Breast Cancer Knowledge and Mammography Use in Samoan Women

Emilie Campos
Department of Biostatistics, UCLA

June 12, 2018

Introduction

Using a modified version of the dataset collected for the Samoan American Health study, reported in Mishra Sl, Bastaini R, Crespi CM, Chang C, Luce PH, Banqet CR (2007) [1], we want to model the effects of an intervention on mammogram usage among Samoan women. The article states that there has been a steady increase in the use of mammography in the United States and that improvements in breast cancer screening aid in earlier stage diagnosis. However, they go on to point out that increases in mammography use have not spread to minority and immigrant populations. Therefore, they focused on Samoan women, who have extremely low rates of mammography use.

The study recruited women who had not received a mammogram in the past 2 years and randomized them to intervention and control conditions. The intervention was a community-based participatory cluster-randomized controlled intervention trial. In order to be effective, health education programs must be culturally tailored to minority groups. This was reflected in the three components of the intervention as they created "specially developed English and Samoan language breast cancer educational booklets; skill building and behavioral exercises; and interactive group discussion sessions." [1]

We will be using several regression techniques to assess whether the intervention increased breast cancer knowledge and mammography use in Samoan communities.

Methods

Sample

There were 61 churches included in this dataset with between 1 and 42 participants from each church, for a total of 776 participants. However, we removed the incomplete cases in order to use some of our models, which left us with 767 participants. Each church was randomly assigned to either the intervention or control conditions. The baseline characteristics of the partcipants are supplied in Table 1. We note that the two groups differ by percent of high school graduates and percent insured at the $\alpha = 0.10$ significance level. This is evidence that we should control for these variables in our future models.

Statistical Analysis and Results

Outcome: Breast Cancer Knowledge

We will create two models, one in which we incorrectly assume that each of the observations are independent of each other and one where we account for the correlation of observations using mixed models. This is important because the intervention was implemented within churches and thus we expect a community effect.

Using `glm()` in R, we fit the incorrect model

$$\text{Post-test} = \beta_0 + \beta_1 * \text{Pre-test} + \beta_2 * \text{Intervention} + \beta_3 * \text{Insured} + \beta_4 * \text{Graduated High School}$$

and using `lmer()` in R, we fit the more appropriate model

$$\begin{aligned}\text{Post-test} = \beta_0 + \beta_1 * \text{Pre-test} + \beta_2 * \text{Intervention} + \beta_3 * \text{Insured} \\ + \beta_4 * \text{Graduated High School} + \text{church}_i.\end{aligned}$$

The estimates and their p-values are given in Table 2. We expect the models to have similar estimates, but the estimates from the incorrect model will underestimate the variance, as illustrated by the narrower confidence intervals. From the correct model, we see that the estimate for the intervention is significant. Thus there is sufficient evidence to suggest that the intervention is associated with an increase of the post-test breast cancer knowledge score of 7.63 points on average, controlling for the pre-test score, insurance status, and education level. All else equal, a one-point increase in the pre-test breast cancer knowledge score is associated with a 0.01 point average increase in the post-test score. Controlling for education level, insurance status, and intervention, a woman who is insured will score 1.94 points lower on the post-test on average. Lastly, a high school graduate will score 2.76 points higher on average than a woman who had not graduated from high school, all else equal.

The random intercept indicates how dispersed the church means are around the grand mean and the residual varaince quartifies the variance of indvidual observations within churches. We find that the estimates of the variance of the church component and the residual component are 34.4 and 267.3, respectively. This gives an intraclass correlation of $\frac{34.4}{34.4+267.3} = 0.114$. This means about 11% of the total variance of the post-test breast cancer knowledge score is attributable to variability in the outcome at the church level.

Next, we wanted to test if there was evidence that the test was differentially effective depending on the age of the woman or whether she had ever had a mammogram in the past. Thus we created a model with an interaction term for age and treatment and a separate model with an interaction term for ever having a mammogram and treatment. The results of these models are given in Table 3 and Table 4. Referring to Table 3, it is important to note that when controlling for age, the intervention is no longer a significant predictor. In addition to this, there is not enough evidence to conclude that the intervention was differentially effective depending on the age of the woman. Now referring to Table 4, we see that the intervention is still a significant predictor, but the interaction term is not significant so there is insufficient evidence to conclude the intervention was differentially effective depending on whether a woman had ever had a mammogram in this set of Samoan women.

Outcome: Receipt of Mammogram During Follow Up

Another focus of our analysis was whether or not the intervention increased mammography use in the Samoan women communities. Thus we will be modeling receipt of mammogram during follow up with the predictors: insurance status, education level, and treatment group. We will be creating three different logistic models: one that incorrectly assumes that the observations are independent, one accounting for the correlation of observations using generalized estimating equations, and lastly one accounting for the correlation of observations using mixed models. The independent and GEE models both take on the form:

$$\text{logit}(P(\text{Receipt of Mammogram})) = \beta_0 + \beta_1 * \text{Intervention} + \beta_2 * \text{Insured} + \beta_3 * \text{High School Graduate}$$

and we will be using an exchangeable correlation structure for the GEE model. We the models with the functions `glm` and `geeglm`. Also, the mixed model will take the form:

$$\begin{aligned}\text{logit}(P(\text{Receipt of Mammogram})) = \beta_0 + \beta_1 * \text{Intervention} + \beta_2 * \text{Insured} \\ + \beta_3 * \text{High School Graduate} + \text{church}_i\end{aligned}$$

using the function `glmer`. The results of all three of these models are in table Table 5. Their results are somewhat similar in that they find the intervention to be associated with a woman being more likely to receive a mammogram during the follow up when controlling for insurance status and education level. However, the confidence interval of the odds ratio in the mixed model does contain 1 so there is some evidence to the contrary.

