Effect of marriage on the efficacy of interventions designed to increase mammography screening

Neyan Deng, Zifei Luo, Xiaomeng Xu, Kalpana Das

Background

Worldwide, female breast cancer stands as the most frequently diagnosed cancer, with approximately 2.3 million new cases reported in 2020, constituting 11.7% of all cancer diagnoses [1]. Mammography is widely regarded as the most effective screening method for detecting breast cancers in asymptomatic individuals early on. Studies suggest that it can lead to a notable decrease in the risk of breast cancer mortality, ranging from 15% to 56% [1]. Research indicates that married individuals are more inclined to undergo breast cancer screenings, with 81.6% of married individuals participating compared to 74.7% of unmarried individuals [2]. This observation underscores the beneficial influence of marital status on health behavior decisions and drives.

However, despite increased public awareness of breast cancer screening, compliance with regular screening has not yet reached the desired level. This study aims to assess the comparative effectiveness of two interventions in enhancing mammography screening, specifically examining how marital status influences the efficacy of these interventions. We anticipate that the findings shall help identify more effective intervention strategies tailored to unmarried and married populations.

Data

The data were sourced from a randomized controlled trial assessing the effect of interventions designed to increase mammography screening among nonadherent women over 50. Mammography compliance, measured by whether screening was done 6-months before enrollment, is our primary outcome variable (*resp6*). The mammography cognitive stage, measured by a summary of 3, 11, and 23-month screening behavior, is our secondary outcome (*stagefwup*). Intervention, categorized as control, phone, mail, and phone and mail, is the main explanatory variable (*treatment*).

Information regarding data cleaning, exploratory data analysis, and supplementary statistical analysis results can be found in the appendix.

Methods

For the primary response, we started by constructing a logistic regression model incorporating all covariates along with an interaction term between *treatment* and *married*. The interaction term was included to assess whether the impact of treatment varies depending on a woman's marital status. The possibility of confounding between marital status and treatment was eliminated after conducting a Chi-squared test for independence. Studies indicate strong correlations between income and factors such as education [3] and employment status [4]. Also, studies have highlighted a significant correlation between the recommendation of mammograms by medical professionals and the baseline stage of mammography screening behavior [5]. Chi-square tests for independence corroborated these findings for our dataset too. To mitigate issues of multicollinearity, we opted to exclude education, working status, and doctor/ nurse recommendations received in the last two

years from the comprehensive model. This choice was validated by both the Likelihood Ratio Test (LRT) and the score test. This reduced model was chosen to be our final model for *resp6* (eq. 1).

For the second outcome, we fit a multinomial regression model (eq. 2), $j = 1, 2$. Due to the significant dependence between the two responses based on a Chi-square test for independence, which we performed during the exploratory data analysis, we maintain the same set of covariates as in (eq. 1). To facilitate interpretation, we centered age around its mean for both the models.

$$\begin{aligned} logit(resp6_i) = \beta_0 + \beta_1 I(mail)_i + \beta_2 I(phone)_i + \beta_3 I(phone\ \&\ mail)_i + \beta_4 (age - 65.56)_i \\ + \beta_5 income_i + \beta_6 married_i + \beta_7 history_i + \beta_8 stagebase_i + \beta_9 I(mail : married)_i \\ + \beta_{10} I(phone : married)_i + \beta_{11} I(phone\ \&\ mail : married)_i \end{aligned} \quad (1)$$

$$\begin{aligned} log(\frac{\pi_{ij}}{\pi_{i0}}) = \beta_{0j} + \beta_{1j} I(mail)_i + \beta_{2j} I(phone)_i + \beta_{3j} I(phone\ \&\ mail)_i + \beta_{4j} (age - 65.56)_i \\ + \beta_{5j} income_i + \beta_{6j} married_i + \beta_{7j} history_i + \beta_{8j} stagebase_i + \beta_{9j} I(mail : married)_i \\ + \beta_{10j} I(phone : married)_i + \beta_{11j} I(phone\ \&\ mail : married)_i \end{aligned} \quad (2)$$

$$\begin{aligned} \pi_{i0} &= P(stagefwup = precontemplation|\mathbf{x_i}) \\ \pi_{i1} &= P(stagefwup = contemplation|\mathbf{x_i}) \\ \pi_{i2} &= P(stagefwup = action|\mathbf{x_i}) \end{aligned} \quad (3)$$

