Stratified Table 1 Analysis Examples

zztable1_nextgen

2025-11-13

Contents

1 Introduction

Stratified analysis is a crucial component of epidemiological and clinical research. This vignette demonstrates how to create stratified Table 1 analyses using the `zztable1_nextgen` package. Stratified tables allow you to examine patterns within subgroups of your data, revealing important differences that might be obscured in overall analyses.

2 Clinical Trial Data Example

Let's create a comprehensive clinical trial dataset that includes multiple potential stratification variables.

```
# Create a realistic multi-center clinical trial dataset
set.seed(42)
n <- 300

clinical_data <- data.frame(
```

```r
  # Primary treatment variable
  treatment = factor(
    sample(c("Placebo", "Low Dose", "High Dose"), n, replace = TRUE,
           prob = c(0.4, 0.3, 0.3)),
    levels = c("Placebo", "Low Dose", "High Dose")
  ),

  # Potential stratification variables
  site = factor(sample(paste("Site", LETTERS[1:4]), n, replace = TRUE)),
  sex = factor(sample(c("Male", "Female"), n, replace = TRUE, prob = c(0.55, 0.45))),
  age_group = factor(
    sample(c("18-44", "45-64", "65+"), n, replace = TRUE, prob = c(0.3, 0.4, 0.3)),
    levels = c("18-44", "45-64", "65+")
  ),
  disease_severity = factor(
    sample(c("Mild", "Moderate", "Severe"), n, replace = TRUE, prob = c(0.4, 0.4, 0.2)),
    levels = c("Mild", "Moderate", "Severe")
  ),

  # Baseline characteristics
  age = round(rnorm(n, 58, 15)),
  bmi = round(rnorm(n, 26.5, 4.2), 1),
  systolic_bp = round(rnorm(n, 135, 18)),

  # Comorbidities
  diabetes = factor(sample(c("No", "Yes"), n, replace = TRUE, prob = c(0.75, 0.25))),
  hypertension = factor(sample(c("No", "Yes"), n, replace = TRUE, prob = c(0.65, 0.35))),

  # Lab values
  hemoglobin = round(rnorm(n, 13.2, 1.8), 1),
  creatinine = round(rnorm(n, 1.1, 0.3), 2)
)

# Add some realistic missing values
clinical_data$bmi[sample(1:n, 8)] <- NA
clinical_data$hemoglobin[sample(1:n, 5)] <- NA
clinical_data$creatinine[sample(1:n, 3)] <- NA

# Show dataset structure
str(clinical_data)
```

'data.frame': 300 obs. of 12 variables: $ treatment : Factor w/ 3 levels "Placebo","Low Dose",..: 2 2 1 2 3 3 2 1 3 2 ... $ site : Factor w/ 4 levels "Site A","Site B",..: 1 4 2 4 1 1 2 3 1 2 ... $ sex : Factor w/ 2 levels "Female","Male": 2 2 1 2 2 2 1 2 2 2 ... $ age_group : Factor w/ 3 levels "18-44","45-64",..: 2 2 3 2 3 1 2 2 3 2 ... $ disease_severity: Factor w/ 3 levels "Mild","Moderate",..: 1 1 1 2 3 3 1 1 2 1 ... $ age : num 34 29 53 61 24 62 64 77 60 61 ... $ bmi : num 27.5 28.3 26.5 29.1 28.3 28.8 28.6 33.5 22.1 32.8 ... $ systolic_bp : num 128 132 148 126 144 145 117 134 125 143 ... $ diabetes : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 1 1 1 1 1 ... $ hypertension : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 1 2 1 1 ... $ hemoglobin : num 13.9 12.4 13.4 12 14.6 10.1 13.5 13.2 15.1 12.8 ... $ creatinine : num 0.94 1.54 1.23 0.79 0.71 1.17 1.01 1.07 1.05 1.3 ...

```r
head(clinical_data, 10)
```

treatment site sex age_group disease_severity age bmi systolic_bp 1 Low Dose Site A Male 45-64 Mild 34 27.5 128 2 Low Dose Site D Male 45-64 Mild 29 28.3 132 3 Placebo Site B Female 65+ Mild 53 26.5 148 4 Low Dose Site D Male 45-64 Moderate 61 29.1 126 5 High Dose Site A Male 65+ Severe 24 28.3 144 6 High

