Trauma-Associated Page Kidney: A Systematic Review and Data Analysis

Anthony Onde Morada, MD

June 29, 2025

Contents

---

1 Required Libraries

```r
library(dplyr)
library(knitr)
library(kableExtra)
library(stringr)
library(tidyr)
library(glue)
library(ggplot2)
```

2 Data Loading and Preprocessing

```r
# Load the raw dataset
df <- read.csv("data/trauma-page-kidney-table.csv") %>%
    select(-review.1, -review.2)

# Identify excluded manuscripts for reporting
exdf <- df %>% filter(!str_starts(include, "y"))

# Process included manuscripts with comprehensive data cleaning
df <- df %>% filter(str_starts(include, "y")) %>%
    select(-pdf, -include) %>%
    mutate(
        reference = as.factor(reference),
        year = as.integer(year),
        age = as.numeric(age),
        gender = factor(gender),
        pmh = as.character(pmh),
        onset.y = as.numeric(onset.y),
        mechanism = factor(mechanism),
        ss = as.character(ss),
        hypertension = if_else(hypertension == "y", TRUE, FALSE),
        ua = factor(ua),
        grade = factor(grade, levels = sort(unique(grade), na.last = TRUE)),
        size.cm = as.numeric(size.cm),
        page.type = factor(page.type),
        laterality = factor(laterality),
        treatment = as.character(treatment),
        fu.status = if_else(str_trim(fu.status) == "s", TRUE, NA),
        fu.y = as.numeric(fu.y)
  )
```

```r
# Clean urinalysis results for standardized reporting
df <- df %>%
  mutate(
    ua_clean = str_to_lower(ua),
    ua_result = case_when(
      ua_clean %in% c("hematuria", "trace blood") ~ "Positive",
      ua_clean == "negative" ~ "Negative",
      TRUE ~ NA_character_  # preserve NA when UA not done
    )
  )

# Standardize laterality reporting
df <- df %>%
  mutate(
    laterality = tolower(as.character(laterality)),
    laterality = gsub(",", "", laterality),
    laterality = str_trim(laterality),
    laterality = factor(laterality)
  )
```

Dataset Processing Summary:

- Total references processed: 35
- References excluded: 11
- References included in analysis: 24
- Unique studies analyzed: 23
- Individual patients: 24

3 Study Selection and Exclusions

```r
# Process exclusion data
exdf <- exdf %>%
  select(reference, include) %>%
  mutate(
    reference = as.factor(reference),
    include = as.factor(include)
  )

# Create exclusion summary table
exclusion_summary <- as.data.frame(table(exdf$include))
colnames(exclusion_summary) <- c("Reason for Exclusion", "Number of References")

exclusion_summary %>%
  kbl(caption = "Summary of Study Exclusions",
      align = "lc",
      booktabs = TRUE) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = FALSE,
                position = "center")
```

Total excluded references: 11

Table 1: Summary of Study Exclusions

Reason for Exclusion	Number of References
duplicate	1
not-available	6
path	2
patient level	2

Excluded references: Arai_1981, Barns_2022, Benendo-Kapuscinska_1999, El Madhoun_2009, Estrada_2019, Hoshiyama_2009, Oliveira_2002, Scott_1976, Shome_2002, Smyth_2012, van Ahlen_1987

4 Study Patterns

```
# Page Kidney Systematic Review: Publication Year Analysis
# Load required libraries
years <- df$year
# Basic descriptive statistics
basic_stats <- data.frame(
  Statistic = c("Total papers", "Min year", "Max year", "Year span", "Mean year", "Median year"),
  Value = c(
    length(years),
    min(years),
    max(years),
    max(years) - min(years),
    round(mean(years), 1),
    median(years)
  )
)

print("=== BASIC STATISTICS ===")
```

```
[1] "=== BASIC STATISTICS ==="
```

```
kable(basic_stats, caption = "Basic Publication Statistics")
```

Table 2: Basic Publication Statistics

Statistic	Value
Total papers	24
Min year	1969
Max year	2024
Year span	55
Mean year	2009
Median year	2016

```r
# Frequency distribution by decade
decade_freq <- years %>%
  sapply(function(x) floor(x/10) * 10) %>%
  table() %>%
  as.data.frame()
names(decade_freq) <- c("Decade", "Papers")
decade_freq$Decade <- paste0(decade_freq$Decade, "s")

print("=== FREQUENCY BY DECADE ===")
```

```
[1] "=== FREQUENCY BY DECADE ==="
```

```r
kable(decade_freq, caption = "Publications by Decade")
```

Table 3: Publications by Decade

Decade	Papers
1960s	1
1970s	1
1990s	3
2000s	4
2010s	9
2020s	6

```r
# Era analysis
era_breaks <- c(1969, 2000, 2010, 2020, 2025)
era_labels <- c("Pre-2000", "2000-2009", "2010-2019", "2020-2024")

era_analysis <- data.frame(
  Era = era_labels,
  Papers = c(
    sum(years < 2000),
    sum(years >= 2000 & years < 2010),
    sum(years >= 2010 & years < 2020),
    sum(years >= 2020)
  )
)
era_analysis$Percentage <- round((era_analysis$Papers / length(years)) * 100, 1)
era_analysis$Rate_per_year <- c(
  era_analysis$Papers[1] / 31,  # 1969-1999 = 31 years
  era_analysis$Papers[2] / 10,  # 2000-2009 = 10 years
  era_analysis$Papers[3] / 10,  # 2010-2019 = 10 years
  era_analysis$Papers[4] / 5    # 2020-2024 = 5 years
)
era_analysis$Rate_per_year <- round(era_analysis$Rate_per_year, 2)

print("=== ERA DISTRIBUTION ===")
```

```
[1] "=== ERA DISTRIBUTION ==="
```

```r
kable(era_analysis, caption = "Publications by Era with Rates")
```

Table 4: Publications by Era with Rates

Era	Papers	Percentage	Rate__per__year
Pre-2000	5	21	0.16
2000-2009	4	17	0.40
2010-2019	9	38	0.90
2020-2024	6	25	1.20

```r
# Recent years analysis (2015-2024)
recent_years <- 2015:2024
recent_counts <- sapply(recent_years, function(x) sum(years == x))
recent_analysis <- data.frame(
  Year = recent_years,
  Papers = recent_counts
)

print("=== RECENT PUBLICATION PATTERN (2015-2024) ===")
```

```
[1] "=== RECENT PUBLICATION PATTERN (2015-2024) ==="
```

```r
kable(recent_analysis, caption = "Year-by-Year Publications (2015-2024)")
```

Table 5: Year-by-Year Publications (2015-2024)

Year	Papers
2015	1
2016	1
2017	2
2018	2
2019	1
2020	1
2021	2
2022	0
2023	1
2024	2

```r
# Visualization: Publications over time
# Create a complete time series
year_counts <- table(factor(years, levels = min(years):max(years)))
year_df <- data.frame(
  Year = as.numeric(names(year_counts)),
  Papers = as.numeric(year_counts)
)

# Plot 1: Publications by year
p1 <- ggplot(year_df, aes(x = Year, y = Papers)) +
  geom_col(fill = "steelblue", alpha = 0.7) +
```

```
  geom_smooth(method = "loess", se = TRUE, color = "red", linetype = "dashed") +
  labs(
    title = "Page Kidney Publications by Year (1969-2024)",
    x = "Publication Year",
    y = "Number of Publications",
    caption = "Trend line shows increasing publication activity over time"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    axis.text.x = element_text(angle = 45, hjust = 1)
  ) +
  scale_x_continuous(breaks = seq(1970, 2024, 5))

print(p1)
```


