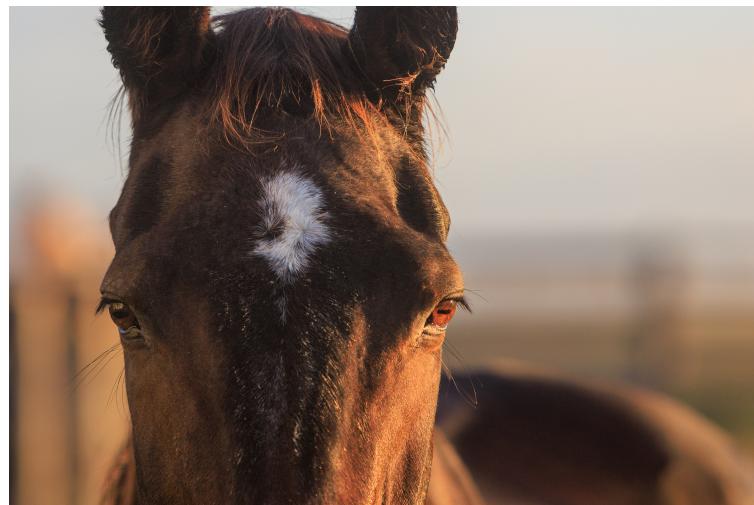


Clinical Clues to Survival in Horses Affected by Colic

A Statistical Analysis of Key Clinical Indicators in Equine Colic

Adam van Rensburg

December 2025



1 Executive Summary

This project analyzes clinical data from 127 horses affected by colic to identify whether pain level, abdominal distention, and mucous membrane color are related to survival outcomes. After cleaning the dataset and examining summary statistics, exploratory heatmaps revealed clear patterns showing that horses with higher pain levels and more abdominal distention had significantly lower survival rates. Logistic regression was then used to quantify these relationships and determine the strength of each predictor. The results showed that pain level and abdominal distention significantly influence survival probability, while mucous membrane color appears to play a weaker but still meaningful role. Overall, this analysis demonstrates how clinical examination data can help identify high-risk cases and support decision-making in veterinary medicine.

2 Introduction

Colic is a very serious medical issue in horses, it is abdominal pain caused by gas buildup in the digestive tract, spasms, and blockages. Colic can progress very rapidly in horses and could potentially become life threatening. Veterinarians rely on clinical signs like pain level, mucous membrane color, abdominal distention, heart rate, and other measurements to evaluate how bad the issue is for the patient. The reason I chose this data set is because I'm studying animal science and being able to see how certain measurable indicators can lead to real-world outcomes is a great way to show my passion for animals and data!

The research question trying to be answered: Are certain clinical signs like pain level, mucous membrane color, and abdominal distention related to survival outcomes in horses with colic? I will be exploring how these variables are different between horses that lived, died, or euthanized.

The data set used in this analysis is the Horse Colic Dataset from the UCI Machine Learning Repository accessed through Kaggle. This data set has over 299 observations and 29 variables to start with. The data set includes medical records collected from horses admitted to a veterinary hospital for colic evaluation. The cases represent a single horse and the information was recorded during a physical examination, diagnostic workup, and if surgery was needed or not. The variables in the data set include categorical such as pain level, abdominal distention, and mucous membrane color. The discrete variables are rectal exam findings and continuous numerical variables such as pulse, respiratory rate, and total protein.

This project will start off with cleaning and preparing the data set for analysis which will include taking care of missing values and making sure each variable is able to be used. The goal of this project is to better understand which indicators may be helpful in predicting if a horse has colic.