Dose Site A Male 18-44 Severe 62 28.8 145 7 Low Dose Site B Female 45-64 Mild 64 28.6 117 8 Placebo Site C Male 45-64 Mild 77 33.5 134 9 High Dose Site A Male 65+ Moderate 60 22.1 125 10 Low Dose Site B Male 45-64 Mild 61 32.8 143 diabetes hypertension hemoglobin creatinine 1 No No 13.9 0.94 2 No No 12.4 1.54 3 No No 13.4 1.23 4 Yes Yes 12.0 0.79 5 No No 14.6 0.71 6 No No 10.1 1.17 7 No No 13.5 1.01 8 No Yes 13.2 1.07 9 No No 15.1 1.05 10 No No 12.8 1.30

3 Stratified Analysis Examples

3.1 Example 1: Stratified by Study Site

Understanding how baseline characteristics vary across different study sites is crucial for multi-center trials.

```
create_table(
  treatment ~ age + sex + bmi + diabetes + systolic_bp,
  data = clinical_data,
  strata = "site",
  theme = "nejm",
  pvalue = TRUE,
  totals = TRUE
)
```

% NEJM theme colors and formatting

variables	Placebo	Low Dose	High Dose	Total	p.value
Site: Site A					
age	57 ± 13	57.2 ± 19.7	61.6 ± 17.1	58.7 ± 16.5	0.822
sex					
Female	10 (35.7%)	8 (38.1%)	13 (46.4%)	31 (10%)	0.5583
Male	18 (64.3%)	13 (61.9%)	15 (53.6%)	46 (15%)	
bmi	26.9 ± 3.6	26.9 ± 4.4	26.4 ± 4.8	26.7 ± 4.2	0.7463
diabetes					
No	17 (60.7%)	13 (61.9%)	21 (75%)	51 (17%)	0.9867
Yes	11 (39.3%)	8 (38.1%)	7 (25%)	26 (9%)	
systolic_bp	138.9 ± 14.8	134 ± 17.3	134.1 ± 14.7	135.8 ± 15.5	0.724
Site: Site D					
age	62.5 ± 16.1	60.6 ± 14.9	56.3 ± 12.6	60.1 ± 14.7	0.822
sex					
Female	8 (36.4%)	5 (35.7%)	6 (37.5%)	19 (6%)	0.5583
Male	14 (63.6%)	9 (64.3%)	10 (62.5%)	33 (11%)	
bmi	26.9 ± 3.5	27.6 ± 5.4	25.9 ± 3	26.8 ± 4	0.7463
diabetes					
No	15 (68.2%)	8 (57.1%)	11 (68.8%)	34 (11%)	0.9867
Yes	7 (31.8%)	6 (42.9%)	5 (31.2%)	18 (6%)	
systolic_bp	131.9 ± 16.7	136 ± 18.9	135.8 ± 12.6	134.2 ± 16	0.724
Site: Site B					
age	54.7 ± 16.6	54.2 ± 12.7	56 ± 19.3	54.8 ± 15.8	0.822
sex					
Female	10 (24.4%)	19 (55.9%)	11 (52.4%)	40 (13%)	0.5583
Male	31 (75.6%)	15 (44.1%)	10 (47.6%)	56 (19%)	
bmi	27.1 ± 4.2	26.2 ± 4.2	27.2 ± 4.4	26.8 ± 4.2	0.7463
diabetes					
No	31 (75.6%)	29 (85.3%)	12 (57.1%)	72 (24%)	0.9867
Yes	10 (24.4%)	5 (14.7%)	9 (42.9%)	24 (8%)	
systolic_bp	135.6 ± 18.2	131.4 ± 18.7	134.2 ± 21.7	133.8 ± 19.1	0.724
Site: Site C					
age	58.2 ± 18.3	58.7 ± 16.4	61 ± 14.3	59.1 ± 16.6	0.822
sex					
Female	20 (62.5%)	10 (40%)	8 (44.4%)	38 (13%)	0.5583
Male	12 (37.5%)	15 (60%)	10 (55.6%)	37 (12%)	
bmi	25.2 ± 4	26.9 ± 3.9	24.9 ± 4.4	25.7 ± 4.1	0.7463
diabetes					
No	24 (75%)	15 (60%)	14 (77.8%)	53 (18%)	0.9867
Yes	8 (25%)	10 (40%)	4 (22.2%)	22 (7%)	
systolic_bp	137.3 ± 15.9	141.2 ± 17.7	129.2 ± 19.9	136.7 ± 17.8	0.724