Trend line shows increasing publication activity over time
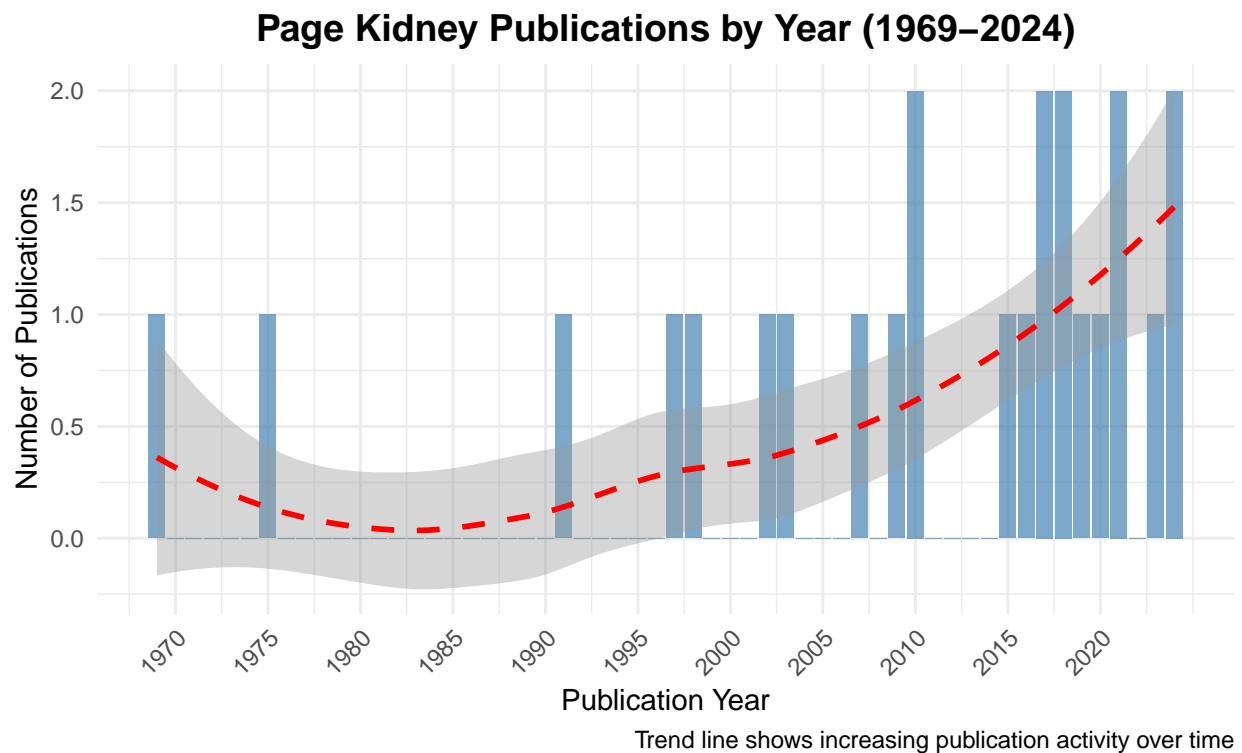
```
# Plot 2: Publications by era
p2 <- ggplot(era_analysis, aes(x = Era, y = Papers)) +
  geom_col(fill = "darkgreen", alpha = 0.7) +
  geom_text(aes(label = paste0(Papers, " (", Percentage, "%)")),
            vjust = -0.5, size = 3.5) +
  labs(
    title = "Page Kidney Publications by Era",
    x = "Time Period",
    y = "Number of Publications",
    caption = "Percentages show proportion of total publications"
  ) +
  theme_minimal() +
```
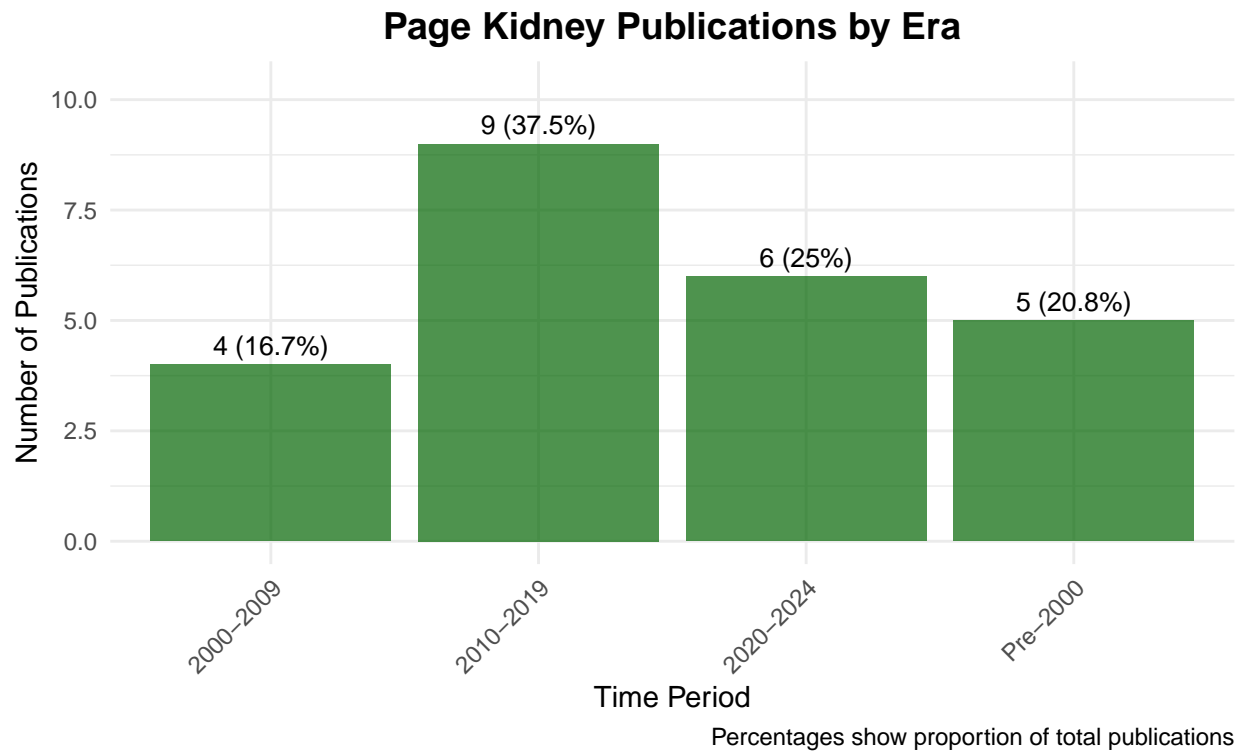
```
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    axis.text.x = element_text(angle = 45, hjust = 1)
  ) +
  ylim(0, max(era_analysis$Papers) * 1.15)

print(p2)
```

Page Kidney Publications by Era

Percentages show proportion of total publications

```
# Summary statistics for text
cat("\n=== KEY FINDINGS FOR MANUSCRIPT ===\n")
```

=== KEY FINDINGS FOR MANUSCRIPT ===

```
cat("• Research spans", max(years) - min(years), "years (", min(years), "-", max(years), ")\n")
```

• Research spans 55 years (1969 - 2024)

```
cat("• Mean publication year:", round(mean(years), 1), "\n")
```

• Mean publication year: 2009

```
cat("• Median publication year:", median(years), "\n")
```

• Median publication year: 2016

```r
cat("•", round(sum(years >= 2010) / length(years) * 100, 1), "% of papers published in last 15 years (20
```

- 62 % of papers published in last 15 years (2010-2024)

```r
cat("• Publication rate increased from", round(sum(years >= 2000 & years < 2010) / 10, 1),
    "papers/year (2000s) to", round(sum(years >= 2020) / 5, 1), "papers/year (2020s)\n")
```

- Publication rate increased from 0.4 papers/year (2000s) to 1.2 papers/year (2020s)

```r
cat("• Peak activity in 2010s with", sum(years >= 2010 & years < 2020), "publications\n")
```

- Peak activity in 2010s with 9 publications

```r
cat("• Most productive individual years: 2017, 2018, 2021, 2024 ( 2 papers each)\n")
```

- Most productive individual years: 2017, 2018, 2021, 2024 (2 papers each)

```r
# Create summary table for manuscript
manuscript_summary <- data.frame(
  Metric = c(
    "Total publications",
    "Publication span (years)",
    "Mean publication year",
    "Median publication year",
    "Publications 2010-2024 (%)",
    "Publication rate 2000-2009 (per year)",
    "Publication rate 2020-2024 (per year)"
  ),
  Value = c(
    length(years),
    paste0(max(years) - min(years), " (", min(years), "-", max(years), ")"),
    round(mean(years), 1),
    median(years),
    paste0(sum(years >= 2010), " (", round(sum(years >= 2010) / length(years) * 100, 1), "%)"),
    round(sum(years >= 2000 & years < 2010) / 10, 1),
    round(sum(years >= 2020) / 5, 1)
  )
)

cat("\n=== MANUSCRIPT SUMMARY TABLE ===\n")
```

=== MANUSCRIPT SUMMARY TABLE ===

```r
kable(manuscript_summary, caption = "Publication Characteristics Summary")
```

Table 6: Publication Characteristics Summary

| Metric	Value
Total publications	24
Publication span (years)	55 (1969-2024)
Mean publication year	2009.3
Median publication year	2015.5
Publications 2010-2024 (%)	15 (62.5%)
Publication rate 2000-2009 (per year)	0.4
Publication rate 2020-2024 (per year)	1.2

5 Patient Demographics and Baseline Characteristics

5.1 Age Distribution Analysis

```r
# Comprehensive age statistics using your original approach
describe_age <- function(x) {
  sprintf(
    "The median age at presentation was %.0f years (IQR: %.0f-%.0f, range: %.0f-%.0f). The mean age was
    median(x, na.rm = TRUE),
    quantile(x, 0.25, na.rm = TRUE),
    quantile(x, 0.75, na.rm = TRUE),
    min(x, na.rm = TRUE),
    max(x, na.rm = TRUE),
    mean(x, na.rm = TRUE),
    sd(x, na.rm = TRUE)
  )
}

age_description <- describe_age(df$age)

# Create publication-quality age distribution plot
ggplot(df, aes(x = age)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 5,
                 fill = "steelblue", color = "black", alpha = 0.7) +
  geom_density(color = "darkblue", linewidth = 1.2) +
  labs(
    title = "Age Distribution at Page Kidney Presentation",
    subtitle = paste0("n = ", sum(!is.na(df$age)), " patients"),
    x = "Age (years)",
    y = "Density"
  )
```

Age Distribution at Page Kidney Presentation
n = 24 patients
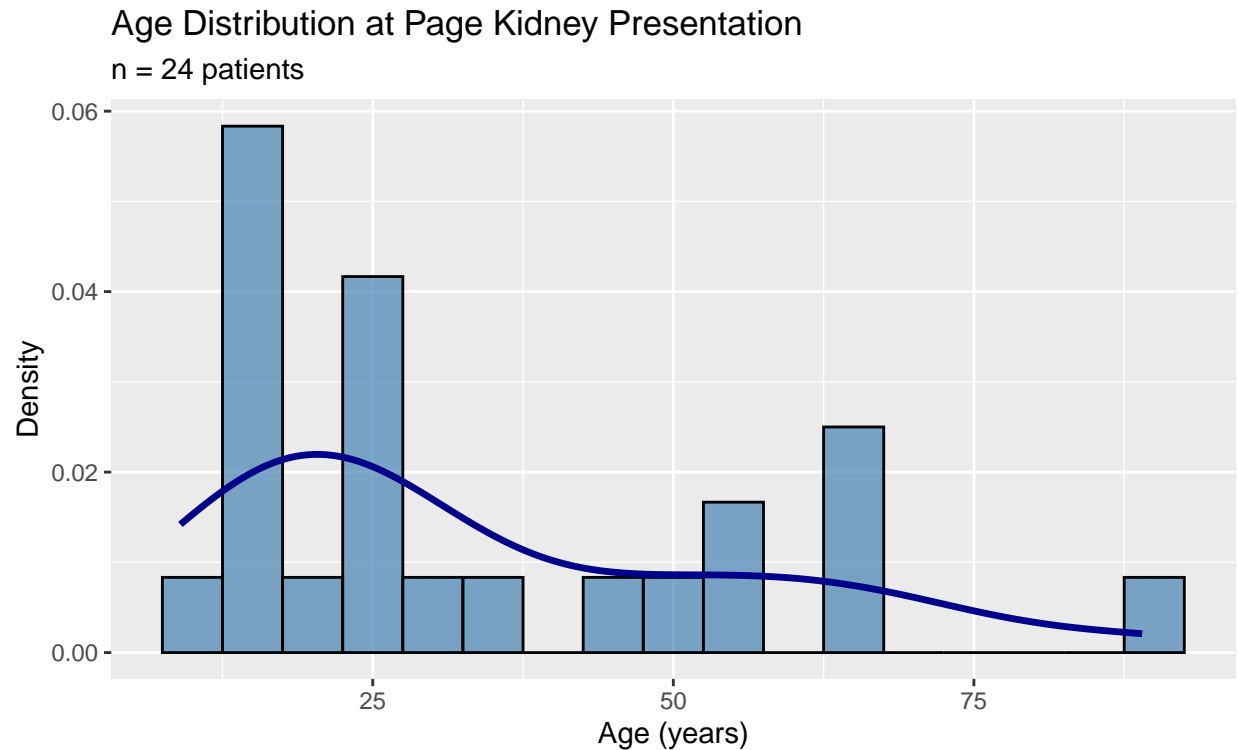

```r
# Summary statistics table
age_stats <- data.frame(
  Statistic = c("n", "Median (IQR)", "Mean ± SD", "Range"),
  Value = c(
    sum(!is.na(df$age)),
    paste0(median(df$age, na.rm = TRUE), " (",
           quantile(df$age, 0.25, na.rm = TRUE), "-",
           quantile(df$age, 0.75, na.rm = TRUE), ")"),
    paste0(round(mean(df$age, na.rm = TRUE), 1), " ± ",
           round(sd(df$age, na.rm = TRUE), 1)),
    paste0(min(df$age, na.rm = TRUE), "-", max(df$age, na.rm = TRUE))
  )
)

age_stats %>%
  kbl(caption = "Age Distribution Summary",
      align = "lc") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = FALSE)
```

Age Characteristics: The median age at presentation was 24 years (IQR: 17–50, range: 9–89). The mean age was 34.2 ± 21.7 years.