2.1 Data Cleaning and Preparation

2.1.1 Load Dataset

```
horse_raw <- read.csv("horse.csv",
                      na.strings = c("?", "", "NA"),
                      stringsAsFactors = FALSE)

str(horse_raw)
dim(horse_raw)
```

2.1.2 Clean Variable Types and Factors

```
library(dplyr, warn.conflicts = FALSE)

horse_clean <- horse_raw %>%
  mutate(
    hospital_number = factor(hospital_number),
    surgery = factor(surgery, levels = c("no", "yes")),
```

```

age = factor(age, levels = c("adult", "young")),
temp_of_extremities = factor(temp_of_extremities),
peripheral_pulse = factor(peripheral_pulse),
mucous_membrane = factor(mucous_membrane),
capillary_refill_time = dplyr::recode(capillary_refill_time, "3" = "three_sec"),
capillary_refill_time = factor(capillary_refill_time,
                               levels = c("less_3_sec", "three_sec", "more_3_sec")),
pain = factor(pain),
peristalsis = factor(peristalsis),
abdominal_distention = factor(abdominal_distention),
nasogastric_tube = factor(nasogastric_tube),
nasogastric_reflux = factor(nasogastric_reflux),
rectal_exam_feces = factor(rectal_exam_feces),
abdomen = factor(abdomen),
abdomo_appearance = factor(abdomo_appearance),
outcome = factor(outcome, levels = c("lived", "died", "euthanized")),
surgical_lesion = factor(surgical_lesion, levels = c("no", "yes")),
cp_data = factor(cp_data, levels = c("no", "yes")),
lesion_1 = factor(lesion_1),
lesion_2 = factor(lesion_2),
lesion_3 = factor(lesion_3)

```

2.2 Missing Data Analysis

```

colSums(is.na(horse_clean))
horse_clean$na_count <- apply(horse_clean, 1, function(x) sum(is.na(x)))
summary(horse_clean$na_count)

```

2.2.1 Filter NA Rows

```

horse_filtered <- horse_clean %>%
  filter(na_count <= 10)
dim(horse_filtered)

## [1] 268 29

```

2.2.2 Create Final Data set

The data set in use now has 127 observations and 13 variables. These variables are surgery, age, rectal_temp, pulse, respiratory_rate, temp_of_extremities, mucous_membrane, pain, peristalsis, abdominal_distention, packed_cell_volume, total_protein, and outcome.

```

vars_for_analysis <- c("surgery", "age",
"rectal_temp", "pulse", "respiratory_rate",
"temp_of_extremities", "mucous_membrane",
"pain", "peristalsis", "abdominal_distention",
"packed_cell_volume", "total_protein",
"outcome")

horse_analysis <- horse_filtered %>%
  select(all_of(vars_for_analysis)) %>%
  filter(complete.cases(.))

dim(horse_analysis)
str(horse_analysis)

```

3 Data Analysis

3.1 Outcome and Predictor Variables

The response Y for this analysis is the outcome variable in the data set, which represents if each horse lived, died, or was euthanized after the colic evaluation. To get a better understanding if the clinical signs are related with survival. The predictor variable used for this project are pain levels, abdominal distention, and mucous membrane color. These variable are observed signs that veterinarians saw during examination and relate directly to our research question, “are certain clinical signs related to the survival outcomes in horses with colic”?

3.2 Preliminary Exploratory Data Analysis

Exploratory data analysis gives us a better understanding of the distribution and predictor variables like mucous membrane color, abdominal pain, and pain levels. Summary statistics for the cleaned data set show the survival outcome are not evenly distributed. Most of the horses survived(n=82), while only 33 died and 12 euthanized. The pain variable ranges from alert to severe pain. The mucous membrane color variable includes colors from bright pink, red, cyanotic, normal pink, and pale. 45 horses had normal pink color which was the highest and only 8 had dark cyanotic.

