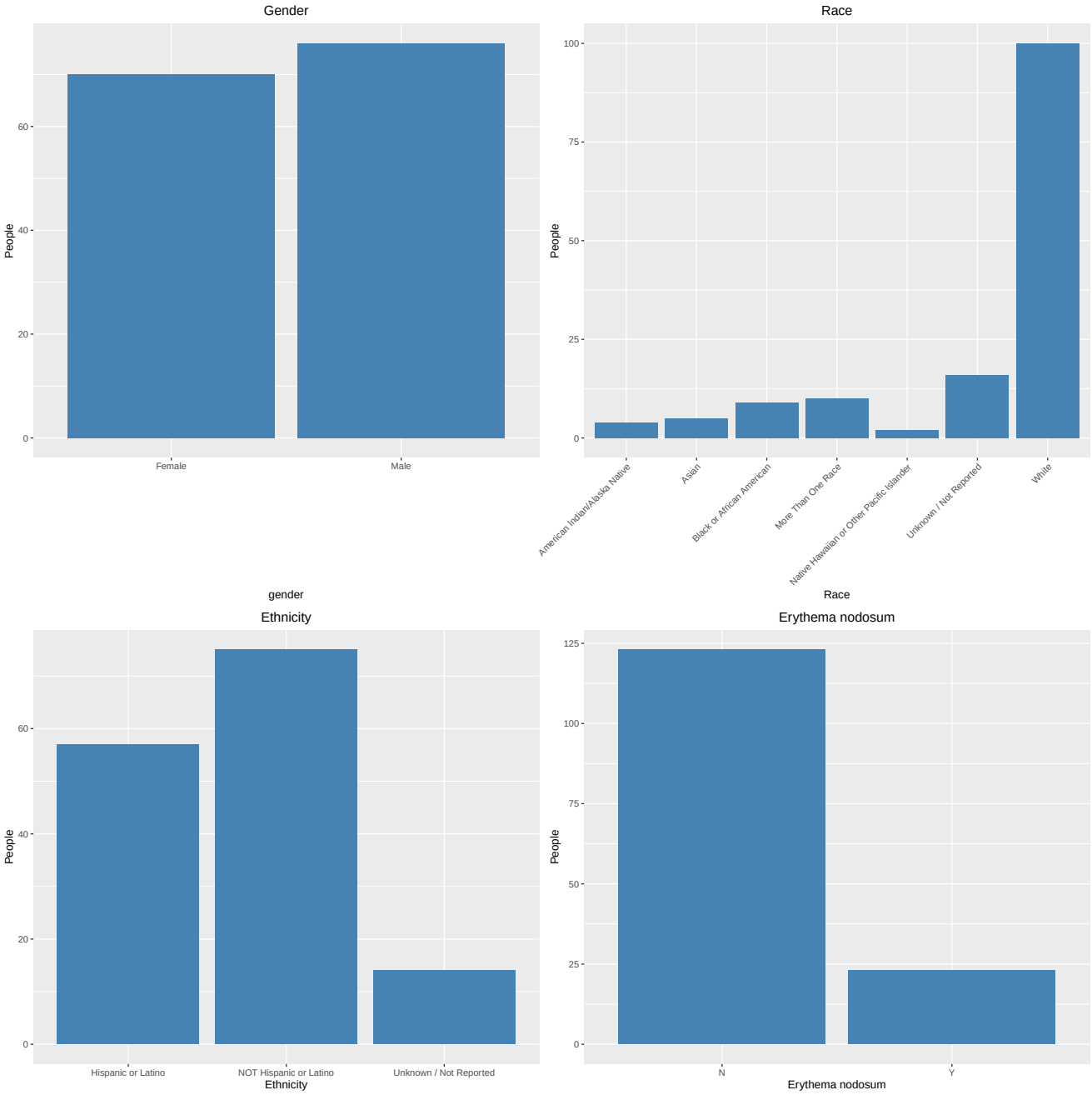
```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':

    filter, lag
```

```
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(patchwork)

gen_count <- rash_df_cleaned %>%
  count(Gender)

race_count <- rash_df_cleaned %>%
  count(Race)

eth_count <- rash_df_cleaned %>%
  count(Ethnicity)

rash_count <- rash_df_cleaned %>%
  count(`Erythema nodosum`)

gen_plot <- ggplot(gen_count, aes(x = Gender, y=n)) +
  geom_col(fill="steelblue") +
  labs(title = "Gender", x="gender",y="People") +
  theme(plot.title = element_text(hjust = 0.5))
```

```r
race_plot <- ggplot(race_count, aes(x = Race, y=n)) +
  geom_col(fill="steelblue") +
  labs(title = "Race", x="Race",y="People") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1))

eth_plot <- ggplot(eth_count, aes(x = Ethnicity, y=n)) +
  geom_col(fill="steelblue") +
  labs(title = "Ethnicity", x="Ethnicity",y="People") +
  theme(plot.title = element_text(hjust = 0.5))

rash_plot <- ggplot(rash_count, aes(x = `Erythema nodosum`, y=n)) +
  geom_col(fill="steelblue") +
  labs(title = "Erythema nodosum", x="Erythema nodosum",y="People") +
  theme(plot.title = element_text(hjust = 0.5))

(gen_plot + race_plot) / (eth_plot + rash_plot)
```




```
library(recipes)
```

```
Attaching package: 'recipes'
```

```
The following object is masked from 'package:stats':

    step
```

```
rash_dummy <- recipe(~., data = rash_df_cleaned) %>%
  step_rm(`Study ID`) %>%
  step_zv(all_nominal_predictors()) %>%
  step_dummy(all_nominal_predictors(), one_hot = TRUE)
```

```
prepared_rec <- prep(rash_dummy, training = rash_df_cleaned)
one_hot_encoded_df <- bake(prepared_rec, new_data = rash_df_cleaned)

rash_cor <- cor(one_hot_encoded_df)
library(ggcorrplot)
ggcorrplot(rash_cor,hc.order = TRUE, outline.col = "white",
        method = "circle")
```