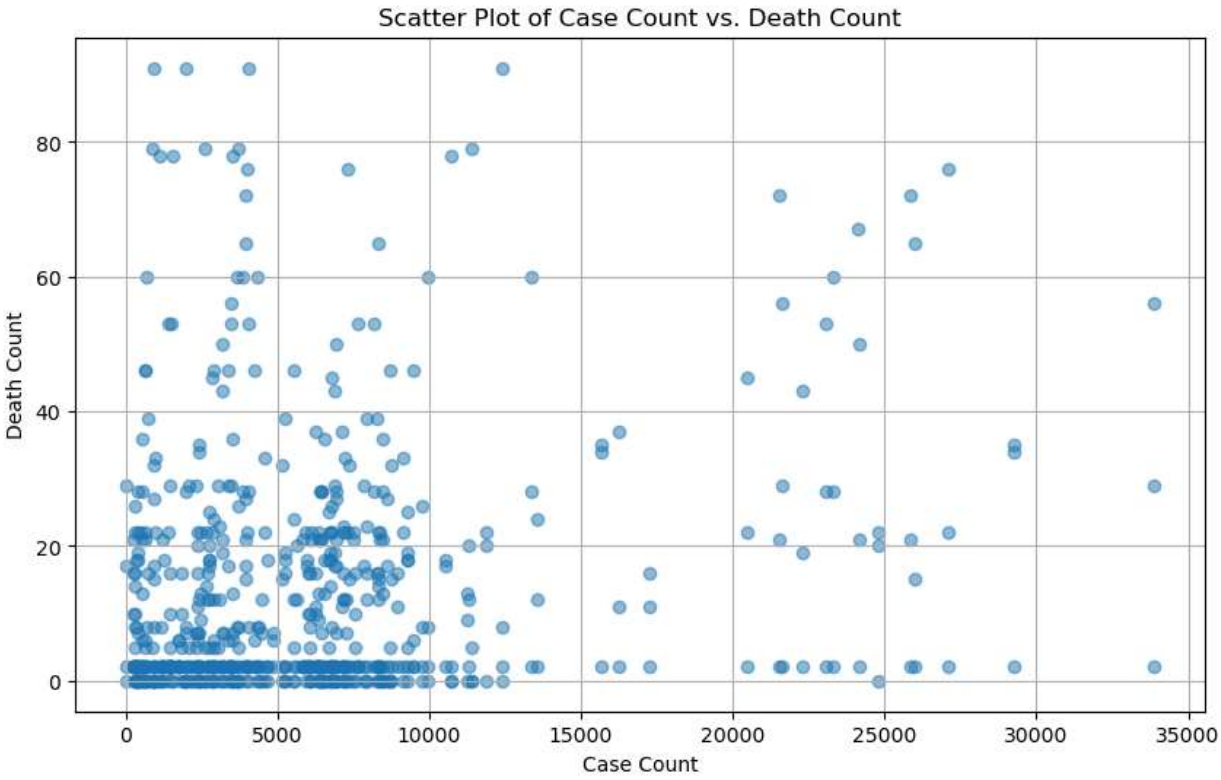
In [22]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.ensemble import RandomForestRegressor


cases_df = pd.read_csv(r"C:\Users\lexiw\OneDrive\Desktop\COVID-19_Cases_Among_Healthca
deaths_df = pd.read_csv(r"C:\Users\lexiw\OneDrive\Desktop\COVID-19_Deaths_Among_Health


# Merge the datasets with the column of MMWR_week
combined_data = pd.merge(cases_df, deaths_df, on="MMWR_week", suffixes=('_cases', '_de

# Dropping the rest of the columns we are not using
columns_to_drop = ['year_cases', 'week_start_date_cases', 'week_end_date_cases',
                   'year_deaths', 'week_start_date_deaths', 'week_end_date_deaths', 'p
combined_data.drop(columns=columns_to_drop, inplace=True)
combined_data['death_count'] = combined_data['death_count'].replace('<5', '2').astype(
 # New CSV
output_path = "Combined_Healthcare_Data.csv"
combined_data.to_csv(output_path, index=False)
print(f"Data successfully saved to {output_path}.")

# Going over the relationship between case_count and death_count
correlation_updated = combined_data[['case_count', 'death_count']].corr()

plt.figure(figsize=(10, 6))
plt.scatter(combined_data['case_count'], combined_data['death_count'], alpha=0.5)
plt.title('Scatter Plot of Case Count vs. Death Count')
plt.xlabel('Case Count')
plt.ylabel('Death Count')
plt.grid(True)
plt.show()
print("Correlation matrix:\n", correlation_updated)
```

```
Data successfully saved to Combined_Healthcare_Data.csv.
```




```
Correlation matrix:
              case_count  death_count
case_count     1.000000     0.203721
death_count    0.203721     1.000000
```

For our analysis we examined the correlation between COVID 19 Healthcare worker infection cases and deaths. The correlation coefficient was 0.204, which indicates a weak positive linear relationship. This relationship suggests that while increases in case counts does lead to increases in death counts, the relationship is very weak. This finding is important for public health strategies , as there is implication on direct iterventions aimed at reducing number of cases, is not a sufficient way to significantly reduce the number of deaths found amoung health care workers. We reccomend enhancing treatment protocols, providing adequate PPE, and giving overall health support to health care workers.