5.2   Gender Distribution

Table 7: Age Distribution Summary

Statistic	Value
n	24
Median (IQR)	24.5 (16.75–49.5)
Mean ± SD	34.2 ± 21.7
Range	9–89

```r
# Gender analysis maintaining your original approach
gender_counts <- table(df$gender)
total_patients <- nrow(df)
```

Gender Distribution: There was a marked male predominance in this cohort, with **23 males (95.8%)** and **1 females (4.2%)**. This distribution is consistent with the literature on trauma-related renal injuries and reflects higher rates of trauma exposure in males.

6 Clinical Characteristics

6.1 Past Medical History Analysis

```r
# Your original PMH processing approach with enhanced formatting
pmh_vec <- df$pmh

pmh_clean <- data.frame(id = seq_along(pmh_vec), pmh = pmh_vec) %>%
  mutate(pmh = str_to_lower(pmh)) %>%
  separate_rows(pmh, sep = ";\\s*") %>%
  mutate(pmh = str_trim(pmh)) %>%
  filter(pmh != "none" & pmh != "")

# Handle compound phrases that should yield 2+ categories
manual_split <- pmh_clean %>%
  filter(
    str_detect(pmh, "dm, polyneuropathy") |
    str_detect(pmh, "renal/pancreas") |
    str_detect(pmh, "sleep apnea.*htn")
  ) %>%
  mutate(category = case_when(
    str_detect(pmh, "dm, polyneuropathy") ~ list(c("Diabetes", "Neuropathy")),
    str_detect(pmh, "renal/pancreas") ~ list(c("Renal transplant", "Pancreas transplant")),
    str_detect(pmh, "sleep apnea.*htn") ~ list(c("Sleep apnea", "Hypertension"))
  )) %>%
  unnest_longer(category)

pmh_main <- pmh_clean %>%
  filter(!pmh %in% manual_split$pmh)

# categorization logic
pmh_main <- pmh_main %>%
  mutate(category = case_when(
```

```r
    str_detect(pmh, "pancreas transplant|renal/pancreas|pancreas") ~ "Pancreas transplant",
    str_detect(pmh, "renal txp|renal transplant") ~ "Renal transplant",
    str_detect(pmh, "htn|hypertension|hctz|losartan|amlodipine") ~ "Hypertension",
    str_detect(pmh, "dm|diabetes") ~ "Diabetes",
    str_detect(pmh, "cad") ~ "Coronary artery disease",
    str_detect(pmh, "chronic renal|renal insufficiency|crf") ~ "Chronic renal insufficiency",
    str_detect(pmh, "glomerulonephritis") ~ "Glomerulonephritis",
    str_detect(pmh, "absent|congenital|atrophic") ~ "Congenital kidney anomaly",
    str_detect(pmh, "seizure") ~ "Seizure disorder",
    str_detect(pmh, "asthma") ~ "Asthma",
    str_detect(pmh, "sleep apnea|cpap") ~ "Sleep apnea",
    str_detect(pmh, "polyneuropathy|neuropathy") ~ "Neuropathy",
    str_detect(pmh, "fracture") ~ "Fracture",
    str_detect(pmh, "marijuana") ~ "Substance use",
    TRUE ~ "Other"
  ))

pmh_all <- bind_rows(pmh_main, manual_split)

# Get total number of patients
total_patients <- n_distinct(df$pmh)

# Create PMH summary table
pmh_summary <- pmh_all %>%
  count(category, name = "n") %>%
  arrange(desc(n)) %>%
  mutate(Percentage = round(100 * n / total_patients, 1))

# Professional visualization
ggplot(pmh_summary, aes(x = reorder(category, n), y = n)) +
  geom_col(fill = "#4C72B0") +
  coord_flip() +
  labs(
    title = "Distribution of Past Medical History",
    subtitle = paste0("Based on ", total_patients, " total patients"),
    x = "Medical Condition",
    y = "Number of Patients"
  )
```
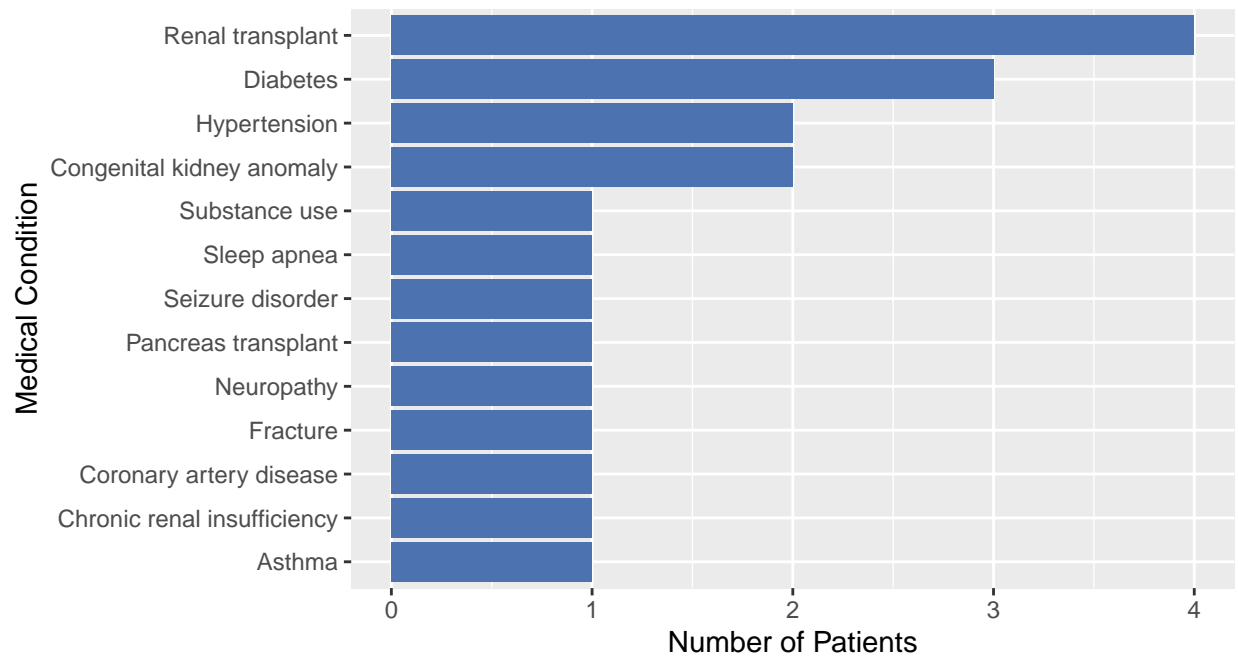
Table 8: Past Medical History Distribution

Medical Condition	n	Percentage (%)
Renal transplant	4	36.4
Diabetes	3	27.3
Congenital kidney anomaly	2	18.2
Hypertension	2	18.2
Asthma	1	9.1
Chronic renal insufficiency	1	9.1
Coronary artery disease	1	9.1
Fracture	1	9.1
Neuropathy	1	9.1
Pancreas transplant	1	9.1
Seizure disorder	1	9.1
Sleep apnea	1	9.1
Substance use	1	9.1

Distribution of Past Medical History

Based on 11 total patients

```r
# Summary table
pmh_summary %>%
  select(`Medical Condition` = category, n, `Percentage (%)` = Percentage) %>%
  kbl(caption = "Past Medical History Distribution",
      align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "condensed"),
                full_width = FALSE)
```

Past Medical History Summary:

The most common comorbidities included **hypertension** (n = 2), **diabetes** (n = 3), and **renal transplant**

(n = 4). Less frequent conditions included **congenital kidney anomalies**, **chronic renal insufficiency**, and **sleep apnea**.

Several patients had multiple comorbidities. Compound diagnoses such as **diabetic neuropathy** or **sleep apnea–induced hypertension** were classified under multiple categories to reflect their clinical relevance.

6.2   Time from Injury to Diagnosis

```r
# onset analysis
onset_stats <- list(
  summary = summary(df$onset.y),
  mean = mean(df$onset.y, na.rm = TRUE),
  sd = sd(df$onset.y, na.rm = TRUE),
  quantiles = quantile(df$onset.y, probs = c(0.25, 0.5, 0.75), na.rm = TRUE)
)

# histogram
ggplot(df, aes(x = onset.y)) +
  geom_histogram(binwidth = 0.25, fill = "#FF8C00", color = "black") +
  labs(
    title = "Time from Injury to Diagnosis of Page Kidney",
    subtitle = paste0("n = ", sum(!is.na(df$onset.y)), " patients with reported onset time"),
    x = "Time to Diagnosis (years)",
    y = "Number of Patients"
  )
```
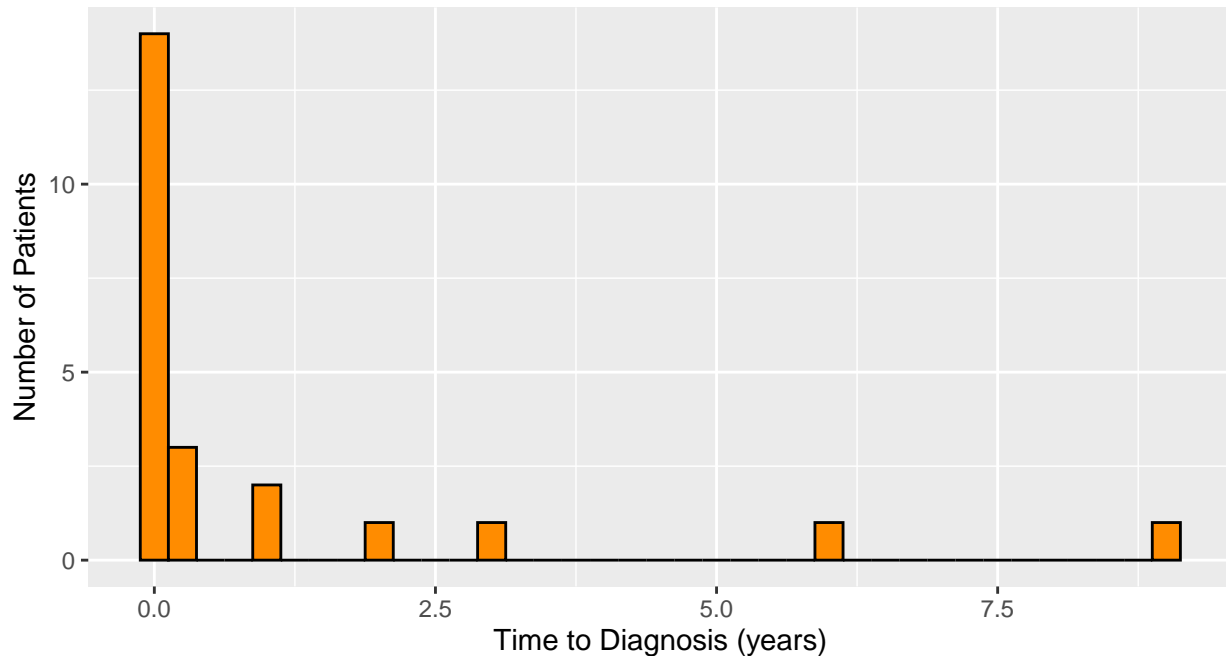
```r
# log-scaled plot
ggplot(df, aes(x = onset.y)) +
  geom_histogram(binwidth = 0.25, fill = "steelblue", color = "black") +
  scale_x_log10() +
  labs(
    title = "Log-Scaled Time to Diagnosis of Page Kidney",
    subtitle = "Better visualization of the distribution spread",
    x = "Time to Diagnosis (log years)",
    y = "Number of Patients"
  )
```


