## HSA 7708 Spring 2021: Replication Study Assignment

##### Andrew S. Cistola MPH

In 2019, *Diabetes Research and Clinical Practice* published the article titled “Correlates of health care use among White and minority men and women with diabetes: An NHANES study” by Jessie J Wong et al1. The purpose of this study was to find patient level factors that are associated with health care utilization among individuals with diabetes among groups delineated by race/ethnicity and gender.

### Replication Results

This study was replicated using the following data, definitons, and methods. The study was cross-sectional and used data available from the 2015-2016 National Health and Nutrition Examination Survey (NHANES) cohort. The population included 447 non-pregnant adults aged 18-64 with diagnosed diabetes. The independent variables were patient level factors including self-reported health status, depression, comorbidities, income, body mass index, marital status, number of children, insurance coverage, and age. The dependent variable was overall health care use. Poisson regression models were created for white men, minority men, white women, and minority women. Subjective factors were significant for white men, financial factors were significant for minority men, and family structure or mental health was significant for both white and minority women.

The replication study (Table 1 and Table 2) provided similar results, with some differences in what factors were significant for the various subgroups. Each subgroup shared one factor that was significant in both studies, however each subgroup included a factor that was significant in original or replicated results only. The relative magnitude and direction of the associations were consistent. with some deviations. The replicated results generally uphold the original conclusions with minor deviations likely due to the differences in computational methods (original study used Math+, replication study used R version 3.6) or definitions used to modify the raw NHANES data. Code, data and documentation are available to review for the replicated study2.

### Strengths and Weaknesses

The methodology used by the original study was appropriate for a number of reasons. The study utilized regression modeling among subgroups to identify population specific estimates rather than adjusting for these demographics. This allowed for a more precise understanding of the possible effects. Due to the skewed distribution of utilization data, Poisson regression as appropriate. However, specific tests of regression assumptions were not reported and the use of Poisson regression was only mentioned in the abstract. Similarly, the primary weakness of the study was the lack of investigation into the possible combined effects of the various independent variables. Seeing that there are variables that are very similar (marital status and children, or income and insurance) it is very possible that these variables may be confounding a more accurate effect. This could be evaluated with a collinearity matrix or other method for displaying collinearity, but the authors did not provide that information.

### Alternative Methods

In order to answer the same research question, the following approach would also be appropriate. Before creating the final Poisson regression models, the collinearity of the independent variables would be evaluated. Those that are clearly colinear would require alternative methods for inclusion or be a candidate for a possible interaction term. Assuming collinearity is identified, a possible approach would be to redefine the variables of interest by:

1) Creating a composite index of the objective health status variables (similar to the neighborhood deprivation index)

2) Recoding the marital and family status variables as one multi-level category defined as 0 = married, no children; 1 = married with children; 2 = unmarried

3) Including an interaction term for income and insurance

4) Age, BMI, and self-rated health would remain the same

These recodings would be validated based on previous studies and evidence of the direction of effect from bivariate statistics. However, these new variables would allow for a more precise understanding of the general effects pertinent to the research question using the same Poisson regression model for each subgroup. Alongside the Poisson regression models for each subgroup, a hierarchical linear model for all groups could be utilized to identify the comparative effect of subgroup identification on utilization. By categorizing the four subgroups as random effects, an intra-class coefficient could be calculated to identify how much of the variation in the data is accounted by these population differences. This information would help compliment the findings from the individual models and show how strong the differences among groups may be when compared to patient level factors.

### Tables

Table 1. Descriptive statistics of study variables within full and subsamples.

|  |  |
| --- | --- |
| Subsample | N |
| Total | 462 |
| White men | 59 |
| White women | 60 |
| Minority men | 177 |
| Minority women | 166 |

Mean

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subsample | Income | Married | Children | BMI | Insurance | Health | Utilization | Depression | Comorbidity |
| Total | 2.1262 | 0.5519 | 0.8593 | 32.3554 | 0.8290 | 3.5844 | 3.3074 | 4.0000 | 2.0779 |
| White men | 1.6671 | 0.7627 | 0.8814 | 29.3627 | 0.7119 | 3.7966 | 2.5593 | 3.2881 | 1.5763 |
| White women | 1.8155 | 0.6167 | 1.3500 | 34.3967 | 0.7333 | 3.9000 | 3.5833 | 4.1167 | 1.7167 |
| Minority men | 2.4733 | 0.5593 | 0.7571 | 31.9531 | 0.8757 | 3.4633 | 3.3051 | 3.4237 | 2.0508 |
| Minority women | 2.0314 | 0.4458 | 0.7831 | 33.1102 | 0.8554 | 3.5241 | 3.4759 | 4.8253 | 2.4157 |

