Xiao Lai

# Breast Cancer Screening Report

## Summary

This analysis includes patient data from 11 clinical teams, comprising 516 observations across eight variables. I excluded one dataset DN01 due to missing 'Employment Status' (ES) data. The consolidated dataset included variables such as breast cancer screening status, age, weight, height, alcohol intake, employment status, number of dependents (Deps), and surveyor identifier. Additionally, 45 missing values were identified in the height and weight variables.

Using the 'metafor' package, I found significant heterogeneity among the surveyors, suggesting variations in the effect sizes across different surveyors. My findings reveal a statistically significant association between patients' employment status and the number of dependents deciding to undergo breast cancer screening. This insight indicates that socioeconomic factors potentially influence patients' health-seeking behaviours, underscoring the importance of understanding patients' backgrounds to enhance breast cancer screening strategies.

**Table 1**. Distribution of missing data across different surveyors. Total means total number of patients each surveyor have surveyed. Percentage means percent of missing values from each surveyor.

| Surveyor | Height | Alcohol | Total | Percentage |
|---|---|---|---|---|
| DN02 | 1 | 1 | 33 | 6.06 |
| DN03 | 0 | 3 | 47 | 6.38 |
| DN04 | 0 | 1 | 18 | 5.56 |
| DN05 | 3 | 8 | 82 | 13.41 |
| DN06 | 0 | 2 | 28 | 7.14 |
| DN07 | 2 | 3 | 35 | 14.29 |
| DN08 | 1 | 5 | 32 | 18.75 |
| DN09 | 0 | 5 | 64 | 7.81 |
| DN010 | 0 | 1 | 18 | 5.56 |
| DN011 | 1 | 4 | 57 | 8.77 |
| DN012 | 0 | 4 | 102 | 3.92 |

## Methodology

I chose the 'metafor' package for this data analysis due to its suite for one study conducted by multiple clinical teams. As 'BUGSnet' specializes in network meta-analysis utilizing Bayesian techniques, it would be better suited if our data were drawn from independent research.

The analysis was conducted in the following steps:

**Data Merging**: Eleven individual datasets collected by different clinical teams were merged. The dataset collected by DN01 was excluded due to a missing column of employment status. To maintain the source information, I introduced an extra column called "Surveyor", which identifies the clinical team that collected each set of patient data.

**Missing Data Exploration**: After merging, I conducted a thorough exploration of the data to identify any missing values. The findings are reported in Table 1, including the distribution of missing data across surveyors.

**Table 2**. Heterogeneity test results with random-effects model (RE) and fixed-effect

| | RE | FE |
|---|---|---|
| tau^2 (estimated amount of total heterogeneity) | 0.6706 | |
| tau (square root of estimated tau^2 value) | 0.8189 | |
| I^2 (total heterogeneity / total variability) | 78.53% | 71.49% |
| H^2 (total variability / sampling variability) | 4.66 | 3.51 |
| p-value | 0.0001 | 0.0001 |

Xiao Lai

**Heterogeneity Test**: I analyzed the heterogeneity across different surveyors concerning the breast cancer screening decision, using both random-effects and fixed-effects models.

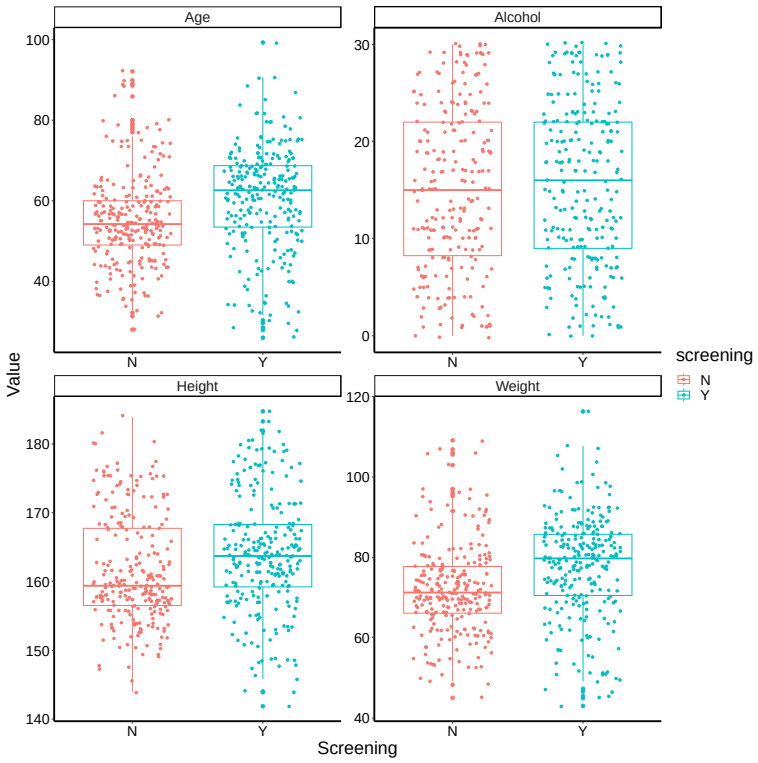
**Association Analysis**: Due to the high heterogeneity of the data, I have used a random-effects model to assess the association between patient characteristics and their decision to undergo breast cancer screening.