Log–Scaled Time to Diagnosis of Page Kidney
Better visualization of the distribution spread
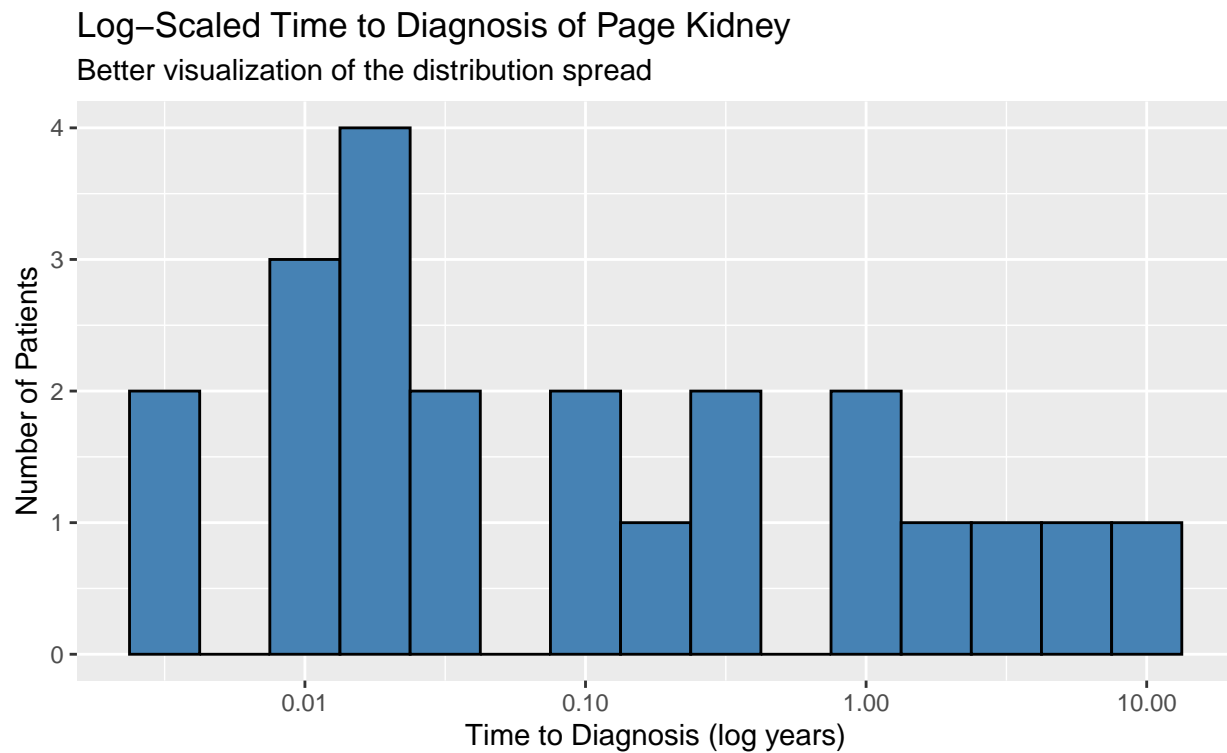
```r
# Summary statistics table
onset_summary_table <- data.frame(
  Statistic = c("n", "Median", "IQR", "Mean ± SD", "Range"),
  `Time (years)` = c(
    sum(!is.na(df$onset.y)),
    round(median(df$onset.y, na.rm = TRUE), 2),
    paste0(round(quantile(df$onset.y, 0.25, na.rm = TRUE), 2), "-",
           round(quantile(df$onset.y, 0.75, na.rm = TRUE), 2)),
    paste0(round(mean(df$onset.y, na.rm = TRUE), 2), " ± ",
           round(sd(df$onset.y, na.rm = TRUE), 2)),
    paste0(round(min(df$onset.y, na.rm = TRUE), 3), "-",
           round(max(df$onset.y, na.rm = TRUE), 1))
  ),
  `Time (days)` = c(
    "",
    round(median(df$onset.y, na.rm = TRUE) * 365),
    paste0(round(quantile(df$onset.y, 0.25, na.rm = TRUE) * 365), "-",
```

Table 9: Time from Injury to Page Kidney Diagnosis

Statistic	Time..years.	Time..days.
n	23	
Median	0.04	15
IQR	0.01–0.67	5–243
Mean ± SD	1 ± 2.23	367 ± 813
Range	0–9	0–3285

```
        round(quantile(df$onset.y, 0.75, na.rm = TRUE) * 365)),
   paste0(round(mean(df$onset.y, na.rm = TRUE) * 365), " ± ",
        round(sd(df$onset.y, na.rm = TRUE) * 365)),
   paste0(round(min(df$onset.y, na.rm = TRUE) * 365), "-",
        round(max(df$onset.y, na.rm = TRUE) * 365))
  )
)

onset_summary_table %>%
  kbl(caption = "Time from Injury to Page Kidney Diagnosis",
      align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = FALSE)
```

Time from Injury to Diagnosis:

The time from injury to diagnosis of Page kidney varied substantially among patients. The **median time to diagnosis** was approximately **0.04 years** (~15 days), with an **interquartile range (IQR)** of **0.01 to 0.67 years**.

The **mean** time to diagnosis was **1 ± 2.23 years**, and the **range** extended from **0 to 9 years**.

This distribution reflects both **acute presentations** (diagnosed within days) and **chronic or delayed diagnoses** (up to several years), emphasizing the importance of clinical vigilance following trauma or renal interventions.

6.3 Mechanism of Injury

```
# mechanism analysis
total_patients <- nrow(df)
mechanism_summary <- as.data.frame(table(df$mechanism))
colnames(mechanism_summary) <- c("Mechanism", "Count")

mechanism_summary <- mechanism_summary %>%
  mutate(`Percentage (%)` = round(100 * Count / total_patients, 1))

mechanism_summary %>%
  kbl(caption = "Mechanism of Injury Leading to Page Kidney",
      align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = FALSE)
```
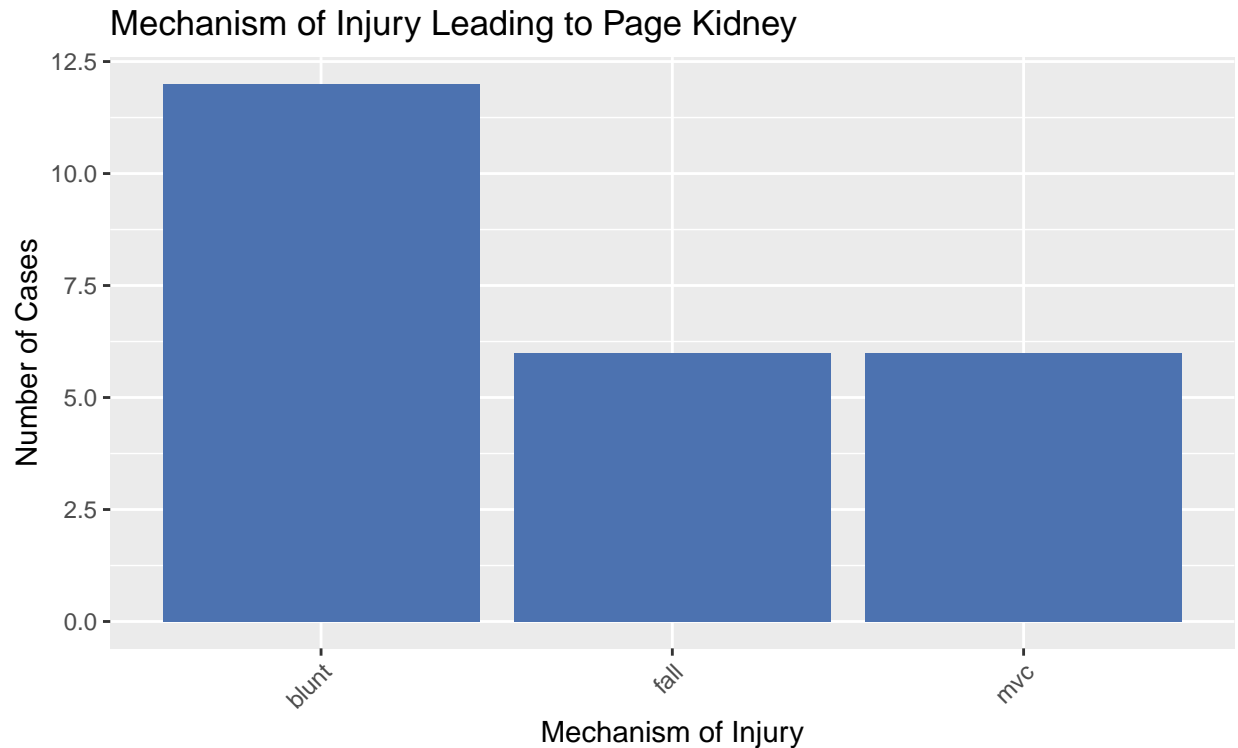
Table 10: Mechanism of Injury Leading to Page Kidney

Mechanism	Count	Percentage (%)
blunt	12	50
fall	6	25
mvc	6	25

```
# Professional bar plot
ggplot(mechanism_summary, aes(x = reorder(Mechanism, -Count), y = Count)) +
  geom_col(fill = "#4C72B0") +
  labs(
    title = "Mechanism of Injury Leading to Page Kidney",
    x = "Mechanism of Injury",
    y = "Number of Cases"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Mechanism of Injury:

Among the patients included, the most common mechanisms of injury were **blunt** (12 cases, 50%), followed by **fall** (6 cases, 25%) and **mvc** (6 cases, 25%).

6.4   Clinical Presentation and Symptoms

