**Adam Lieberman 5/19/18**

**EEH 506 Data Analysis Project**

**Methods**

This analysis was performed using a subset of the 1987 National Indonesia Contraceptive Prevalence Survey designed to predict the current contraceptive methods used by women based on socio-demographic and economic factors. The primary hypothesis of this analyses is to determine if there is an association between contraception use and wife’s educational attainment. The secondary question is to observe if this association is modified by wife’s religion. Multivariable logistic regression was used to estimate the relative risk (odds ratio (OR), 95% confidence interval (CI) of contraception use in relation to wife’s educational attainment. To assess confounding, I sequentially took out one variable at a time to compare the full model to the partial model for each exposure variable. If the OR changed by more than ten percent it was considered a possible confounder and kept in the final adjusted model. The covariates listed in Table 1 were considered as possible confounders for the association. Parameterizations of age, parity, contraception, and wife’s educational attainment use were created. Age was categorized into a dummy variable with four groups, parity into a dichotomous variable (nulliparous and parous), contraceptive use into a dichotomous variable (any use, or no use) and wife’s educational attainment into a dummy variable with 4 groups (none, primary, secondary, and post-secondary). Logistic diagnostics was then run on the final model to test for model fit, departures from linearity, influential observations, and collinearity. If any outliers or influential observations were detected they were removed from the model. To evaluate the second question multiple logistic regression was used on the final model including the variable religion, as well as an interaction variable for religion and wife’s educational attainment. No participants were missing any data, so none were excluded from the analyses.

Table 1. Distributions of maternal and household characteristics and distributions of contraceptive use according to those characteristics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Characteristic** | **Population Distribution**  **N %** | **Contraception Use** % (N=844) 42.7%  **N %** | **No Contraception Use** %  (N= 629) 57.3%  **N %** | **χ2 p-value** |
| **Age (years)** |  |  |  |  |
| ≤19 | 36 2.4 | 18 2.1 | 18 2.9 | <.0001 |
| 20-29 | 570 38.7 | 334 39.6 | 236 37.5 |  |
| 30-39 | 527 35.8 | 345 40.9 | 182 28.9 |  |
| ≥40 | 340 23.1 | 147 17.4 | 193 30.7 |  |
| **Wife’s Education** |  |  |  | <.0001 |
| None | 152 10.3 | 49 5.8 | 103 16.4 |  |
| Primary | 334 22.7 | 158 18.7 | 176 28.0 |  |
| Secondary | 410 27.8 | 235 27.8 | 175 27.8 |  |
| Post-Secondary | 577 39.2 | 402 47.6 | 175 27.8 |  |
| **Husband’s Education** |  |  |  |  |
| None | 44 3.0 | 13 1.5 | 31 4.9 | <.0001 |
| Primary | 178 12.1 | 79 9.4 | 99 15.7 |  |
| Secondary | 352 23.9 | 191 22.6 | 161 25.6 |  |
| Post-Secondary | 899 61.0 | 561 66.5 | 338 53.7 |  |
| **Parity** |  |  |  | <.0001 |
| Parous | 1,376 93.4 | 842 99.8 | 534 84.9 |  |
| Nulliparous | 97 6.6 | 2 .2 | 95 15.1 |  |
| **Religion** |  |  |  | .0051 |
| Non-Islam | 220 14.9 | 145 17.2 | 75 12.0 |  |
| Islam | 1,253 85.1 | 699 82.8 | 554 88.0 |  |
| **Wife’s Working** |  |  |  | .1308 |
| Yes | 369 25.1 | 199 23.6 | 170 27.0 |  |
| No | 1,104 74.9 | 645 76.4 | 459 73.0 |  |
| **Standard of Living** |  |  |  | <.0001 |
| Low | 129 8.8 | 49 5.8 | 80 12.7 |  |
| Semi-Average | 229 15.5 | 112 13.3 | 117 18.6 |  |
| Average | 431 29.3 | 247 29.3 | 184 29.3 |  |
| High | 684 46.4 | 436 51.7 | 248 39.4 |  |
| **Media Exposure** |  |  |  | <.0001 |
| Good | 1364 92.6 | 809 95.8 | 555 88.2 |  |
| Not Good | 109 37.4 | 35 4.2 | 74 11.8 |  |

**Results**

Descriptive characteristics of maternal socio-demographic and economic factors and characteristics stratified by contraceptive use are shown in Table 1. When observing the raw data you can see a positive trend for women’s educational attainment and use of contraceptives, where the more education these women had, the more likely they were to have used contraceptives. Among women who did not use any contraceptives, there seems to be no trend. Contraceptive users were also more likely to be between the ages of 20-39, have a husband with high educational attainment, have one or more than one child, be Islamic, not currently working, have a high standard of living, and have good media exposure. Based on the χ2 p-values, there are significant differences between contraceptive users and non-users in each category except wife’s working status.

