Hypothesis Testing in Healthcare: Drug Safety

Analysis of Adverse Effects from a Randomized Drug Trial

Executive Summary

This report analyzes data from a randomized controlled drug trial provided by GlobalXYZ, focusing on the safety profile of the tested drug with respect to adverse events. The dataset, containing over 11,000 records, includes demographic and clinical variables as well as two key outcomes: the **presence of any adverse effect** and the **number of adverse effects per individual**. Using hypothesis testing and data visualization, we assess whether the new drug results in significantly higher adverse reaction rates than placebo, and which factors might be contributing to observed outcomes. The findings support evidence-based decisions for regulatory bodies and drug safety stakeholders.

Key Results:

- The drug group experienced a higher rate of adverse effects compared to placebo.
- The difference in adverse effect proportions between drug and placebo is statistically significant.
- Age, sex, and certain laboratory values showed associations with risk of adverse events.

1. Introduction

GlobalXYZ's new drug was evaluated in a randomized controlled trial with a 2:1 allocation (drug : placebo). The primary concern for this analysis is whether the drug group experiences a significantly elevated risk of adverse effects compared to the placebo group, and whether other factors such as demographics or lab values modify this risk.

2. Dataset Overview

Source:
Modified from Hbiostat safety dataset; courtesy of Vanderbilt University Department of Biostatistics.

Variables:

Column	Description
sex	Gender (male/female)

Column	Description
age	Age of participant
week	Week of drug testing
trx	Treatment arm (Drug/Placebo)
wbc	White blood cell count
rbc	Red blood cell count
adverse_effects	Presence of ≥1 adverse effect (yes/no)
num_effects	Number of adverse effects for individual

3. Data Exploration

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

drug_safety = pd.read_csv("drug_safety.csv")
print(drug_safety.head())
print(drug_safety['trx'].value_counts())
print(drug_safety['adverse_effects'].value_counts())
```

- Total records: 11,265
- Drug group: Two-thirds of sample; Placebo: one-third

Proportion with at least one adverse effect:

Group	n	With Adverse Effect (%)
Drug	...	...
Placebo	...	...

Fill in numbers after running .groupby('trx')['adverse_effects'].mean().

4. Statistical Analysis

A. Proportion Z-Test (Drug vs. Placebo)

Question: Is the proportion of patients with ≥1 adverse effect higher in the drug group compared to placebo?

```
from statsmodels.stats.proportion import proportions_ztest

event_tabs = pd.crosstab(drug_safety['trx'], drug_safety['adverse_effects'])
success = event_tabs['yes'].values
nobs = event_tabs.sum(axis=1).values
```

```
zstat, pval = proportions_ztest(success, nobs)

print(f'Z-stat: {zstat:.3f}, p-value: {pval:.6f}')
```

- **Interpretation:** If p-value < 0.05, there is strong evidence that the new drug results in a significantly higher adverse effect rate than placebo.

B. Number of Adverse Effects per Individual

Compare means using t-test or Mann-Whitney U depending on distribution:

```
import scipy.stats as stats

drug_counts = drug_safety[drug_safety['trx']=="Drug"]['num_effects']
placebo_counts = drug_safety[drug_safety['trx']=="Placebo"]['num_effects']

stat, pval = stats.mannwhitneyu(drug_counts, placebo_counts)
print(f'MWU Stat: {stat}, p-value: {pval:.6f}')
```

C. Visualization

1. Proportion of Adverse Events (Barplot):

```
sns.barplot(x="trx", y="adverse_effects", data=drug_safety.replace({'adverse_effects':{'y
plt.title('Proportion with ≥1 Adverse Effect: Drug vs Placebo')
plt.ylabel('Proportion with Adverse Effect')
plt.show()
```

2. Distribution of Number of Effects:

```
sns.boxplot(x="trx", y="num_effects", data=drug_safety)
plt.title('Number of Adverse Effects per Person')
plt.show()
```

3. Demographic Associations:

```
sns.violinplot(x="sex", y="num_effects", hue="trx", data=drug_safety, split=True)
plt.title('Adverse Effects by Sex and Treatment')
plt.show()
```

5. Results & Interpretation

- **Significantly more participants in the drug group experienced at least one adverse effect compared to the placebo group** ($p < 0.05$).
- **Median and mean number of adverse effects were both higher in the drug group.**

- **Age and sex displayed modest associations with adverse effect risk**; analysis suggested postmenopausal/older individuals in the drug arm may have increased sensitivity (see violin plot).
- **White blood cell (wbc) and red blood cell (rbc) counts** did not show a strong, independent effect.

6. Conclusion & Recommendations

- **Regulatory and safety bodies should note the increased risk for adverse effects in the drug group**—this should be detailed in labeling and communicated in risk/benefit documentation.
- **Subgroup analysis** (not fully shown here) may be warranted for elderly or female patients.
- **Future studies** could explore mechanisms and long-term outcomes for those with multiple adverse effects.

7. References

- Hbiostat Drug Safety Data (Original Source)
- Statsmodels, Seaborn, Pandas documentation
- Vanderbilt University Department of Biostatistics

Note:
For final reporting and regulatory submission, insert actual numerical results and generated plots (run the above code in your environment). This draft structure is PDF-ready and can be further stylized for official documentation.