Results

For the logit model, the intercept, as well as the coefficients corresponding to income, baseline stage, and the indicator for the interaction between phone and mail treatment and marital status, were found to be statistically significant predictors for mammography compliance (6-month screening) at $\alpha = 0.05$ (based on the Wald test's p-values). Interpreting the intercept, the estimated log odds of compliance at a 6-month mammography screening for a woman of mean age (65.56 years) with control treatment, high income, no family history, who contemplated at baseline stage is -1.53. Compared to high-income individuals, the log odds of compliance at a 6-month mammography screening decreased by 0.56 for a low-income individual, holding all the other covariates constant. Also, compared to individuals who precontemplated at the baseline stage, the log odds of compliance at a 6-month mammography screening increased by 1.45 for individuals who contemplated at the baseline stage, holding all the other covariates constant. Finally, the effect of phone and mail treatment on the log odds of compliance at 6-month mammography screening is additionally increased by 0.86 for unmarried women, adjusting for other covariates. The same can be seen in the upper subfigure in figure 1.

For the multinomial model, the intercept corresponding to *Action* level, and the coefficients corresponding to *Action* and *Contemplation*, were found to be statistically significant predictors for mammography cognitive stage (follow-up) at $\alpha = 0.05$. Interpreting the intercept, the estimated log odds of having a mammogram in the past 12 months for a woman of mean age (65.56 years) with control treatment, high income, no family history who contemplated at baseline stage is -1.34. Compared to individuals who precontemplated at the baseline stage, the log odds of having a mammogram in the past 12 months increased by 3.86 for individuals who contemplated at the baseline stage, given they had a mammogram in the past 12 months, holding all the other covariates constant. Similarly, compared to individuals who precontemplated at the baseline stage, the log odds

of planning to have their first mammogram in the next 6 months increased by 3.31 for individuals who contemplated at the baseline stage, given they never had a mammogram but planned to have one in the next 6 months, holding all the other covariates constant. None of the interaction terms were found to be significant. The same can be seen in the lower subfigure in figure 1.