The code chunk below displays summary statistics for all the variables in the analysis. These statistics give an insight on the distribution of variables and which categories have enough support.

```
summary(horse_analysis)
```

```
##   surgery      age      rectal_temp      pulse      respiratory_rate
##   no :54    adult:116   Min.   :36.40   Min.   : 36.00   Min.   : 9.00
##   yes:73    young: 11   1st Qu.:37.80   1st Qu.: 48.00   1st Qu.:19.00
##                                         Median :38.20   Median : 60.00   Median :28.00
##                                         Mean   :38.20   Mean   : 70.89   Mean   :30.17
##                                         3rd Qu.:38.55   3rd Qu.: 88.00   3rd Qu.:38.00
##                                         Max.   :40.80   Max.   :184.00   Max.   :88.00
##   temp_of_extremities      mucous_membrane      pain      peristalsis
##   cold  :13      bright_pink   :21      alert     :24      absent    :29
##   cool   :48      bright_red   :11      depressed  :36      hypermotile:24
##   normal :52      dark_cyanotic: 8      extreme_pain:14      hypomotile:65
##   warm   :14      normal_pink  :45      mild_pain  :39      normal    : 9
##                           pale_cyanotic:17      severe_pain:14
##                           pale_pink   :25
##   abdominal_distention packed_cell_volume total_protein      outcome
##   moderate:36          Min.   :30.00      Min.   : 3.30   lived    :82
##   none    :46          1st Qu.:38.00      1st Qu.: 6.50   died     :33
##   severe   :13          Median :45.00      Median : 7.20   euthanized:12
##   slight   :32          Mean   :45.92      Mean   :22.04
##                           3rd Qu.:50.00      3rd Qu.:48.00
##                           Max.   :75.00      Max.   :82.00
```

3.3 Heatmap Visualization of Survival Proportion

To visually show how pain level and abdominal distention relate to survival outcomes in horses, I created a heat map using ggplot, showing the proportion of horses that survived.

```
library(dplyr)
library(ggplot2)
```

```
horse_plot <- horse_analysis %>%
```

```

mutate( survived = ifelse(outcome == "lived", 1, 0),
pain = factor(pain,
levels = c("alert", "mild_pain", "depressed", "severe_pain", "extreme_pain")),
abdominal_distention = factor(abdominal_distention, levels = c("none", "slight", "moderate", "severe")))

heatmap_data1 <- horse_plot %>%
group_by(abdominal_distention, pain) %>%

summarise(survival_rate = mean(survived), .groups = "drop")
ggplot(heatmap_data1, aes(x = pain, y = abdominal_distention, fill = survival_rate)) +
geom_tile(color = "white", linewidth = 0.4) +
geom_text(aes(label = sprintf("%.2f", survival_rate)), size = 2.8) +
scale_fill_gradientn(colors = c("red", "yellow", "green"),
limits = c(0, 1), name = "Proportion\nSurvived") +
labs(title = "Survival Proportion by Pain Level and Abdominal Distention",
x = "Pain Level", y = "Abdominal Distention") +
theme_minimal(base_size = 10) +
theme( plot.title = element_text(size = 12, hjust = 0.5),
axis.title = element_text(size = 10),
axis.text.x = element_text(size = 8, angle = 45, hjust = 1),
axis.text.y = element_text(size = 9),
legend.title = element_text(size = 10),
legend.text = element_text(size = 8))