Key Observations: - Each study site shows as a separate table section - Within-site treatment group comparisons - Helps identify site-specific recruitment patterns - Essential for assessing treatment balance across sites

3.2 Example 2: Stratified by Sex

Sex-stratified analysis is important for understanding treatment effects and baseline differences between male and female participants.

```
create_table(
  treatment ~ age + age_group + bmi + diabetes + hypertension + hemoglobin,
  data = clinical_data,
```

```
  strata = "sex",
  theme = "lancet",
  pvalue = TRUE,
  totals = TRUE
)
```

% Lancet theme formatting

variables	Placebo	Low Dose	High Dose	Total	p.value
Sex: Male					
age	58.2 (17.7)	54.4 (16)	59.6 (15.8)	57.4 (16.7)	0.822
age_group					
18-44	21 (28%)	13 (25%)	12 (26.7%)	46 (15%)	[Error]
45-64	33 (44%)	25 (48.1%)	18 (40%)	76 (25%)	
65+	21 (28%)	14 (26.9%)	15 (33.3%)	50 (17%)	
bmi	26.9 (4.1)	26.9 (4.3)	26.3 (3.3)	26.7 (4)	0.7463
diabetes					
No	50 (66.7%)	36 (69.2%)	32 (71.1%)	118 (39%)	0.9867
Yes	25 (33.3%)	16 (30.8%)	13 (28.9%)	54 (18%)	
hypertension					
No	45 (60%)	31 (59.6%)	33 (73.3%)	109 (36%)	0.3364
Yes	30 (40%)	21 (40.4%)	12 (26.7%)	63 (21%)	
hemoglobin	13.3 (1.7)	13 (1.8)	13.7 (1.7)	13.3 (1.7)	0.3287
Sex: Female					
age	56.5 (13.8)	60.2 (14.9)	58.4 (17)	58.3 (15.2)	0.822
age_group					
18-44	17 (35.4%)	12 (28.6%)	13 (34.2%)	42 (14%)	[Error]
45-64	16 (33.3%)	17 (40.5%)	18 (47.4%)	51 (17%)	
65+	15 (31.2%)	13 (31%)	7 (18.4%)	35 (12%)	
bmi	26.1 (3.6)	26.5 (4.4)	26.1 (5.3)	26.2 (4.4)	0.7463
diabetes					
No	37 (77.1%)	29 (69%)	26 (68.4%)	92 (31%)	0.9867
Yes	11 (22.9%)	13 (31%)	12 (31.6%)	36 (12%)	
hypertension					
No	27 (56.2%)	28 (66.7%)	24 (63.2%)	79 (26%)	0.3364
Yes	21 (43.8%)	14 (33.3%)	14 (36.8%)	49 (16%)	
hemoglobin	13.3 (2)	13.1 (1.9)	12.4 (1.7)	13 (1.9)	0.3287

Key Observations: - Separate baseline characteristics for males and females - Age and BMI distributions may differ by sex - Hemoglobin levels typically differ between sexes - Treatment allocation balance within each sex

3.3 Example 3: Stratified by Disease Severity

Disease severity stratification helps understand how patient characteristics and treatment allocation vary by baseline disease status.

```
create_table(
  treatment ~ age + sex + bmi + systolic_bp + diabetes + hypertension + creatinine,
  data = clinical_data,
  strata = "disease_severity",
  theme = "jama",
  pvalue = TRUE,
  totals = TRUE
)
```