There is a significant increase in crude odds of contraceptive use among women who have had some education compared to those who have not had any education (OR 1.88, CI 1.26-2.82) (Table 2). This association seems to be a dose-response relationship where odds of contraceptive use increases with more education attained. The women with the highest level of education had a much higher odds (OR 4.83, CI 3.29-7.08) of contraceptive use compared to those in the third highest category (OR 2.82, CI 1.90-4.18). When adjusting for confounding, three variables modified the odds ratio by more than 10% and were included in the final model. After adjusting for parity, standard of living, and media exposure, the results are similar to the unadjusted model. Odds of contraceptive use among those with some education to those with none were significantly higher (OR 1.90, CI 1.23-2.94). An analogous dose-response relationship is shown among the different educational levels: highest level (OR 4.86, CI 3.13-7.57) and third highest level (OR 2.89, CI 1.87-4.46).

To identify if wife’s religion modified the relationship between contraception use and wife’s educational attainment, an interaction term was added to the model (Table 3). The odds of contraceptive use was 54% higher in women with any educational attainment relative to none, but this conclusion is not significant p=.73. We can conclude that religion is not an effect modifier and remove it from the final model.

Table 2: Unadjusted and adjusted OR and 95% CI of contraceptive use according to wife’s educational attainment.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Characteristic** | **No. of Cases** | **Total No.** | **Unadjusted**  **OR 95% CI** | **Adjusted a**  **OR 95% CI** |
| **Total** | 844 | 1,473 |  |  |
| **Wife’s Education** |  |  |  |  |
| None | 49 | 152 | 1.0 Referent | 1.0Referent |
| Primary | 158 | 334 | 1.88 1.26-2.82 | 1.90 1.23-2.94 |
| Secondary | 235 | 410 | 2.82 1.90-4.18 | 2.89 1.87-4.46 |
| Post-Secondary | 402 | 577 | 4.83 3.29-7.08 | 4.86 3.13-7.57 |

aAdjusted for parity, standard of living, and media exposure

Table 3: Adjusted OR and 95% CI of contraceptive use according to wife’s educational attainment modified by wife’s religion using an interaction term

|  |  |  |
| --- | --- | --- |
| **Characteristic** | **Interaction OR 95% CI** | **p-value for interaction** |
| **Wife’s Education** | 1.54 1.05-2.25 | .73 |

To evaluate logistic model fit, deviation and pearson residual tests were conducted. Deviance and pearson p-values were both <.0001, which is a general indication for poor model fit. The Hosemer and Lemeshow test gives a p-value of .6328, which signifies no evidence of poor fit. The R2 is .0968, which tells us that about 10% of the deviance in contraceptive use is explained by the set of predictors in our model. This may mean that although contraceptive use may be associated with educational attainment, it may not be a good predictor of it. There was no indication of high leverage points or Dfbetas. The C bar plot had a few observations that were high that may seem influential, but were ruled out as being erroneous after further analyses and kept in the model. When testing for collinearity, all VIF values were close to 1.0 and did not show any sign of collinearity. Overall, we can conclude that although we observe an association between our exposure and outcome variable, this model may not be the best predictor of contraception use.