**Data Visualization**: Finally, to aid in the interpretation of the results, I used the ggplot2 to create informative and aesthetic visualizations of our data.

## Results and Discussion

**Summary of missing data**: This project involved the analysis of twelve distinct datasets. However, dataset DN01 was found to be missing the 'Employment Status' column, possibly due to data collection oversights or data entry errors. In the remaining eleven datasets, a total of 45 missing values were noted in patients' 'Height' and 'Weight' data. Such missing values may be due to non-responses from patients, data entry mishaps, or data losses during cleaning or processing. Table 1 provides a detailed summary of these missing values as per the clinical team. It was observed that teams 8, 7, and 5 exhibited more than a 10% incidence of missing values, marking a higher rate compared to the other teams.

**Figure 1**. Distribution of patients' age, height, weight, and alcohol consumption patterns in relation to their decisions about breast cancer screening.




**Heterogeneity test**: The heterogeneity test reveals high variability across patients surveyed by different clinical teams, as confirmed by both random-effects and fixed-effects models. With I^2 scores exceeding 70% and a p-value less than 0.0001 (Table 2), the test underscores significant heterogeneity. This suggests the patient populations seen by different teams may not be directly comparable due to their distinct characteristics. Consequently, subsequent analysis will employ only the random-effects model to account for this high variability.

**Table 3**. p-values for each patient characteristics

| | p-value |
|---|---|
| intercept | 0.9728 |
| Age | 0.4624 |
| Weight | 0.2961 |
| Height | 0.9565 |
| Alcohol | 0.3039 |
| ES | <.0001*** |
| Deps | <.0001*** |

**Association test**: The association test reveals a strong correlation between employment status, number of dependents, and breast cancer screening decisions. Table 3 displays p-values for each variable. Age, Weight, Height, and Alcohol intake show p-values above 0.05, indicating no significant association, whereas Employment Status (ES) and Dependents (Deps) present p-values less than 0.0001, signifying significant association at an 99.99% confidence interval. Additionally, visualizations were created, illustrating each patient's characteristics versus their screening decisions (Figure 1 and 2), particularly focusing on employment status, number of dependents, and screening decisions across different clinical teams (Figure 3). As we can see, there were less participants who do part-time work or with 3 dependents. When look at clinic team level, not all employment status and number of dependents were included in each teams.

Xiao Lai

## Conclusion and Clinical Implications

The study identifies a high level of heterogeneity among surveyors and a significant association between patients' employment status, number of dependents, and their decisions to undertake breast cancer screening. This insight could guide health professionals in modifying their communication and interventions to increase screening participation.

## Potential Limitations

Several limitations should be noted. Bias may have been introduced by the missing values in the height and weight columns and the exclusion of the dataset owing to the absence of the 'ES' variable. High heterogeneity means that respondents had different patient characteristics, which makes comparisons difficult. The results might not be applicable to other populations as they mostly represent the sampled population. Verifying these results and examining additional decision-influencing factors for breast cancer screening requires more investigation.

**Figure 2**. Distribution of patients' employment status (ES) and number of dependents (Deps) patterns in relation to their decisions about breast

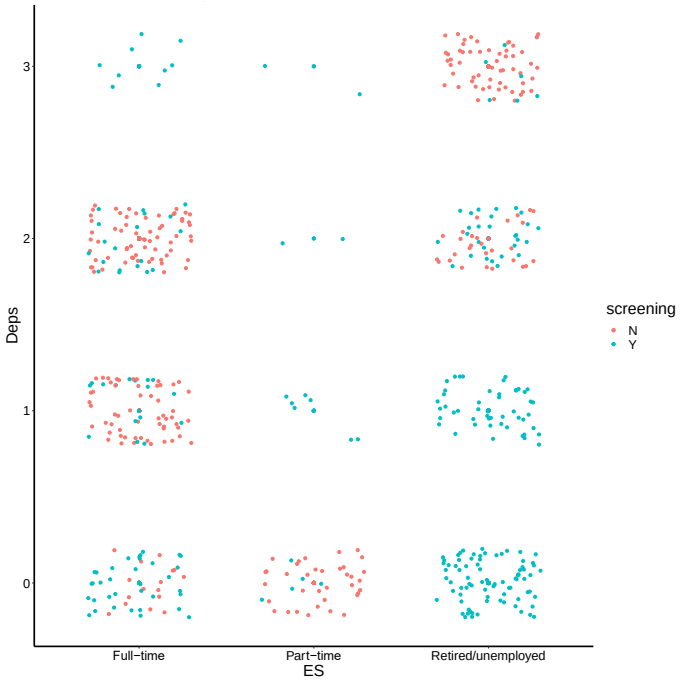



**Figure 3**. Distribution patterns of patients' employment status and number of dependents, as sorted by each clinical team, in relation to their