# Biostatistics 200C: Project 2

*Chad Pickering*

*6/19/2018*

**Introduction.**

It is of global and national interest to understand how effective various cancer education programs are for different groups of individuals. Specifically, we would like to hone our interest to BC (breast cancer) knowledge, particularly in terms of mammographic screenings of Samoan women. A randomized trial was conducted in which Samoan churches in southern California were randomized into intervention and control groups; Samoan women who had not had a mammogram in the past two years were recruited into the study, placed into one of the two groups depending on which church they attend. The intervention group members were placed in a group-based program, where they were contacted 8 months after to record outcomes. We need to understand if the intervention group saw a differential effect of mammogram receival compared to the control group, and whether or not their breast cancer knowledge score was significantly different. To achieve this, we will use GEEs and mixed models that account for the data correlated by church.

**Methods and Discussion.**

The overall mean for baseline BC knowledge score is 46.7 (with a median of 50), and the overall age at baseline is 55.1; no significant differences are detected between intervention and control groups in either variable (Table 1; t-tests). No significant differences are found in the proportions of Samoan women who have had a mammography in the past two years between intervention and control groups (chi-square test, $p = 0.689$). However, the proportions of Samoan women who have health insurance are significantly different between groups (chi-square test, $p = 0.042$), whereas levels of education are borderline significantly different between groups (chi-square test, $p = 0.060$). Using a significance threshold of $\alpha = 0.05$, only health insurance status will be controlled for out of these variables. (It should be noted that controlling for education in addition to insurance status does not significantly change any of the following results.)

When fitting a model to test whether there is a difference between treatment groups for BC knowledge disregarding dependence structure, we find strong evidence of significance ($\beta = 8.547$, $p < 0.001$) when controlling for baseline BC knowledge score and insurance status. When correlation is considered via a mixed model, there is still a significant difference between treatment groups for follow-up BC knowledge when considering churches as clusters ($\beta = 7.882$, $p < 0.001$). This suggests that Samoan women who are allocated into the intervention group score about 8 points higher on follow-up breast cancer knowledge score compared to those in the control group, controlling for baseline knowledge score and insurance status. The intervention effect appears to be quite strong, although not as strong as it is when churches are erroneously not considered as clusters; standard errors increased, thereby reducing significance. The random intercept has variance 33.42, indicating how dispersed the church means are around the grand mean, whereas the residual variance quantifies the variance of individual women within churches, amounting to 265.82. Thus, about 11.2% of the total variance of the BC knowledge score after follow-up is attributable to variability in the outcome at the church level ($ICC = \frac{33.4201}{33.4201+265.82} \approx 0.112$). When the mixed model that considers correlation between churches introduces an interaction term, once for the interaction of age of the woman with treatment and once for the interaction of a woman's history of past mammograms, neither affect whether the intervention is effective, adjusting for insurance status and treatment.

We next fit models to test whether there is a significant difference between groups for receiving a mammogram by the follow-up date. First, we fit a logistic regression disregarding dependence structure and we find strong evidence of a treatment effect ($\beta = 0.680$, $p < 0.001$) when controlling for insurance status. When a generalized estimating equation (GEE) is used, clustering by church is treated as a nuisance, and mean responses are estimated while choosing a covariance structure of the correlated responses. Using this method,

there seems to be borderline significance of a treatment effect ($\beta = 0.507$, $p = 0.050$); the odds of receiving a mammogram by the follow-up date for those in the treatment group is approximately 1.659 times the odds of mammogram receipt for those in the control group, irregardless of church, and controlling for insurance status. In the GEE model, those with insurance had 1.227 times the odds of receiving a mammogram before the follow-up date than those without, controlling for treatment group, and again, irregardless of cluster membership. Additionally, the odds of having mammogram receipt by follow-up given a member of the population has no insurance and is a member of the control group is 0.599. The exchangeable working correlation is 0.1457, suggesting that there is minimal correlation between subjects irregardless of cluster membership.