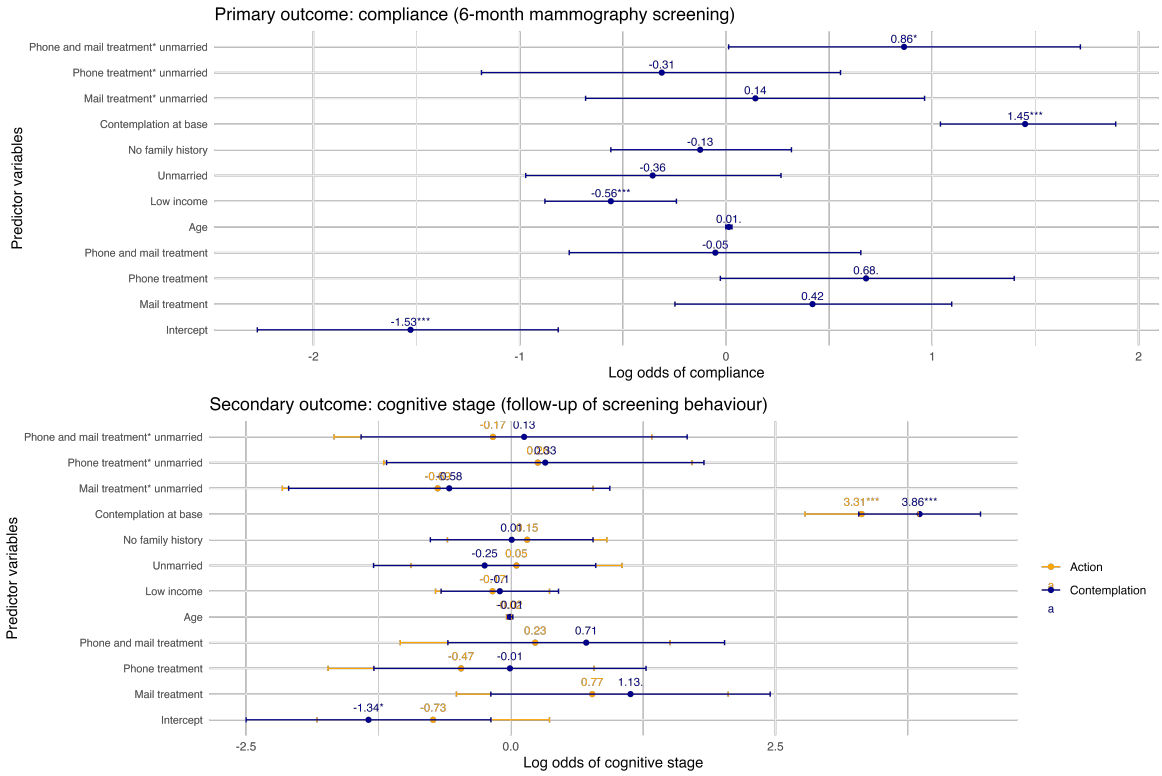

Figure 1: Forest plots for the models for compliance and cognitive stage

Conclusion

The results from logistic regression indicate that while treatment alone doesn't significantly predict either of the outcomes, the interaction term involving phone and mail treatment and marital status emerges as a significant predictor for mammography screening compliance. The effectiveness of phone and mail treatment is contingent upon the marital status of women, aligning closely with findings in existing literature [2]. However, such a relationship isn't observed for other interventions. Furthermore, our analysis suggests that individuals with higher income levels and those at the contemplative stage during baseline are more inclined to adhere to mammography screening at the 6-month mark.

Analysis from the multinomial model reveals that the effectiveness of none of the treatments is contingent upon the marital status of women. Additionally, we infer that individuals in the contemplative stage at baseline are more inclined to have undergone a mammogram in the past 12 months or to have contemplated undergoing one in the next 6 months. A potential avenue for improvement could involve fitting a multinomial model with different covariates tailored specifically for the secondary outcome, rather than utilizing the same set of covariates. This customized approach may yield different results regarding the impact of marital status on the efficacy of interventions.

References

[1] I. J. Nduka, I. L. Ejie, C. E. Okafor, G. U. Eleje, and O. I. Ekwunife, "Interventions to increase mammography screening uptake among women living in low-income and middle-income countries: a systematic review," *BMJ open*, vol. 13, no. 2, p. e066928, 2023.

[2] J. Hanske, C. P. Meyer, J. D. Sammon, T. K. Choueiri, M. Menon, S. R. Lipsitz, J. Noldus, P. L. Nguyen, M. Sun, and Q.-D. Trinh, "The influence of marital status on the use of breast, cervical, and colorectal cancer screening," *Preventive medicine*, vol. 89, pp. 140–145, 2016.

[3] W. Mou, "A quantitative analysis of the relationship between education level and income," *Journal of Education, Humanities and Social Sciences*, vol. 12, pp. 160–166, 2023.

[4] K. Stronks, H. Van De Mheen, J. Van den Bos, and J. P. Mackenbach, "The interrelationship between income, health and employment status." *International journal of epidemiology*, vol. 26, no. 3, pp. 592–600, 1997.

[5] C.-C. Chang, T.-C. Ho, C.-Y. Lien, D. H.-Y. Shen, K.-P. Chuang, H.-P. Chan, M.-H. Yang, and Y.-C. Tyan, "The effects of prior mammography screening on the performance of breast cancer detection in taiwan," in *Healthcare*, vol. 10, no. 6. MDPI, 2022, p. 1037.

Appendix

Data cleaning

We transformed all variables, except *age* (the sole continuous variable), into factor variables. *stagefwup* and *stagebase* were given an order. Finally, as 92% of the dataset were complete observations, we omitted those with missing information for simplicity.

Exploratory data analysis

We looked at the demographics of all variables (table 1). We crafted boxplots to visualize the distribution of age across different outcome categories and to illustrate the distribution of responses, stratified by treatment (figure 2). We also created cross tables for each outcome against *treatment*. Furthermore, we performed a Chi-square test for independence to assess the relationship between the two outcomes, *resp6* and *stagefwup*, and found a dependency between the two outcomes. Additionally, Chi-square tests of independence between marital status and compliance suggested dependence; between marital status and cognitive stage suggested independence; and between marital status and treatment suggested indpendence. The later eliminates the doubt of confounding.