% JAMA theme formatting

variables	Placebo	Low Dose	High Dose	Total	p.value
Disease_severity: Mild					
age	56.9 (15.8)	57.8 (15.5)	58.5 (17.2)	57.6 (15.9)	0.822
sex					
Female	23 (43.4%)	22 (50%)	9 (34.6%)	54 (18%)	0.5583
Male	30 (56.6%)	22 (50%)	17 (65.4%)	69 (23%)	
bmi	26.7 (3.7)	26.8 (4.4)	25.2 (4.7)	26.4 (4.2)	0.7463
systolic_bp	137.2 (16.4)	135.9 (19.3)	135.7 (19.5)	136.4 (18)	0.724
diabetes					
No	43 (81.1%)	32 (72.7%)	15 (57.7%)	90 (30%)	0.9867
Yes	10 (18.9%)	12 (27.3%)	11 (42.3%)	33 (11%)	
hypertension					
No	30 (56.6%)	26 (59.1%)	17 (65.4%)	73 (24%)	0.3364
Yes	23 (43.4%)	18 (40.9%)	9 (34.6%)	50 (17%)	
creatinine	1.1 (0.3)	1 (0.4)	1.1 (0.3)	1.1 (0.3)	0.7768
Disease_severity: Moderate					
age	56.4 (16)	53.8 (14.8)	59.2 (16.4)	56.6 (15.8)	0.822
sex					
Female	20 (39.2%)	12 (35.3%)	21 (52.5%)	53 (18%)	0.5583
Male	31 (60.8%)	22 (64.7%)	19 (47.5%)	72 (24%)	
bmi	26.7 (4.3)	26.5 (4.6)	27 (4.1)	26.7 (4.3)	0.7463
systolic_bp	135.9 (17.8)	134.2 (16)	133.2 (16.5)	134.5 (16.8)	0.724
diabetes					
No	31 (60.8%)	24 (70.6%)	30 (75%)	85 (28%)	0.9867
Yes	20 (39.2%)	10 (29.4%)	10 (25%)	40 (13%)	
hypertension					
No	35 (68.6%)	24 (70.6%)	32 (80%)	91 (30%)	0.3364
Yes	16 (31.4%)	10 (29.4%)	8 (20%)	34 (11%)	
creatinine	1.1 (0.3)	1.2 (0.3)	1.1 (0.3)	1.1 (0.3)	0.7768
Disease_severity: Severe					
age	62.1 (18.5)	61.9 (17.5)	59.5 (15.4)	61.2 (16.9)	0.822
sex					
Female	5 (26.3%)	8 (50%)	8 (47.1%)	21 (7%)	0.5583
Male	14 (73.7%)	8 (50%)	9 (52.9%)	31 (10%)	
bmi	26 (3.7)	27.3 (3.8)	25.7 (3.9)	26.3 (3.8)	0.7463
systolic_bp	133.8 (14)	136.1 (20.8)	130.5 (16.6)	133.4 (17)	0.724
diabetes					
No	13 (68.4%)	9 (56.2%)	13 (76.5%)	35 (12%)	0.9867
Yes	6 (31.6%)	7 (43.8%)	4 (23.5%)	17 (6%)	
hypertension					
No	7 (36.8%)	9 (56.2%)	8 (47.1%)	24 (8%)	0.3364
Yes	12 (63.2%)	7 (43.8%)	9 (52.9%)	28 (9%)	
creatinine	1.3 (0.3)	1.2 (0.3)	1.2 (0.2)	1.2 (0.3)	0.7768

Key Observations: - Baseline characteristics across mild, moderate, and severe disease - Treatment allocation may vary by severity - Comorbidity prevalence often increases with disease severity - Important for stratified randomization assessment

3.4 Example 4: Stratified by Age Group

Age-group stratified analysis reveals how baseline characteristics and treatment allocation vary across different age ranges.

```
create_table(
  treatment ~ sex + bmi + systolic_bp + diabetes + hypertension + hemoglobin + creatinine,
  data = clinical_data,
  strata = "age_group",
  theme = "nejm",
  pvalue = TRUE,
  totals = TRUE
)
```