When a mixed model is used to model the binary responses of mammogram receipt by follow-up instead, estimates of the odds are quite similar even though clustering by church is considered. There seems to be borderline significance of a treatment effect ($\beta = 0.5497$, $p = 0.066$); the odds of receiving a mammogram by the follow-up date for those in the treatment group is approximately 1.733 times the odds of mammogram receipt for those in the control group with church grouping considered, and controlling for insurance status. Those with insurance had 1.262 times the odds of receiving a mammogram before the follow-up date than those without, controlling for treatment group, and again, considering cluster membership. Additionally, the odds of having mammogram receipt by follow-up given a member of the population has no insurance and is a member of the control group is 0.577.

Lastly, when we test via a mixed model whether the intervention is differentially effective depending on the age of the woman or whether she had a mammogram in the past before the study began, we find non-significant results (using a GEE as well, see Table 2).

**Limitations and conclusion.**

There are a few limitations worth mentioning. Churches target a substantially more religious demographic, which introduces a bias in the results; it may be of interest to cluster by another facility type that is not so polarizing. in an additional arm of the study. Also, one could argue that the study has a limited number of explanatory variables; those that are not considered such as family history of cancer, income, or availability of care cannot be adjusted for in the model, and some variability is not explained well.

In conclusion, we find strong evidence of significance in the difference between treatment groups for BC knowledge when controlling for baseline BC knowledge score and insurance status and when correlation is considered via a mixed model. Samoan women who are allocated into the intervention group score about 8 points higher on follow-up breast cancer knowledge score compared to those in the control group, controlling for baseline knowledge score and insurance status. Additionally, there seems to be borderline significance of a treatment effect for receiving a mammogram by the follow-up date regardless of considering church clustering, and controlling for insurance status.

**Tables.**

**Table 1. Summary of baseline characteristics by treatment group.**

| Variable | Interv. Mean (SD) | Ctrl. Mean (SD) | Overall Mean (SD) | p-value |
|---|---|---|---|---|
| Baseline BC knowledge score | 46.3 (15.6) | 47.0 (15.8) | 46.7 (15.7) | 0.534 |
| Mammography in past 2+ years | 166 (0.425) | 158 (0.410) | 324 (0.418) | 0.689 |
| No mammography in past 2+ years | 225 (0.575) | 227 (0.590) | 452 (0.582) | |
| Has health insurance | 319 (0.816) | 291 (0.756) | 610 (0.786) | **0.042** |
| Has no health insurance | 72 (0.184) | 94 (0.244) | 166 (0.214) | |
| High school educ. or better* | 76 (0.194) | 56 (0.145) | 132 (0.170) | 0.060 |
| Less than high school educ.* | 309 (0.790) | 327 (0.849) | 636 (0.820) | |
| Age at baseline | 54.8 (10.6) | 55.4 (10.6) | 55.1 (10.6) | 0.403 |
| Total | 391 | 385 | 776 | |

(*) - Missing data (8 obs.)

**Table 2. Effect of treatment on outcomes**

| Outcome | Model type | Feature | $\beta$ Estimate (SE) | p-value |
|---|---|---|---|---|
| BC knowledge at follow-up | Linear reg. | No correlation structure | 8.547 (1.245) | **<0.001** |
| BC knowledge at follow-up | Mixed | Churches correlated | 7.882 (2.028) | **<0.001** |
| BC knowledge at follow-up | Mixed | Treatment Intxn. w/ age | 0.118 (0.113) | 0.301 |
| BC knowledge at follow-up | Mixed | Treatment Intxn. w/ past mam. | -0.654 (2.442) | 0.789 |
| Mammogram by follow-up | Logistic reg. | No correlation structure | 0.680 (0.146) | **<0.001** |
| Mammogram by follow-up | GEE | Churches correlated (exch.) | 0.507 (0.259) | 0.050 |
| Mammogram by follow-up | Mixed | Churches correlated | 0.550 (0.299) | 0.066 |
| Mammogram by follow-up | GEE | Treatment Intxn. w/ age | -0.012 (0.013) | 0.369 |
| Mammogram by follow-up | Mixed | Treatment Intxn. w/ age | -0.015 (0.015) | 0.334 |
| Mammogram by follow-up | GEE | Treatment Intxn. w/ past mam. | -0.216 (0.264) | 0.413 |
| Mammogram by follow-up | Mixed | Treatment Intxn. w/ past mam. | -0.158 (0.332) | 0.633 |