Referring the the results of the GEE model, the intervention is associated with a woman being approximately 1.6 times more likely to receive a mammogram during the follow up when controlling for insurance status and education level. The other two predictors, insured and high school graduate are not statistically significant when controlling for each other.

Referring to the results of the mixed model, the intervention is associated with a woman being approximately 1.7 times more likely to receive a mammogram during the follow up when controlling for insurance status and education level. This is not a significant predictor though. Again, the other two predictors are also not statistically significant when controlling for each other.

Another question we had was whether the intervention was differentially effective based on the age of the woman. From the results in Table 6, we find that the interaction term is insignificant so there is insufficient evidence to reject the hypothesis that the intervention is not differentially effective based on the age of the woman. Lastly, we tested whether previous mammography use affected the effectiveness of the intervention. In Table 7, we find that while controlling for previous mammography use is associated with a woman being 2% more likely to receive a mammogram in the follow up, but it is not associated with making the intervention differentially effective.

Conclusion

To summarize all of the results found in this analysis, we find that the intervention is effective in increasing the breast cancer knowledge score, but does not have a large effect on mammography use in Samoan communities when taking into account the correlation of responses within churches and controlling for insurance status and education level. We also found that age and mammography history do not affect the effectiveness of the intervention.

Tables and Figures

Table 1: Participant Characteristics at Baseline

	Control	Intervention	p
N	382	385	
Age (mean(sd))	55.37 (10.59)	54.61 (10.48)	0.319
Breast Cancer Knowledge Score (mean(sd))	46.93 (15.76)	46.17 (15.58)	0.499
Insured (%)	288 (75.4)	313 (81.3)	0.058
High school graduate or more (%)	56 (14.7)	76 (19.7)	0.077
Ever had a mammogram (%)	158 (41.4)	166 (43.1)	0.675

Table 2: Linear Model Estimates for Breast Cancer Knowledge

	Independent		Mixed	
	Estimate [95% CI]	p	Estimate [95% CI]	p
Pre-test Knowledge Score	-0.01 [-0.08, 0.07]	0.885	0.01 [-0.07, 0.08]	0.899
Intervention	8.38 [5.91, 10.84]	<0.001	7.63 [3.60, 11.64]	<0.001
Insured	-2.00 [-5.01, 1.01]	0.193	-1.94 [-4.83, 0.95]	0.189
High School Graduate or more	2.55 [-0.74, 5.84]	0.129	2.76 [-0.44, 5.95]	0.091

Table 3: Differential Intervention Affect by Age

Predictor	Estimate	Standard Error	p
Pre-test Knowledge Score	0.00	0.04	0.46
Intervention	0.87	6.66	0.45
Insured	-1.76	1.48	0.88
High School Graduate or more	2.27	1.69	0.09
Age	-0.14	0.08	0.96
Intervention x Age	0.12	0.12	0.14

Table 4: Differential Intervention Affect by Previous Mammography

Predictor	Estimate	Standard Error	p
Pre-test Knowledge Score	0.00	0.04	0.453
Intervention	7.87	2.30	<0.001
Insured	-2.01	1.49	0.911
High School Graduate or more	2.71	1.65	0.05
Ever had a mammogram	0.64	1.78	0.36
Intervention x Ever had a mammogram	-0.60	2.46	0.596

Table 5: Odds Ratio Estimates for Mammography During Follow Up

	Independent		GEE		Mixed	
	OR [95% CI]	p	OR [95% CI]	p	OR [95% CI]	p
Intervention	1.93 [1.45, 2.58]	<0.001	1.63 [1.06, 2.49]	0.059	1.74 [0.94, 3.21]	0.069
Insured	1.19 [0.84, 1.71]	0.327	1.20 [0.85, 1.67]	0.382	1.24 [0.84, 1.84]	0.287
High School Graduate	1.38 [0.94, 2.03]	0.099	1.39 [1.01, 1.91]	0.090	1.49 [0.97, 2.31]	0.071

Table 6: Differential Intervention Affect by Age

Predictor	Estimate	Standard Error	p
Intervention	1.43	0.92	0.12
Insured	0.24	0.20	0.24
High School Graduate or more	0.33	0.23	0.15
Age	0.00	0.01	0.91
Intervention x Age	-0.02	0.02	0.31

Table 7: Differential Intervention Affect by Previous Mammography

Predictor	Estimate	Standard Error	p
Intervention	0.62	0.34	0.07
Insured	0.06	0.21	0.78
High School Graduate or more	0.24	0.23	0.29
Ever had a mammogram	1.02	0.24	<0.001
Intervention x Ever had a mammogram	-0.14	0.34	0.67

References

[1] S. I. Mishra, R. Bastani, C. M. Crespi, L. C. Chang, P. H. Luce, and C. R. Baquet, "Results of a randomized trial to increase mammogram usage among Samoan women," *Cancer Epidemiology Biomarkers and Prevention*, vol. 16, no. 12, pp. 2594–2604, 2007.