% NEJM theme colors and formatting

variables	Placebo	Low Dose	High Dose	Total	p.value
Age_group: 45-64					
sex					
Female	16 (32.7%)	17 (40.5%)	18 (50%)	51 (17%)	0.5583
Male	33 (67.3%)	25 (59.5%)	18 (50%)	76 (25%)	
bmi	26.9 ± 3.9	25.7 ± 4.5	25.5 ± 4.5	26.1 ± 4.3	0.7463
systolic_bp	135.7 ± 15.7	134.8 ± 19.9	131.8 ± 18.5	134.3 ± 17.9	0.724
diabetes					
No	34 (69.4%)	31 (73.8%)	28 (77.8%)	93 (31%)	0.9867
Yes	15 (30.6%)	11 (26.2%)	8 (22.2%)	34 (11%)	
hypertension					
No	30 (61.2%)	27 (64.3%)	20 (55.6%)	77 (26%)	0.3364
Yes	19 (38.8%)	15 (35.7%)	16 (44.4%)	50 (17%)	
hemoglobin	13.6 ± 1.7	13.1 ± 1.5	13.2 ± 1.8	13.3 ± 1.7	0.3287
creatinine	1.1 ± 0.3	1.2 ± 0.4	1 ± 0.3	1.1 ± 0.3	0.7768
Age_group: 65+					
sex					
Female	15 (41.7%)	13 (48.1%)	7 (31.8%)	35 (12%)	0.5583
Male	21 (58.3%)	14 (51.9%)	15 (68.2%)	50 (17%)	
bmi	26.1 ± 4.1	28 ± 4.9	26.4 ± 3.6	26.8 ± 4.3	0.7463
systolic_bp	134.2 ± 15.1	133.6 ± 17.3	132.5 ± 17.9	133.6 ± 16.4	0.724
diabetes					
No	22 (61.1%)	19 (70.4%)	15 (68.2%)	56 (19%)	0.9867
Yes	14 (38.9%)	8 (29.6%)	7 (31.8%)	29 (10%)	
hypertension					
No	17 (47.2%)	19 (70.4%)	17 (77.3%)	53 (18%)	0.3364
Yes	19 (52.8%)	8 (29.6%)	5 (22.7%)	32 (11%)	
hemoglobin	13.1 ± 1.8	13 ± 2.1	13.3 ± 1.7	13.1 ± 1.9	0.3287
creatinine	1.2 ± 0.3	1 ± 0.3	1.1 ± 0.3	1.1 ± 0.3	0.7768
Age_group: 18-44					
sex					
Female	17 (44.7%)	12 (48%)	13 (52%)	42 (14%)	0.5583
Male	21 (55.3%)	13 (52%)	12 (48%)	46 (15%)	
bmi	26.6 ± 3.8	27.1 ± 3	27 ± 4.6	26.8 ± 3.8	0.7463
systolic_bp	138.5 ± 19	138 ± 16.8	136.5 ± 15.4	137.8 ± 17.2	0.724
diabetes					
No	31 (81.6%)	15 (60%)	15 (60%)	61 (20%)	0.9867
Yes	7 (18.4%)	10 (40%)	10 (40%)	27 (9%)	
hypertension					
No	25 (65.8%)	13 (52%)	20 (80%)	58 (19%)	0.3364
Yes	13 (34.2%)	12 (48%)	5 (20%)	30 (10%)	
hemoglobin	13.1 ± 1.9	13 ± 2.1	12.7 ± 1.9	13 ± 1.9	0.3287
creatinine	1.1 ± 0.3	1.1 ± 0.2	1.2 ± 0.3	1.1 ± 0.3	0.7768

Key Observations: - Younger vs middle-aged vs older participants - Comorbidity prevalence increases with age - Lab values may vary by age group - Treatment allocation balance across age groups

4 Advanced Stratified Analysis

4.1 Multiple Variables with Missing Data

Let's examine how stratified analysis handles missing data and multiple variable types.

```
create_table(
  treatment ~ age + bmi + hemoglobin + creatinine + diabetes + hypertension,
  data = clinical_data,
  strata = "sex",
  theme = "lancet",
  missing = TRUE,  # Show missing value patterns
  pvalue = TRUE,
  totals = TRUE
)
```