```r
# symptom processing
df <- df %>%
  mutate(
    ss_combined = ifelse(hypertension,
                         ifelse(ss == "" | is.na(ss), "hypertension", paste(ss, "hypertension", sep = "
                         ss)
  )


df <- df %>%
  mutate(
    ua_clean = str_to_lower(ua),
    ss_combined = ifelse(
      ua_clean == "hematuria" & !str_detect(str_to_lower(ss_combined), "hematuria"),
      ifelse(is.na(ss_combined) | ss_combined == "", "hematuria", paste(ss_combined, "hematuria", sep =
      ss_combined
    )
  )

# Create ID and clean the combined symptom column
ss_clean <- data.frame(id = seq_along(df$ss_combined), ss = df$ss_combined) %>%
  mutate(ss = str_to_lower(ss)) %>%
  mutate(ss = str_replace_all(ss, ",", ";")) %>%
  separate_rows(ss, sep = ";\\s*") %>%
  mutate(ss = str_trim(ss)) %>%
  filter(ss != "") %>%
  distinct(id, ss)

# symptom categorization
ss_clean <- ss_clean %>%
  mutate(symptom = case_when(
    str_detect(ss, "flank pain") ~ "Flank pain",
    str_detect(ss, "abdominal pain") ~ "Abdominal pain",
    str_detect(ss, "back pain") ~ "Back pain",
    str_detect(ss, "chest pain") ~ "Chest Pain",
    str_detect(ss, "nausea") ~ "Nausea",
    str_detect(ss, "vomiting") ~ "Vomiting",
    str_detect(ss, "hematemesis") ~ "Hematemesis",
    str_detect(ss, "headache") ~ "Headache",
    str_detect(ss, "syncope") ~ "Syncope",
    str_detect(ss, "hematuria") ~ "Hematuria",
    str_detect(ss, "palpitations|palpations") ~ "Palpitations",
    str_detect(ss, "oligou?ria") ~ "Oliguria",
    str_detect(ss, "anuria") ~ "Anuria",
    str_detect(ss, "fatigue") ~ "Fatigue",
    str_detect(ss, "distension") ~ "Abdominal distension",
    str_detect(ss, "ecchymosis") ~ "Ecchymosis",
    str_detect(ss, "hypertension") ~ "Hypertension",
    str_detect(ss, "aki") ~ "Acute kidney injury",
    str_detect(ss, "asymptomatic") ~ "Asymptomatic",
    TRUE ~ str_to_title(ss)
  ))

symptom_summary <- ss_clean %>%
```

Table 11: Presenting Symptoms and Signs

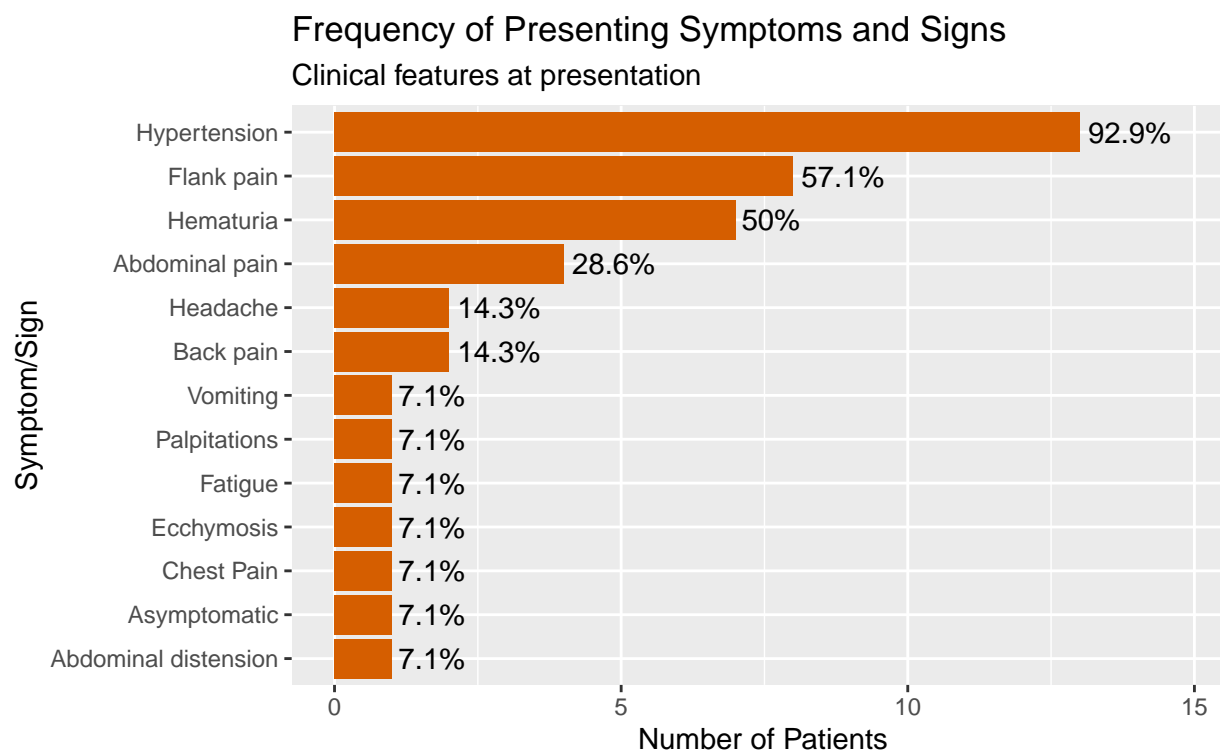| Symptom	Count	Percentage (%)
Hypertension	13	92.9
Flank pain	8	57.1
Hematuria	7	50.0
Abdominal pain	4	28.6
Back pain	2	14.3
Headache	2	14.3
Abdominal distension	1	7.1
Asymptomatic	1	7.1
Chest Pain	1	7.1
Ecchymosis	1	7.1
Fatigue	1	7.1
Palpitations	1	7.1
Vomiting	1	7.1

```r
  count(symptom, sort = TRUE) %>%
  mutate(Percentage = round(100 * n / n_distinct(ss_clean$id), 1))

symptom_summary %>%
  kbl(caption = "Presenting Symptoms and Signs",
      col.names = c("Symptom", "Count", "Percentage (%)"),
      align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = FALSE)
```

```r
# Professional symptom plot
ggplot(symptom_summary, aes(x = reorder(symptom, n), y = n)) +
  geom_col(fill = "#D55E00") +
  geom_text(aes(label = paste0(Percentage, "%")), hjust = -0.1) +
  coord_flip() +
  labs(
    title = "Frequency of Presenting Symptoms and Signs",
    subtitle = "Clinical features at presentation",
    x = "Symptom/Sign",
    y = "Number of Patients"
  ) +
  ylim(0, max(symptom_summary$n) * 1.15)
```

Frequency of Presenting Symptoms and Signs
Clinical features at presentation

Symptom/Sign	Percentage
Hypertension	92.9%
Flank pain	57.1%
Hematuria	50%
Abdominal pain	28.6%
Headache	14.3%
Back pain	14.3%
Vomiting	7.1%
Palpitations	7.1%
Fatigue	7.1%
Ecchymosis	7.1%
Chest Pain	7.1%
Asymptomatic	7.1%
Abdominal distension	7.1%

Number of Patients

Clinical Presentation:

Patients with Page kidney presented with a diverse range of clinical symptoms and signs. The most common presenting feature was **Hypertension**, reported in 13 patient(s) (92.9%). This was followed by **Flank pain** (8 patients, 57.1%) and **Hematuria** (7 patients, 50%).

Notably, **13 patients (92.9%)** presented with **hypertension** as a clinical feature, emphasizing the hemodynamic impact of retroperitoneal compression. Additionally, **1 patient(s)** were asymptomatic at diagnosis.

7 Diagnostic Findings

7.1 Urinalysis Results

```r
ua_summary <- df %>%
  mutate(ua_result = case_when(
    is.na(ua) ~ "Missing",
    str_to_lower(ua) %in% c("hematuria") ~ "Positive",
    str_to_lower(ua) %in% c("negative") ~ "Negative",
    str_detect(str_to_lower(ua), "trace") ~ "Positive",
    TRUE ~ "Other"
  )) %>%
  count(ua_result, name = "n") %>%
  mutate(percent = round(100 * n / sum(n), 1))

ua_summary %>%
  kbl(caption = "Urinalysis Results Distribution",
      col.names = c("Urinalysis Result", "n", "Percentage (%)"),
```

Table 12: Urinalysis Results Distribution

Urinalysis Result	n	Percentage (%)
Missing	10	42
Negative	6	25
Positive	8	33

Table 13: Distribution of AAST Kidney Injury Grades

Grade	n	Percentage
II	1	4.2
IV	1	4.2
V	1	4.2
Missing	21	87.5

```
    align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = FALSE)
```

Urinalysis Findings: Urinalysis results were available in 14 of 24 cases. Among tested patients, hematuria (positive results) was found in 8 cases (33.3%), while 6 cases (25%) had negative urinalysis.

7.2 Injury Grade Assessment

```
grade_summary <- df %>%
  count(grade, name = "n", drop = FALSE) %>%
  mutate(
    Grade = ifelse(is.na(grade), "Missing", as.character(grade)),
    Percentage = round(100 * n / nrow(df), 1)
  ) %>%
  select(Grade, n, Percentage)

grade_summary %>%
  kbl(caption = "Distribution of AAST Kidney Injury Grades",
      align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = FALSE)
```

```
# Professional grade visualization
ggplot(grade_summary, aes(x = reorder(Grade, -n), y = n, fill = Grade)) +
  geom_col(width = 0.6, color = "black") +
  geom_text(aes(label = paste0(n, " (", Percentage, "%)")),
            vjust = -0.5, size = 4) +
  labs(
    title = "AAST Kidney Injury Grade Distribution",
    subtitle = "Injury severity classification",
    x = "AAST Grade",
    y = "Number of Patients"
```