**SAS CODE**

/\*Create variables: contraception into dichotomous variable, age into dummy variable, parity into dichotomous variable, and education into dummy variable\*/

**data** analysis.contraception;

set analysis.contraception;

if contra GE **2** then contrause= **1**;

if contra LE **1** then contrause= **0**;

if age LE **19** then age1 =**1**;

else age1=**0**;

if age GE **20** and age LE **29** then age2 =**1**;

else age2=**0**;

if age GE **30** and age LE **39** then age3 =**1**;

else age3=**0**;

if age GE **40** then age4 =**1**;

else age4=**0**;

if parity LE **0** then parity1=**0**;

if parity GE **1** then parity1=**1**;

if educ=**1** then educ1=**1**;

else educ1=**0**;

if educ=**2** then educ2=**1**;

else educ2=**0**;

if educ=**3** then educ3=**1**;

else educ3=**0**;

if educ=**4** then educ4=**1**;

else educ4=**0**;

**run**;

/\*Frequency of each variable for Table 1\*/

**proc** **freq** data=analysis.contraception;

tables age1 age2 age3 age4 educ hus\_age parity1 religion working sol media contrause;

**run**;

/\*Sort data by contraception use\*/

**proc** **sort** data=analysis.contraception;

by contrause;

**run**;

/\*Frequency of each variable stratified by contraception use: Table 1\*/

**proc** **freq** data=analysis.contraception;

tables age1 age2 age3 age4 educ hus\_age parity1 religion working sol media contrause;

by contrause;

**run**;

/\*X2 p-values to test for significant differences between contraceptive users and non-users: Table 1\*/

**proc** **freq** data=analysis.contraception;

tables age\*contrause educ\*contrause hus\_age\*contrause parity1\*contrause religion\*contrause working\*contrause

sol\*contrause media\*contrause/all;

**run**;

/\*Crude association between contraception use and wife's educational attainment, with low education as the referent: Table 2\*/

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ2 educ3 educ4;

**run**;

/\*Multiple regression model with all covariates included\*/

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ age hus\_age parity working sol media;

**run**;

/\*Sequentially taking out one variable at a time to see if the OR changes by >10%. If so it will be added to the final model\*/

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ hus\_age parity working sol media;

**run**;

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ age parity working sol media;

**run**;

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ age hus\_age working sol media;

**run**;

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ age hus\_age parity sol media;

**run**;

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ age hus\_age parity working media;

**run**;

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ age hus\_age parity working sol;

**run**;

/\*Final Model after adjusting for parity, standard of living, and media, low education as referent group: Table 2\*/

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ2 educ3 educ4 parity sol media;

**run**;

/\*Testing Model fit\*/

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ parity sol media/iplots lackfit scale=none aggregate rsq;

output out=diagnostics resdev=resdev reschi=reschi;

**run**;

/\*Evaluate specific departures from linearity

resdev for deviation and resci for pearson residuals\*/

**proc** **sgplot** data=diagnostics;

scatter y=resdev x=educ;

title "residual deviances vs. educ";

**proc** **sgplot** data=diagnostics;

scatter y=reschi x=educ;

title "X2 deviances vs. educ";

**proc** **sgplot** data=diagnostics;

scatter y=resdev x=parity;

title "residual deviances vs. parity";

**proc** **sgplot** data=diagnostics;

scatter y=reschi x=parity;

title "X2 deviances vs. parity";

**proc** **sgplot** data=diagnostics;

scatter y=resdev x=sol;

title "residual deviances vs. sol";

**proc** **sgplot** data=diagnostics;

scatter y=reschi x=sol;

title "X2 deviances vs. sol";

**proc** **sgplot** data=diagnostics;

scatter y=resdev x=media;

title "residual deviances vs. media";

**proc** **sgplot** data=diagnostics;

scatter y=reschi x=media;

title "X2 deviances vs. media";

**run**;

/\*Testing for any Influential Observations\*/

**proc** **logistic** data=analysis.contraception;

model contrause=educ parity sol media/influence iplots lackfit scale=none aggregat rsq;

output out=diagnostics2 cbar=cbar;

**run**;

/\*A few observations had high C-Bar values. I will find these observations and look at them individually\*/

**proc** **print** data=diagnostics2;

where cbar>**.04**;

**run**;

/\*we can conclude that these observations are not influential and can keep them in the original model\*/

/\*Testing for collinearity, looking at VIF's\*/

**proc** **reg** data=analysis.contraception;

model contrause=educ parity sol media/vif;

**run**;

/\*All are close to 1.0 no signs of collinearity\*/

/\*Model to test if the association is modified by wife's religion: Table 3\*/;

**proc** **logistic** data=analysis.contraception descending;

model contrause=educ parity religion sol media religion\*educ;

**run**;