% Lancet theme formatting

variables	Placebo	Low Dose	High Dose	Total	p.value
Sex: Male					
age	58.2 (17.7)	54.4 (16)	59.6 (15.8)	57.4 (16.7)	0.822
bmi	26.9 (4.1)	26.9 (4.3)	26.3 (3.3)	26.7 (4)	0.7463
hemoglobin	13.3 (1.7)	13 (1.8)	13.7 (1.7)	13.3 (1.7)	0.3287
creatinine	1.1 (0.3)	1.2 (0.3)	1.1 (0.3)	1.1 (0.3)	0.7768
diabetes					
No	50 (66.7%)	36 (69.2%)	32 (71.1%)	118 (39%)	0.9867
Yes	25 (33.3%)	16 (30.8%)	13 (28.9%)	54 (18%)	
hypertension					
No	45 (60%)	31 (59.6%)	33 (73.3%)	109 (36%)	0.3364
Yes	30 (40%)	21 (40.4%)	12 (26.7%)	63 (21%)	
Sex: Female					
age	56.5 (13.8)	60.2 (14.9)	58.4 (17)	58.3 (15.2)	0.822
bmi	26.1 (3.6)	26.5 (4.4)	26.1 (5.3)	26.2 (4.4)	0.7463
hemoglobin	13.3 (2)	13.1 (1.9)	12.4 (1.7)	13 (1.9)	0.3287
creatinine	1.2 (0.3)	1 (0.3)	1.1 (0.3)	1.1 (0.3)	0.7768
diabetes					
No	37 (77.1%)	29 (69%)	26 (68.4%)	92 (31%)	0.9867
Yes	11 (22.9%)	13 (31%)	12 (31.6%)	36 (12%)	
hypertension					
No	27 (56.2%)	28 (66.7%)	24 (63.2%)	79 (26%)	0.3364
Yes	21 (43.8%)	14 (33.3%)	14 (36.8%)	49 (16%)	

4.2 Site and Sex Combined Analysis

For comprehensive analysis, we might want to examine the interaction of multiple stratification factors.

```
# Create a combined stratification variable for demonstration
clinical_data$site_sex <- interaction(clinical_data$site, clinical_data$sex, sep = " - ")

# Show the distribution
table(clinical_data$site_sex, clinical_data$treatment)
```

```
               Placebo Low Dose High Dose
```

Site A - Female 10 8 13 Site B - Female 10 19 11 Site C - Female 20 10 8 Site D - Female 8 5 6 Site A - Male

18 13 15 Site B - Male 31 15 10 Site C - Male 12 15 10 Site D - Male 14 9 10

```
create_table(
  treatment ~ age + bmi + diabetes + systolic_bp,
  data = clinical_data,
  strata = "site_sex",
  theme = "jama",
  pvalue = TRUE
)
```

% JAMA theme formatting

variables	Placebo	Low Dose	High Dose	p.value
Site_sex: Site A - Male				
age	58 (12.5)	51.9 (16.4)	58.9 (19.2)	0.822
bmi	25.9 (3.5)	27.2 (4.7)	26.9 (3.6)	0.7463
diabetes				
No	8 (44.4%)	9 (69.2%)	12 (80%)	0.9867
Yes	10 (55.6%)	4 (30.8%)	3 (20%)	
systolic_bp	137.5 (16.3)	133.2 (18.5)	129.1 (14.4)	0.724
Site_sex: Site D - Male				
age	64.9 (17.6)	59.7 (16.2)	59.5 (13.6)	0.822
bmi	27.7 (3.9)	26.3 (4.8)	26.3 (2.5)	0.7463
diabetes				
No	10 (71.4%)	6 (66.7%)	8 (80%)	0.9867
Yes	4 (28.6%)	3 (33.3%)	2 (20%)	
systolic_bp	133.5 (15.5)	134.3 (18.6)	137.7 (15.3)	0.724
Site_sex: Site B - Female				
age	52.1 (13.8)	54.3 (11.1)	50.5 (19.4)	0.822
bmi	25.8 (3.8)	25.8 (3.6)	28.5 (4.3)	0.7463
diabetes				
No	8 (80%)	17 (89.5%)	7 (63.6%)	0.9867
Yes	2 (20%)	2 (10.5%)	4 (36.4%)	
systolic_bp	135.5 (12.5)	123.7 (18.3)	129.5 (22.6)	0.724
Site_sex: Site C - Male				
age	57.4 (24.2)	53.9 (17.4)	58.1 (11.2)	0.822
bmi	25.4 (4.4)	27.3 (3.4)	25.6 (2.8)	0.7463
diabetes				
No	9 (75%)	9 (60%)	7 (70%)	0.9867
Yes	3 (25%)	6 (40%)	3 (30%)	
systolic_bp	131.2 (12.3)	140.3 (20.5)	131.3 (15.5)	0.724
Site_sex: Site B - Male				
age	55.5 (17.5)	54 (14.9)	62 (18.2)	0.822
bmi	27.6 (4.3)	26.8 (4.9)	25.9 (4.3)	0.7463
diabetes				
No	23 (74.2%)	12 (80%)	5 (50%)	0.9867
Yes	8 (25.8%)	3 (20%)	5 (50%)	
systolic_bp	135.7 (19.9)	141.2 (14.7)	139.5 (20.4)	0.724
Site_sex: Site C - Female				
age	58.7 (14.4)	65.9 (12.1)	64.6 (17.5)	0.822
bmi	25.1 (3.9)	26.3 (4.7)	24.2 (5.9)	0.7463
diabetes				
No	15 (75%)	6 (60%)	7 (87.5%)	0.9867
Yes	5 (25%)	4 (40%)	1 (12.5%)	
systolic_bp	140.9 (16.9)	142.5 (13.2)	126.5 (25.2)	0.724
Site_sex: Site D - Female				
age	58.2 (12.9)	62.4 (13.6)	51 (9.5)	0.822
bmi	25.5 (2.3)	29.9 (6.2)	25.2 (3.8)	0.7463
diabetes				
No	5 (62.5%)	2 (40%)	3 (50%)	0.9867
Yes	3 (37.5%)	3 (60%)	3 (50%)	
systolic_bp	129 (19.4)	139 (21.2)	132.5 (6.2)	0.724
Site_sex: Site A - Female				
age	55.1 (14.3)	65.9 (22.7)	64.7 (14.5)	0.822
bmi	28.7 (3.2)	26.3 (4.3)	25.8 (5.9)	0.7463
diabetes				
No	9 (90%)	4 (50%)	9 (69.2%)	0.9867
Yes	1 (10%)	4 (50%)	4 (30.8%)	
systolic_bp	141.4 (12)	135.4 (16.2)	139.8 (13.4)	0.724