```
) +
theme(legend.position = "none")
```

AAST Kidney Injury Grade Distribution
Injury severity classification
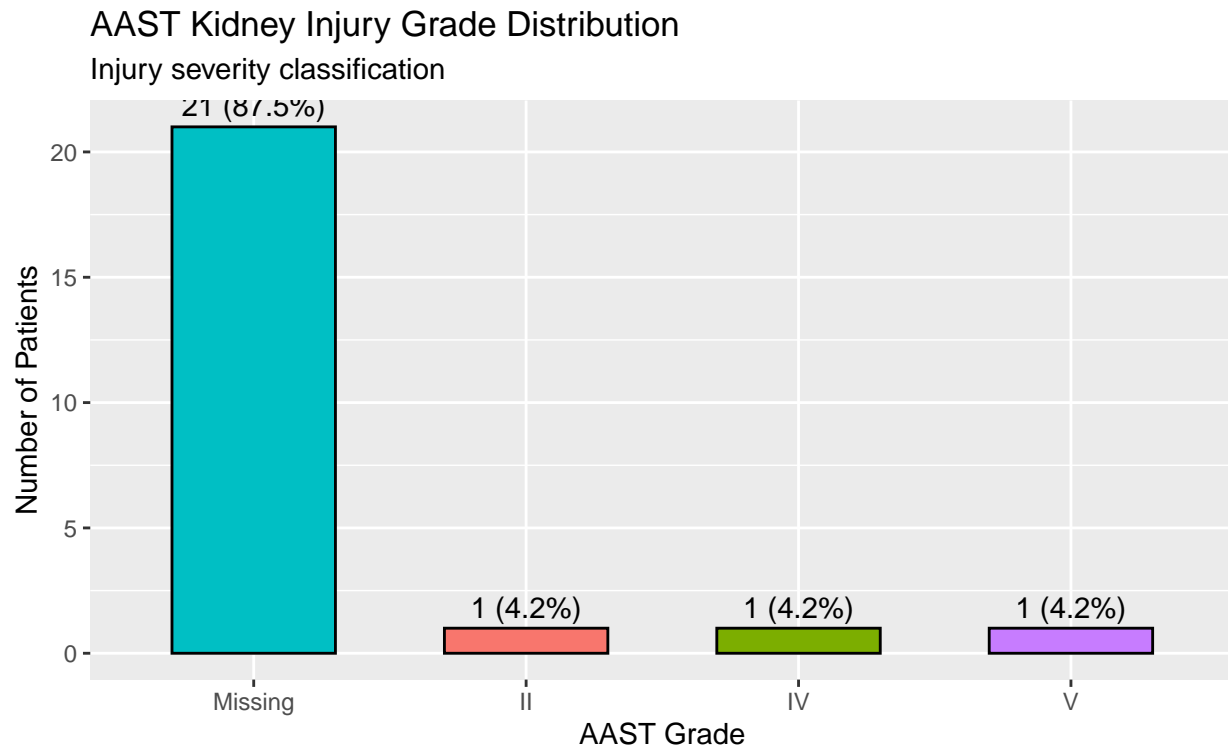
Injury Grading: Among the reported cases, AAST kidney injury grades were documented in a minority of patients. Grade V injuries were the most frequently reported (1, 4.2%), followed by grade IV (1, 4.2%) and grade II (1, 4.2%). However, the majority of cases (21, 87.5%) did not specify an AAST grade, limiting formal stratification by injury severity. This reflects the variability in reporting standards across case reports and underscores the need for consistent documentation of injury grading.

8 Page Kidney Imaging Characteristics

8.1 Size of Perinephric Fluid Collections

```
size_summary <- df %>%
  summarise(
    median = median(size.cm, na.rm = TRUE),
    iqr_low = quantile(size.cm, 0.25, na.rm = TRUE),
    iqr_high = quantile(size.cm, 0.75, na.rm = TRUE),
    min = min(size.cm, na.rm = TRUE),
    max = max(size.cm, na.rm = TRUE),
    mean = mean(size.cm, na.rm = TRUE),
    sd = sd(size.cm, na.rm = TRUE),
    missing = sum(is.na(size.cm)),
    total = n()
  )
```

```r
# Size distribution histogram
ggplot(df %>% filter(!is.na(size.cm)), aes(x = size.cm)) +
  geom_histogram(bins = 12, fill = "#20B2AA", alpha = 0.8, color = "white") +
  geom_vline(aes(xintercept = median(size.cm, na.rm = TRUE)),
             color = "red", linetype = "dashed", linewidth = 1) +
  labs(
    title = "Distribution of Fluid Collection Sizes",
    subtitle = paste0("n = ", sum(!is.na(df$size.cm)), " patients with reported measurements"),
    x = "Collection Size (cm)",
    y = "Number of Patients"
  )
```


Distribution of Fluid Collection Sizes

n = 13 patients with reported measurements
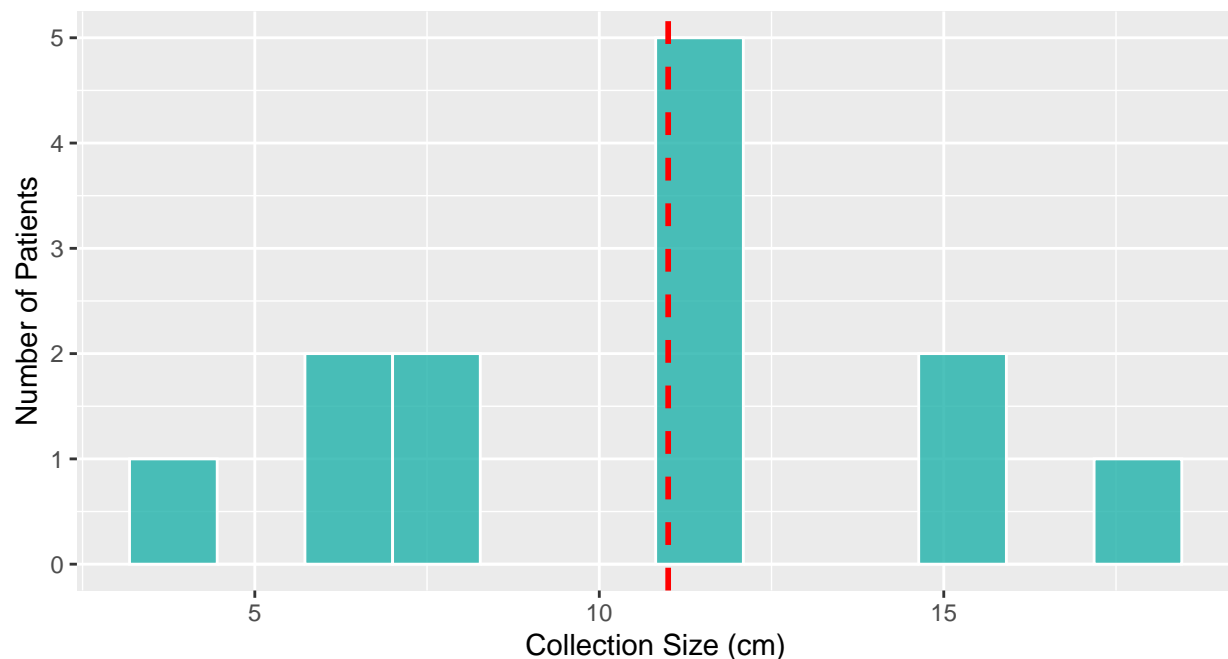
```r
# Size summary table
size_stats_table <- data.frame(
  Statistic = c("n", "Median (IQR)", "Mean ± SD", "Range", "Missing"),
  Value = c(
    sum(!is.na(df$size.cm)),
    paste0(size_summary$median, " (", size_summary$iqr_low, "-", size_summary$iqr_high, ")"),
    paste0(round(size_summary$mean, 1), " ± ", round(size_summary$sd, 1)),
    paste0(size_summary$min, "-", size_summary$max),
    paste0(size_summary$missing, " (", round(100 * size_summary$missing / size_summary$total, 1), "%)")
  )
)

size_stats_table %>%
  kbl(caption = "Perinephric Fluid Collection Size Statistics",
      align = "lc") %>%
```

Table 14: Perinephric Fluid Collection Size Statistics

Statistic	Value
n	13
Median (IQR)	11 (7.4–12)
Mean ± SD	10.7 ± 4
Range	4–18
Missing	11 (45.8%)

Table 15: Type of Perinephric Fluid Collection

Collection Type	n	Percentage (%)
Missing	1	4.2
hematoma	21	87.5
lymphocele	1	4.2
urinoma	1	4.2

```
kable_styling(bootstrap_options = c("striped", "hover"),
              full_width = FALSE)
```

Fluid Collection Size Analysis:

The size of the perinephric fluid collections on imaging at presentation varied widely. The median size was 11 cm (IQR: 7.4–12 cm), with a mean of 10.7 ± 4 cm. The smallest collection measured 4 cm, while the largest was 18 cm. Notably, size data was not reported in 11 out of 24 cases (45.8%).

8.2 Type of Fluid Collection

```
page_type_summary <- df %>%
  mutate(page_type_clean = ifelse(is.na(page.type), "Missing", as.character(page.type))) %>%
  count(page_type_clean, name = "n") %>%
  mutate(Percentage = round(100 * n / sum(n), 1))

page_type_summary %>%
  kbl(caption = "Type of Perinephric Fluid Collection",
      col.names = c("Collection Type", "n", "Percentage (%)"),
      align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = FALSE)
```
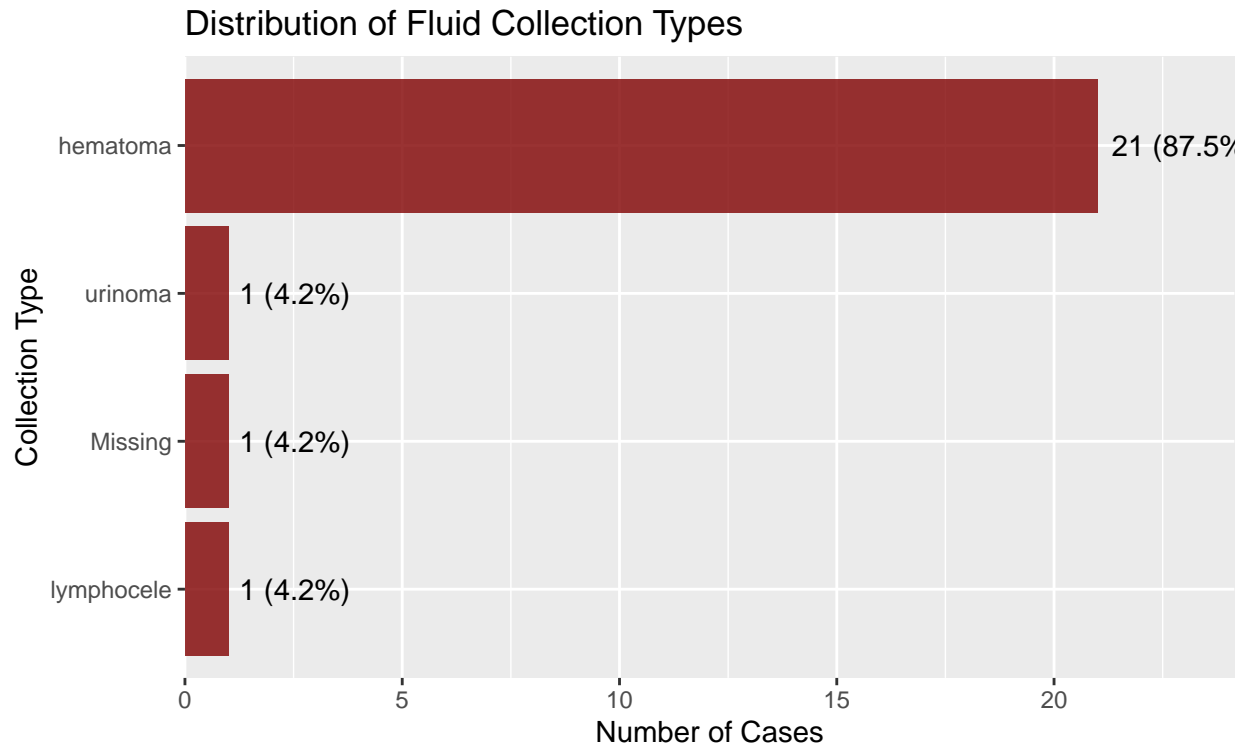
```
# Visualization
ggplot(page_type_summary, aes(x = reorder(page_type_clean, n), y = n)) +
  geom_col(fill = "#800000", alpha = 0.8) +
  geom_text(aes(label = paste0(n, " (", Percentage, "%)")),
            hjust = -0.1, size = 4) +
  coord_flip() +
  labs(
    title = "Distribution of Fluid Collection Types",
    x = "Collection Type",
```