```

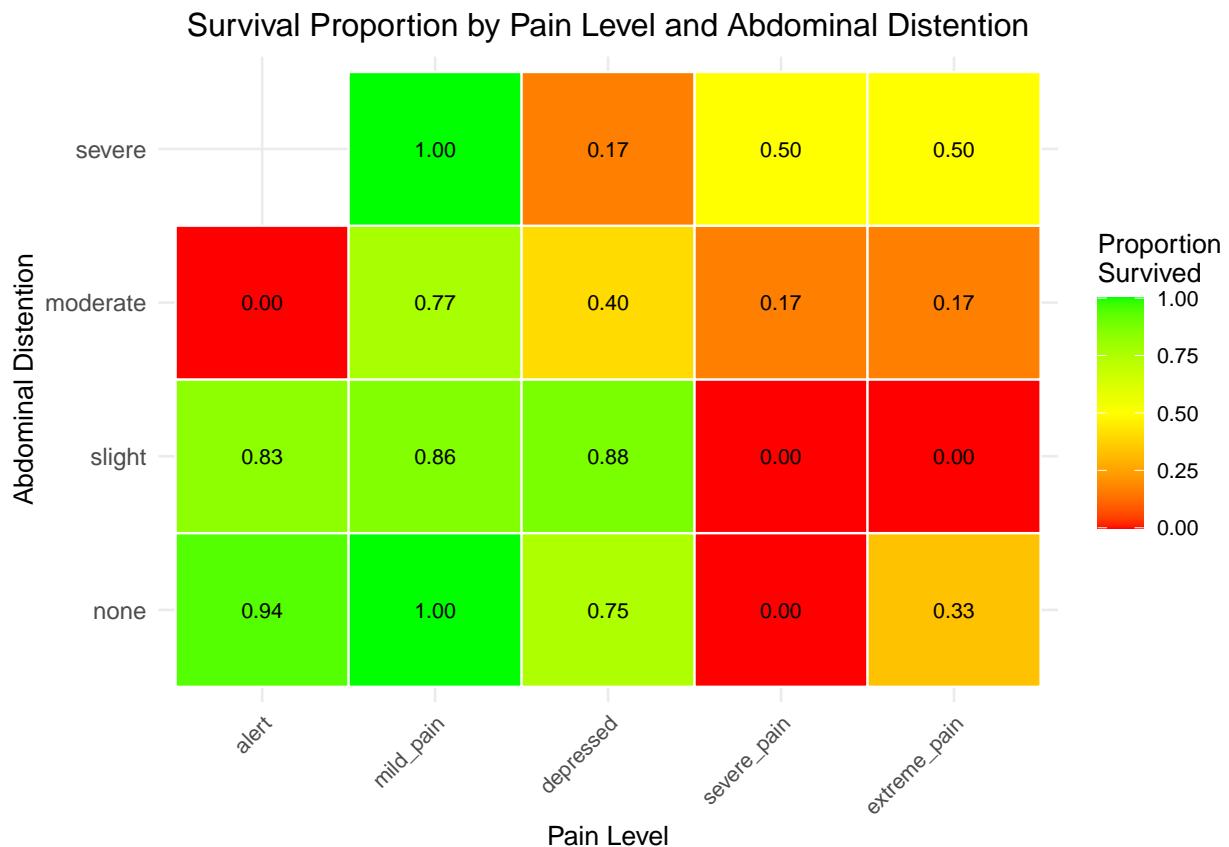


Figure 1

In this heat map each of the squares show the proportion of horses that survived for a certain pain level and

abdominal distention combination. The pain level represents the x-axis while the abdominal distention is on the y-axis. The color scheme is from green which is a high survival rate to yellow which is a medium survival rate. Last is Red which means low/zero survival rate. Overall, this map displays horses with lower pain levels like alert or mild and no abdominal distention have the highest survival rates. As the pain levels increase and abdominal distention, the tiles shift towards a red color. This is showing severe pain or moderate to severe abdominal distention are less likely to survive.

This plot makes it easy to see that high abdominal distention and pain is very dangerous for the horse. This plot suggest that both pain levels and distention are strongly related to the outcome of the horse.

3.4 Survival Rate by Pain Level and Mucous Membrane Color

To visually show how pain level and mucous membrane color relate to survival outcomes in horses, I created a heat map using ggplot, showing the proportion of horses that survived.

```
horse_plot2 <- horse_analysis %>%
  mutate( survived = ifelse(outcome == "lived", 1, 0),
  pain = factor(pain, levels = c("alert", "mild_pain", "depressed", "severe_pain", "extreme_pain")),
  mucous_membrane = factor(mucous_membrane,
  levels = c("normal_pink", "bright_pink", "pale_pink",
  "bright_red", "pale_cyanotic", "dark_cyanotic")))

heatmap_data2 <- horse_plot2 %>%
  group_by(mucous_membrane, pain) %>%
  summarise(survival_rate = mean(survived), .groups = "drop")

ggplot(heatmap_data2,
  aes(x = pain, y = mucous_membrane, fill = survival_rate)) +
  geom_tile(color = "white", linewidth = 0.4) +
  geom_text(aes(label = sprintf("%.2f", survival_rate)), size = 2.8) +
  scale_fill_gradientn(colors = c("red", "yellow", "green"),
  limits = c(0, 1), name = "Proportion\nSurvived") +
  labs( title = "Survival Proportion by Pain Level and Mucous Membrane Color",
  x = "Pain Level",
  y = "Mucous Membrane Color") +
  theme_minimal(base_size = 10) +
  theme( plot.title = element_text(size = 12, hjust = 0.5),
  axis.title = element_text(size = 10),
  axis.text.x = element_text(size = 8, angle = 45, hjust = 1),
  axis.text.y = element_text(size = 8),
  legend.title = element_text(size = 10),
  legend.text = element_text(size = 8))
```