5 Comparative Analysis

5.1 Before and After Stratification

Let's compare an overall analysis with a stratified analysis to see how stratification reveals important patterns.

5.1.1 Overall Analysis (Non-stratified)

```
create_table(
  treatment ~ age + sex + bmi + diabetes + hypertension + systolic_bp,
  data = clinical_data,
  theme = "console",
  pvalue = TRUE,
  totals = TRUE
)
```

variables	Placebo	Low Dose	High Dose	Total	p.value
age	57.5 (16.3)	57 (15.7)	59 (16.3)	57.8 (16.1)	0.822
sex					
Female	48 (39%)	42 (45%)	38 (46%)	128 (43%)	0.5583
Male	75 (61%)	52 (55%)	45 (54%)	172 (57%)	
bmi	26.6 (3.9)	26.8 (4.3)	26.2 (4.3)	26.5 (4.2)	0.7463
diabetes					
No	87 (71%)	65 (69%)	58 (70%)	210 (70%)	0.9867
Yes	36 (29%)	29 (31%)	25 (30%)	90 (30%)	
hypertension					
No	72 (59%)	59 (63%)	57 (69%)	188 (63%)	0.3364
Yes	51 (41%)	35 (37%)	26 (31%)	112 (37%)	
systolic_bp	136.1 (16.6)	135.3 (18.3)	133.4 (17.4)	135.1 (17.3)	0.724

5.1.2 Stratified by Disease Severity

```
create_table(
  treatment ~ age + sex + bmi + diabetes + hypertension + systolic_bp,
  data = clinical_data,
  strata = "disease_severity",
  theme = "console",
  pvalue = TRUE,
  totals = TRUE
)
```