Variable	Overall (N=1042)
Age	
Mean (SD)	65.6 (10.4)
Median [Min, Max]	64.0 [51.0, 97.0]
Treatment	
Control	284 (27.3%)
Mail	275 (26.4%)
Phone	240 (23.0%)
Phone and mail	243 (23.3%)
6-month mammography	
Didn't screen	683 (65.5%)
Screened	359 (34.5%)
Education	
High education	351 (33.7%)
Low education	688 (66.0%)
Missing	3 (0.3%)
Income	
High income	427 (41.0%)
Low income	533 (51.2%)
Missing	82 (7.9%)
Workpay	
Not working	735 (70.5%)
Working	307 (29.5%)
Married	
Married	317 (30.4%)
Unmarried	725 (69.6%)
Mammogram recommendation in the last year or two	
Recommended	773 (74.2%)
Wasn't recommended	269 (25.8%)
Family history	
Family history	123 (11.8%)
No family history	919 (88.2%)
Baseline stage of mammography	
Precontemplation	251 (24.1%)
Contemplation	791 (75.9%)
Follow-up stage of mammography	
Precontemplation	170 (16.3%)
Action	433 (41.6%)
Contemplation	439 (42.1%)

Table 1: Descriptive statistics of the study participants

Additional results

For the logistic regression model, the p-values obtained from LRT (0.3455) and score test (0.3443) were close to each other and greater than 0.05. See table 2 for the corresponding deviance and Akaike Information Criterion (AIC) values. Among the measures of R-squared, Nagelkerke's value

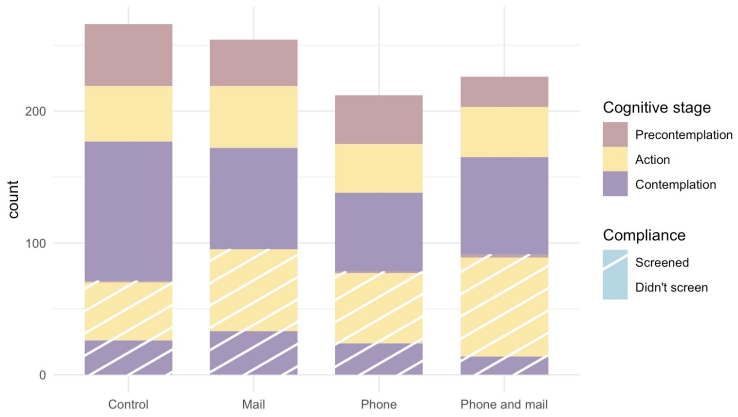

Figure 2: Distribution of both the outcomes and treatment

was the highest at 0.1321; implying that only 13% of the total variation in compliance (6-month screening) is explained by eq. 1 (table 2). Furthermore, the assessment of variance inflation factor (VIF) values within the reduced model indicated no alarming multicollinearity concerns, with all adjusted VIF $\in (1, 5)$. In addition, the Hosmer-Lemeshow test (p-value $= 0.9629$) yielded that the reduced model fit the dataset well. Moreover, the comparison of deviance and AIC values suggested only marginal differences between the reduced and the full model, implying comparable goodness of fit.

For the multinomial regression model as well, both LRT (p-value $= 0.5266$) and Wald test (p-value $= 0.5333$) failed to provide sufficient evidence to conclude that the full model offered a better fit than the reduced model. The Hosmer-Lemeshow test (p-value $= 0.2850$) indicated that the reduced model adequately fits the dataset. Additionally, a comparison of deviance and AIC values suggested a comparable goodness of fit between the reduced and full models. In line with the approach taken for the first model, R-squared values were computed for the second model as well (table 2).

Model	**DoF**	**Deviance**	**AIC**	R^2 **Nagelkerke**
Logistic regression model	946	1143.590	1167.590	0.132
Multinomial regression model	1892	1608.002	1656.002	0.340

Table 2: Goodness of fit measures for the two models

Contributions

Neyan Deng — *code*: logit model, part of data description; *report*: data processing; *poster*: part of background, exploratory data analysis, method
Zifei Luo — *code*: basic multinomial model; *report*: background part; *poster*: part of background and EDA part
Xiaomeng Xu — *code*: 50% logit model; *report*: result part; *poster*: result and conclusion part
Kalpana Das — *code*: exploratory data analysis, multinomial model; *report*: methods, results, conclusion, appendix; *poster*: background, exploratory data analysis, statistical analysis, results, discussion