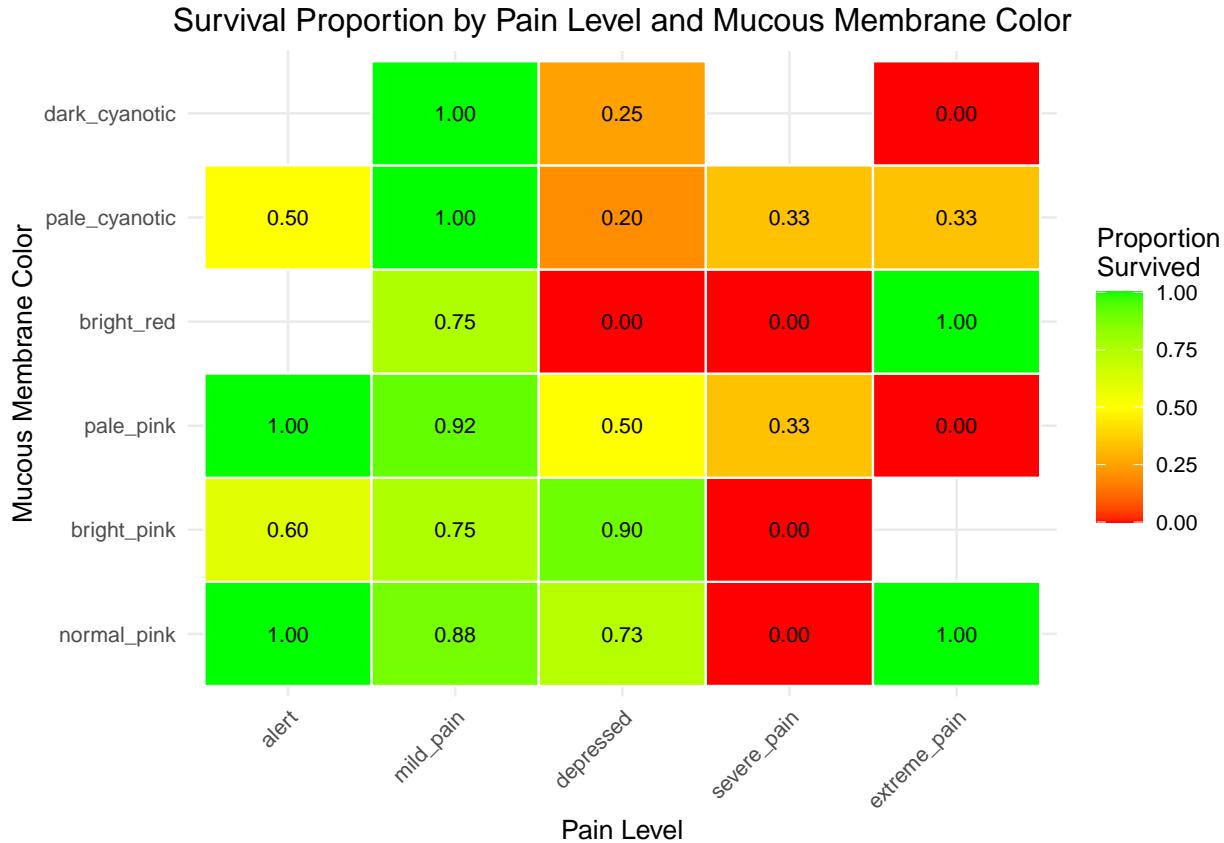


Figure 2

In this second heat plot the squares represent a combination of pain level and mucous membrane color. The green tiles are high survival and yellow/red are low or no survival for the horse. Normal pink or bright pink membrane color means a higher survival rate. While other colors like pale pink, pale cyanotic, and bright red mean a lower chance of survival. The variable pain levels have a big part in survival with most horses having severe pain have no chance to live. On the other hand mild pain there is a higher survival chance for the horse. Overall, mucous membrane color isn't as reliable to predict the horses outcome than pain levels, so it's a good supporting variable not the main predictor. This plot suggest a relationship with membrane color, but the abdominal distention heat map showed a stronger relationship.