variables	Placebo	Low Dose	High Dose	Total	p.value
Disease_severity: Mild					
age	56.9 (15.8)	57.8 (15.5)	58.5 (17.2)	57.6 (15.9)	0.822
sex					
Female	23 (43.4%)	22 (50%)	9 (34.6%)	54 (18%)	0.5583
Male	30 (56.6%)	22 (50%)	17 (65.4%)	69 (23%)	
bmi	26.7 (3.7)	26.8 (4.4)	25.2 (4.7)	26.4 (4.2)	0.7463
diabetes					
No	43 (81.1%)	32 (72.7%)	15 (57.7%)	90 (30%)	0.9867
Yes	10 (18.9%)	12 (27.3%)	11 (42.3%)	33 (11%)	
hypertension					
No	30 (56.6%)	26 (59.1%)	17 (65.4%)	73 (24%)	0.3364
Yes	23 (43.4%)	18 (40.9%)	9 (34.6%)	50 (17%)	
systolic_bp	137.2 (16.4)	135.9 (19.3)	135.7 (19.5)	136.4 (18)	0.724
Disease_severity: Moderate					
age	56.4 (16)	53.8 (14.8)	59.2 (16.4)	56.6 (15.8)	0.822
sex					
Female	20 (39.2%)	12 (35.3%)	21 (52.5%)	53 (18%)	0.5583
Male	31 (60.8%)	22 (64.7%)	19 (47.5%)	72 (24%)	
bmi	26.7 (4.3)	26.5 (4.6)	27 (4.1)	26.7 (4.3)	0.7463
diabetes					
No	31 (60.8%)	24 (70.6%)	30 (75%)	85 (28%)	0.9867
Yes	20 (39.2%)	10 (29.4%)	10 (25%)	40 (13%)	
hypertension					
No	35 (68.6%)	24 (70.6%)	32 (80%)	91 (30%)	0.3364
Yes	16 (31.4%)	10 (29.4%)	8 (20%)	34 (11%)	
systolic_bp	135.9 (17.8)	134.2 (16)	133.2 (16.5)	134.5 (16.8)	0.724
Disease_severity: Severe					
age	62.1 (18.5)	61.9 (17.5)	59.5 (15.4)	61.2 (16.9)	0.822
sex					
Female	5 (26.3%)	8 (50%)	8 (47.1%)	21 (7%)	0.5583
Male	14 (73.7%)	8 (50%)	9 (52.9%)	31 (10%)	
bmi	26 (3.7)	27.3 (3.8)	25.7 (3.9)	26.3 (3.8)	0.7463
diabetes					
No	13 (68.4%)	9 (56.2%)	13 (76.5%)	35 (12%)	0.9867
Yes	6 (31.6%)	7 (43.8%)	4 (23.5%)	17 (6%)	
hypertension					
No	7 (36.8%)	9 (56.2%)	8 (47.1%)	24 (8%)	0.3364
Yes	12 (63.2%)	7 (43.8%)	9 (52.9%)	28 (9%)	
systolic_bp	133.8 (14)	136.1 (20.8)	130.5 (16.6)	133.4 (17)	0.724

Key Differences: - Overall analysis may mask important subgroup differences - Stratified analysis reveals severity-specific patterns - P-values may differ when accounting for stratification - Treatment balance assessment within severity levels

6 Summary and Best Practices

6.1 When to Use Stratified Analysis

1. **Multi-center Studies**: Always stratify by study site
2. **Sex Differences**: Important for most clinical studies
3. **Age Groups**: Especially relevant for studies spanning wide age ranges
4. **Disease Severity**: Critical for understanding baseline risk
5. **Geographic Regions**: For studies spanning different populations

6.2 Interpretation Guidelines

1. **Sample Sizes**: Check adequate sample sizes within strata
2. **Missing Data**: Consider missing data patterns within strata
3. **P-values**: Interpret within-strata comparisons carefully
4. **Clinical Relevance**: Focus on clinically meaningful differences
5. **Multiple Comparisons**: Consider adjustment for multiple testing

6.3 Available Stratification Variables in This Dataset

Table 1: Available Stratification Variables

Variable	Description	Use Case
site	Study site (A, B, C, D)	Multi-center trial balance
sex	Participant sex (Male, Female)	Sex-specific effects
age_group	Age groups (18-44, 45-64, 65+)	Age-related patterns
disease_severity	Disease severity (Mild, Moderate, Severe)	Baseline risk stratification
diabetes	Diabetes status (No, Yes)	Comorbidity analysis
hypertension	Hypertension status (No, Yes)	Cardiovascular risk factors

The stratified analysis capabilities of `zztable1_nextgen` provide powerful tools for understanding complex clinical trial data. By examining baseline characteristics within meaningful subgroups, researchers can better assess treatment allocation, identify potential confounders, and plan appropriate statistical analyses.