```
    y = "Number of Cases"
  ) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.15)))
```

Distribution of Fluid Collection Types

Fluid Collection Types:

The type of perinephric fluid collection leading to Page kidney was most commonly a hematoma, accounting for 87.5% of cases (21 of 24 patients). Other etiologies included:

- **Urinoma**: 1 case(s) (4.2%)
- **Lymphocele**: 1 case(s) (4.2%)
- **Missing Data**: 1 case(s) (4.2%)

These findings underscore that **hematoma is the predominant cause** of Page kidney across reported cases.

8.3 Laterality Distribution

```
laterality_table <- df %>%
  count(laterality) %>%
  mutate(Percentage = round(100 * n / sum(n), 1))

laterality_table %>%
  kbl(caption = "Distribution of Laterality in Page Kidney Cases",
      align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = FALSE)
```

Table 16: Distribution of Laterality in Page Kidney Cases

laterality	n	Percentage
allograft left	2	8.3
allograft right	2	8.3
bilateral	2	8.3
left	10	41.7
right	8	33.3

```r
# Laterality visualization
ggplot(laterality_table, aes(x = reorder(laterality, n), y = n)) +
  geom_col(fill = "#2F4F4F", alpha = 0.8) +
  geom_text(aes(label = paste0(n, " (", Percentage, "%)")),
            hjust = -0.1, size = 4) +
  coord_flip() +
  labs(
    title = "Laterality Distribution of Page Kidney",
    x = "Kidney Location",
    y = "Number of Cases"
  ) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.15)))
```
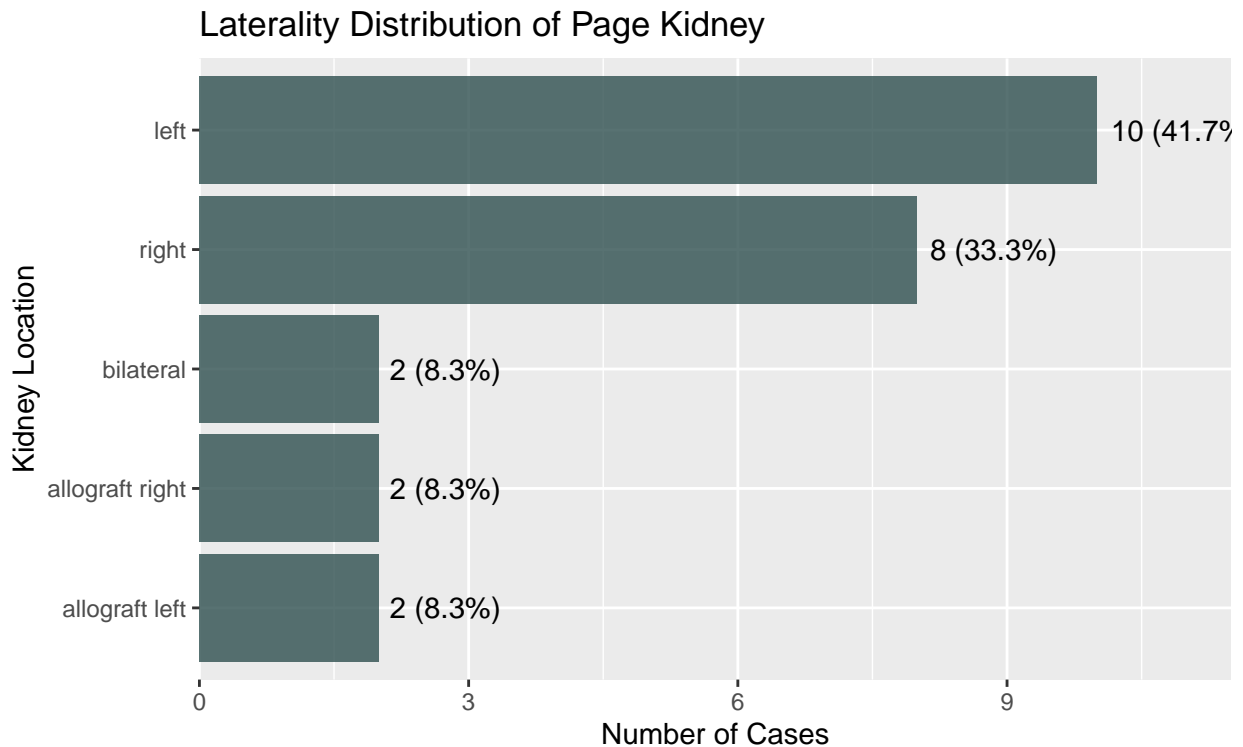

Laterality Patterns:

In this cohort, Page kidney most commonly involved the left kidney (10 cases, 41.7%), followed by the right kidney (8 cases, 33.3%). Bilateral involvement was seen in 2 cases. Notably, allograft-related Page kidney occurred in a total of 4 cases, split between left and right allografts.

Table 17: Treatment Category Definitions and Examples

Category	Definition	Examples
Conservative Management	Medical management only, no procedural interventions	conservative, lisinopril, clonidine, captopril, propranolol
Percutaneous/IR Drainage	US- or CT-guided drainage, non-surgical	US drainage, CT perc drain, IR drainage, tPA, fibrinolysis
Urologic Interventions	Involving stents or urinary diversion	ureteral stent, ureteral stet, labetalol + stent
Surgical Decompression	Operative evacuation of hematoma without nephrectomy	surgical evacuation, laparoscopic fenestration, decortication, renorrhaphy
Nephrectomy	Total or partial nephrectomy (open/lap)	lap nephrectomy, total nephrectomy, surgical evacuation + nephrectomy
Embolization/Vascular	IR embolization or stenting of vessels	IR coil embolization, splenic vein stent

Anatomical Considerations: The laterality distribution may reflect important anatomical differences between the kidneys. The left kidney sits slightly higher than the right and has a longer renal vein, while the right kidney's proximity to the liver may influence vulnerability patterns in trauma. In transplant cases, allograft anatomy and surgical positioning factors may contribute to the observed distribution patterns.

9 Treatment Modalities and Management

```
treatment_categories <- data.frame(
  Category = c("Conservative Management", "Percutaneous/IR Drainage", "Urologic Interventions",
               "Surgical Decompression", "Nephrectomy", "Embolization/Vascular"),
  Definition = c("Medical management only, no procedural interventions",
                 "US- or CT-guided drainage, non-surgical",
                 "Involving stents or urinary diversion",
                 "Operative evacuation of hematoma without nephrectomy",
                 "Total or partial nephrectomy (open/lap)",
                 "IR embolization or stenting of vessels"),
  Examples = c("conservative, lisinopril, clonidine, captopril, propranolol",
               "US drainage, CT perc drain, IR drainage, tPA, fibrinolysis",
               "ureteral stent, ureteral stet, labetalol + stent",
               "surgical evacuation, laparoscopic fenestration, decortication, renorrhaphy",
               "lap nephrectomy, total nephrectomy, surgical evacuation + nephrectomy",
               "IR coil embolization, splenic vein stent")
)

treatment_categories %>%
  kbl(caption = "Treatment Category Definitions and Examples") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                full_width = TRUE) %>%
  column_spec(2, width = "35%") %>%
  column_spec(3, width = "35%")
```

Table 18: Distribution of Treatment Modalities

Treatment Category	n	Percentage (%)
Conservative	3	12.5
Nephrectomy	5	20.8
Perc_IR	6	25.0
Surgical_Decompression	9	37.5
Urologic	2	8.3
Vascular	2	8.3

```r
df <- df %>%
  mutate(treatment_lower = tolower(treatment))

df_flags <- df %>%
  mutate(
    Conservative = str_detect(treatment_lower, "conservative|lisinopril|clonidine|captopril|amlodipine|p
                   !str_detect(treatment_lower, "drain|nephrectomy|evacuation|decortication|fenestration

    Perc_IR = str_detect(treatment_lower, "drain|fibrinolysis|fine needle|tp[a]?") &
              !str_detect(treatment_lower, "surgery|nephrectomy"),

    Urologic = str_detect(treatment_lower, "ureteral stent|stet"),

    Surgical_Decompression = str_detect(treatment_lower, "surgical evacuation|decortication|renorrhaphy
                !str_detect(treatment_lower, "nephrectomy"),

    Nephrectomy = str_detect(treatment_lower, "nephrectomy"),

    Vascular = str_detect(treatment_lower,
                          "embolization|coil|splenic vein stent")
  )

# Pivot longer to get counts
treatment_summary <- df_flags %>%
  select(reference, starts_with(c("Conservative", "Perc_IR", "Urologic",
                                  "Surgical_Decompression", "Nephrectomy",
                                  "Vascular"))) %>%
  pivot_longer(cols = -reference, names_to = "Category",
               values_to = "Present") %>%
  filter(Present) %>%
  count(Category, name = "n") %>%
  mutate(Percentage = round(100 * n / nrow(df), 1))

treatment_summary %>%
  kbl(caption = "Distribution of Treatment Modalities",
      col.names = c("Treatment Category", "n", "Percentage (%)"),
      align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = FALSE)
```