3.5 Methods for Statistical Analysis

To test whether the clinical signs in the research question are related to survival I will use logistic regression. This statistical methods is good for this project because the outcome variable can be treated as the binary measure of survival which is lived, died, or euthanized. This method will give me an estimate on how each of the predictors(pain level, abdominal distention, mucous membrane color) affect the survival rate while having the other variables at a constant.

3.6 Required Results to Support

The results needed from the logistic regression model will be p-values this will show whether each predictor is statistically significant. Next is coefficient and odds ratio which will give an indicator if a variable increase/decrease the survival odds and by how much. Confidence intervals are also important and will gives us the uncertainty of these estimates. All of these will help determine if pain levels, abdominal distention, and mucous membrane color could predict the survival in horses with colic.

3.7 Logistic Regression Table

```

library(dplyr)
library(broom)
library(stringr)
library(knitr)

horse_logistic <- horse_analysis %>%
  mutate(survived_bin = ifelse(outcome == "lived", 1, 0))
logit_model <- glm(
  survived_bin ~ pain + abdominal_distention + mucous_membrane,
  data = horse_logistic,
  family = binomial)
logit_table <- tidy(logit_model) %>%
  mutate(
    term = str_replace(term, "^pain", "Pain: "),
    term = str_replace(term, "^abdominal_distention", "Abd Dist: "),
    term = str_replace(term, "^mucous_membrane", "Muc Mem: "),
    term = str_replace_all(term, "_", " "),
    odds_ratio = exp(estimate),
    odds_low = exp(estimate - 1.96 * std.error),
    odds_high = exp(estimate + 1.96 * std.error))
kable(
  logit_table,
  digits = 3,
  caption = "Logistic Regression Results: Predictors of Survival")

```

Table 1: Logistic Regression Results: Predictors of Survival

term	estimate	std.error	statistic	p.value	odds_ratio	odds_low	odds_high
(Intercept)	0.510	0.995	0.512	0.609	1.665	0.237	11.703
Pain: depressed	-0.787	0.793	-0.992	0.321	0.455	0.096	2.155
Pain: extreme pain	-1.435	1.012	-1.418	0.156	0.238	0.033	1.732
Pain: mild pain	0.837	0.903	0.926	0.354	2.309	0.393	13.565
Pain: severe pain	-2.731	1.072	-2.547	0.011	0.065	0.008	0.533
Abd Dist: none	1.627	0.682	2.385	0.017	5.087	1.336	19.366
Abd Dist: severe	0.845	0.832	1.016	0.310	2.328	0.456	11.888
Abd Dist: slight	1.192	0.672	1.773	0.076	3.294	0.882	12.308
Muc Mem: bright red	-1.449	1.025	-1.414	0.157	0.235	0.031	1.751
Muc Mem: dark	-1.892	1.082	-1.749	0.080	0.151	0.018	1.256
cyanotic							
Muc Mem: normal	0.335	0.746	0.449	0.654	1.398	0.324	6.029
pink							
Muc Mem: pale	-0.751	0.857	-0.876	0.381	0.472	0.088	2.532
cyanotic							
Muc Mem: pale pink	0.017	0.802	0.021	0.983	1.017	0.211	4.903

Figure 3

The logistic regression model indicates that pain level and abdominal distention have strong negative associations with survival, with several categories showing significant p-values. Odds ratios below 1 suggest decreased survival probability for horses in higher pain or greater abdominal distention categories. Mucous membrane color shows weaker effects, consistent with patterns observed in the exploratory visualizations.

4 Closing Remarks

This project explores if clinical signs are related to survival outcomes in horses affected by colic. Through cleaning the data set, analysis, and visualizations the results showed strong patterns in higher pain levels and more abdominal distention cause lower survival chances. Mucous membrane color wasn't as significant as the other two predictors. In the end understanding how these signs related to survival may improve the outcomes for horses with colic.