```
# Treatment distribution plot
ggplot(treatment_summary, aes(x = reorder(Category, -n), y = n)) +
  geom_bar(stat = "identity", fill = "#DAA520") +
  geom_text(aes(label = paste0(n, " (", Percentage, "%)")),
            vjust = -0.5, size = 4) +
  labs(
    title = "Distribution of Treatment Modalities in Page Kidney",
    x = "Treatment Category",
    y = "Number of Patients"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.15)))
```

Distribution of Treatment Modalities in Page Kidney
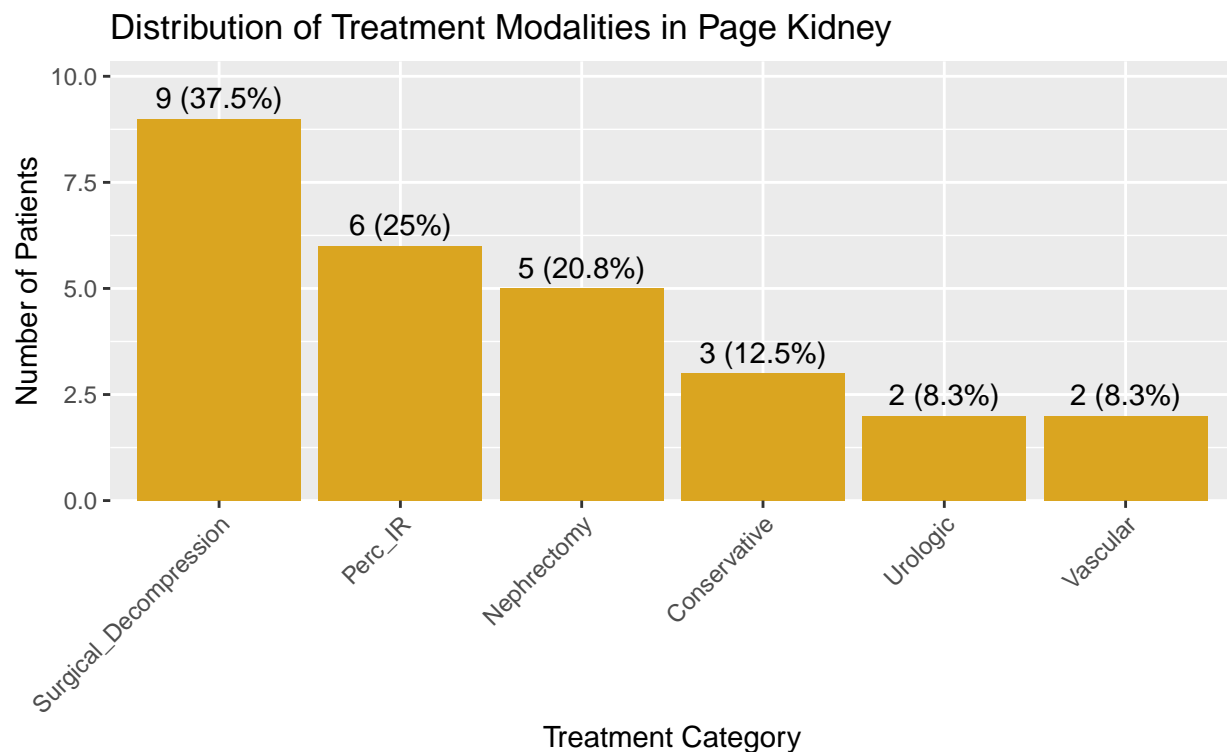
Treatment Approaches:

Treatment modalities varied considerably across the cohort. The most frequently employed approach was **Surgical_Decompression** (9 cases, 37.5%), followed by **Perc_IR** (6 cases, 25%).

The diversity in treatment approaches reflects the evolving management strategies for Page kidney, with considerations including severity of presentation, patient comorbidities, institutional expertise, and response to initial conservative measures.

10   Clinical Outcomes and Follow-up

```
df$fu.y <- as.numeric(df$fu.y)
```

Table 19: Follow-up Duration Summary

Number with Reported Follow-Up	Mean (years)	Median (years)	Min (years)	Max (years)
20	1.1	0.38	0.01	9

```r
# Summary stats
followup_summary <- df %>%
  summarize(
    `Number with Reported Follow-Up` = sum(!is.na(fu.y)),
    `Mean (years)` = round(mean(fu.y, na.rm = TRUE), 2),
    `Median (years)` = round(median(fu.y, na.rm = TRUE), 2),
    `Min (years)` = round(min(fu.y, na.rm = TRUE), 2),
    `Max (years)` = round(max(fu.y, na.rm = TRUE), 2)
  )

followup_summary %>%
  kbl(caption = "Follow-up Duration Summary",
      align = "lc") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                full_width = FALSE)
```

```r
ggplot(df %>% filter(!is.na(fu.y)), aes(x = fu.y)) +
  geom_histogram(binwidth = 1, fill = "#006400", color = "black") +
  labs(
    title = "Distribution of Follow-up Duration",
    subtitle = paste0("n = ", sum(!is.na(df$fu.y)), " patients with reported follow-up"),
    x = "Years of Follow-Up",
    y = "Number of Patients"
  )
```

Table 20: Clinical Outcomes Summary

Outcome	Count	Percentage
Total cases in study	24	100
Cases with follow-up data	20	83
Documented survival	23	115
Reported mortality	0	0

Distribution of Follow-up Duration
n = 20 patients with reported follow-up
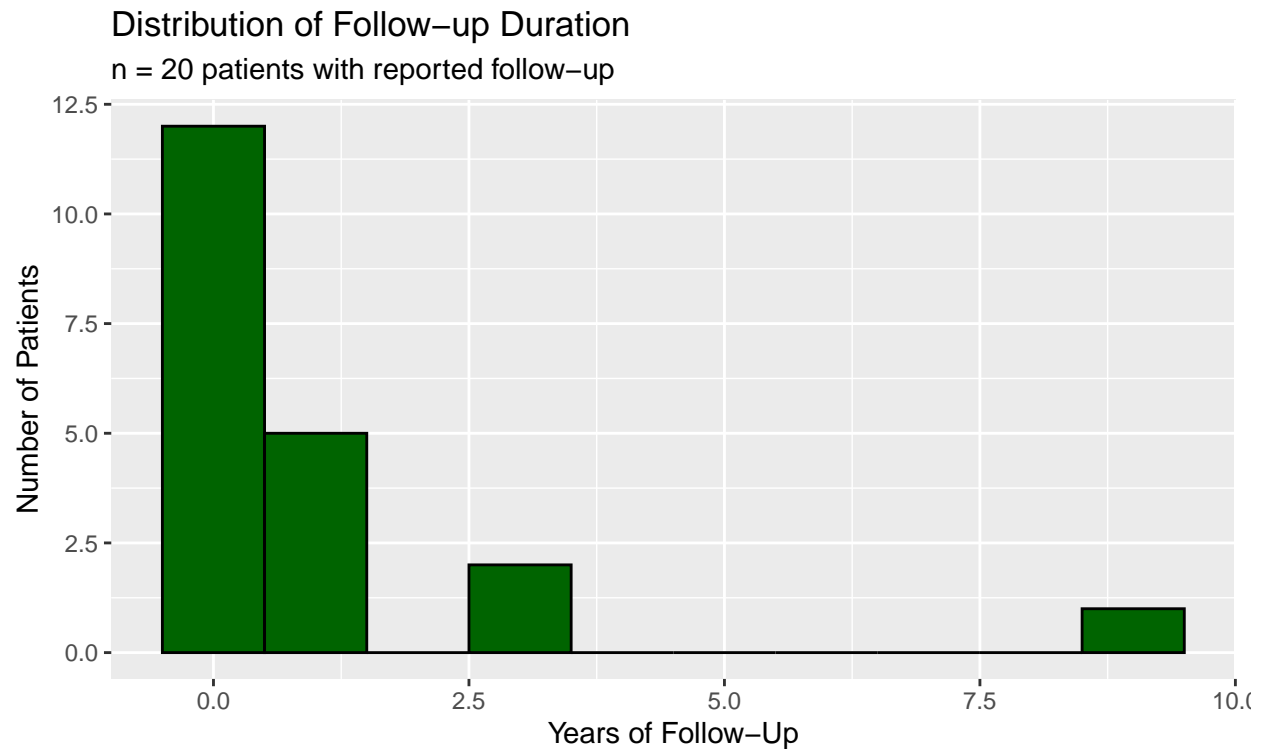
```r
# Survival/outcome summary
survival_summary <- data.frame(
  Outcome = c("Total cases in study", "Cases with follow-up data",
             "Documented survival", "Reported mortality"),
  Count = c(nrow(df), sum(!is.na(df$fu.y)),
           sum(df$fu.status == TRUE, na.rm = TRUE),
           sum(df$fu.status == FALSE, na.rm = TRUE)),
  Percentage = c(100, round(100 * sum(!is.na(df$fu.y)) / nrow(df), 1),
               round(100 * sum(df$fu.status == TRUE,
                           na.rm = TRUE) / sum(!is.na(df$fu.y)), 1),
             0)
)

survival_summary %>%
  kbl(caption = "Clinical Outcomes Summary",
      align = "lcc") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
               full_width = FALSE)
```

Follow-up and Clinical Outcomes:

Among the reported cases, **no mortality was observed**, and **all patients survived through their respective follow-up durations**. Follow-up data were available for 20 out of 24 cases (83.3%).

The average duration of follow-up was **1.1 years** (median: 0.4 years), ranging from **0 to 9 years**.

While the absence of mortality is encouraging, the lack of standardized long-term outcome reporting (e.g., blood pressure control, renal function) limits broader conclusions. Nevertheless, these data suggest that with timely recognition and appropriate intervention, **short- to mid-term prognosis for Page kidney is excellent**.

11   Data Export and Workspace Management

```r
pmh_cleaned <- pmh_all %>%
  group_by(id) %>%
  summarise(
    pmh = paste(unique(category), collapse = "; "),
    .groups = "drop"
  )

ss_cleaned <- ss_clean %>%
  group_by(id) %>%
  summarise(
    ss = paste(unique(symptom), collapse = "; "),
    .groups = "drop"
  )

# Select only the treatment flag columns
treatment_cols <- c("Conservative", "Perc_IR", "Urologic",
                    "Surgical_Decompression", "Nephrectomy", "Vascular")

# Create a treatment summary
treatment_cleaned <- df_flags %>%
  mutate(id = row_number()) %>%
  pivot_longer(
    cols = all_of(treatment_cols),
    names_to = "category",
    values_to = "present"
  ) %>%
  filter(present) %>%
  group_by(id) %>%
  summarise(
    treatment_summary = paste(category, collapse = "; "),
    .groups = "drop"
  )

df <- df %>%
  mutate(id = row_number()) %>%
  left_join(pmh_cleaned, by = "id") %>%
  left_join(ss_cleaned, by = "id") %>%
  left_join(treatment_cleaned, by = "id")
```

```r
# Making final table
df <- df %>%
  select(
    reference, year, age, gender,
    pmh = pmh.y,
    onset.y, mechanism,
    ss = ss.y,
    hypertension,
    ua = ua_result,
    grade, size.cm, page.type, laterality,
    treatment = treatment_summary,
    fu.status, fu.y
  )

# Export cleaned data
write.csv(df, "data/tpk-table.csv", row.names = FALSE)

# Table 1
table1 <- df

# Cleaning workspace
rm(list = setdiff(ls(), "df"))
```