Standard Deviations

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subsample | Income | Married | Children | BMI | Insurance | Health | Utilization | Depression | Comorbidity |
| Total | 1.5103 | 0.4978 | 1.2126 | 10.9400 | 0.3769 | 0.9948 | 1.9860 | 5.1847 | 1.7334 |
| White men | 1.1808 | 0.4291 | 1.3272 | 9.8899 | 0.4568 | 0.7830 | 1.4887 | 4.8745 | 1.7734 |
| White women | 1.3099 | 0.4903 | 1.4241 | 11.0593 | 0.4459 | 0.8577 | 2.1885 | 4.7339 | 1.6060 |
| Minority men | 1.6194 | 0.4979 | 1.2307 | 10.0324 | 0.3309 | 1.1130 | 2.0247 | 5.0403 | 1.5604 |
| Minority women | 1.4911 | 0.4986 | 1.0216 | 11.9770 | 0.3527 | 0.9452 | 1.9747 | 5.5147 | 1.8754 |

Table 2. Standardized regression coefficients and p-values for health care use among White and minority men and women with diabetes.

Coefficients

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subsample | Income | Married | Children | BMI | Insurance | Health | Utilization | Depression | Comorbidity |
| Total | 0.3897 | -0.0194 | -0.0810 | 0.0054 | -0.0003 | 0.4375 | 0.0783 | 0.0156 | 0.0734 |
| White men | -0.1740 | -0.0491 | -0.3525 | -0.0355 | 0.0148 | 0.5323▼ | 0.1357 | 0.0056 | 0.0598 |
| White women | 0.2591 | 0.0063\* | -0.1741 | 0.0666 | -0.0004 | 0.4364┼ | 0.0928 | 0.0195 | 0.1221▼ |
| Minority men | 0.5217▼ | -0.062┼ | 0.0693 | -0.0566 | -0.0041\* | 0.5562┼ | 0.0702 | 0.0147 | 0.0662┼ |
| Minority women | 0.4766 | 0.0111\* | -0.1165 | 0.0424 | 0.0003\* | 0.2230 | 0.0838 | 0.0135\* | 0.0721▼ |

Standard Errors

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subsample | Income | Married | Children | BMI | Insurance | Health | Utilization | Depression | Comorbidity |
| Total | 0.1549 | 0.0193 | 0.0532 | 0.0227 | 0.0023 | 0.0825 | 0.0291 | 0.0049 | 0.0143 |
| White men | 0.6184 | 0.0906 | 0.2270 | 0.0762 | 0.0094 | 0.2261▼ | 0.1331 | 0.0204 | 0.0494 |
| White women | 0.4328 | 0.0607\* | 0.1494 | 0.0527 | 0.0066 | 0.1886┼ | 0.0972 | 0.0146 | 0.0415▼ |
| Minority men | 0.2602▼ | 0.0306 | 0.0904 | 0.0389 | 0.0040\* | 0.1658 | 0.0407 | 0.0085 | 0.0274 |
| Minority women | 0.2651 | 0.0319 | 0.0886 | 0.0413 | 0.0036 | 0.1357 | 0.0546 | 0.0079 | 0.0219▼ |

*p-*values

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subsample | Income | Married | Children | BMI | Insurance | Health | Utilization | Depression | Comorbidity |
| Total | 0.0119 | 0.3130 | 0.1273 | 0.8119 | 0.8940 | 0.0000 | 0.0072 | 0.0015 | 0.0000 |
| White men | 0.7784 | 0.5878 | 0.1204 | 0.6413 | 0.1136 | 0.0186▼ | 0.3079 | 0.7841 | 0.2254 |
| White women | 0.5494 | 0.9176\* | 0.2440 | 0.2060 | 0.9577 | 0.0207┼ | 0.3398 | 0.1828 | 0.0033▼ |
| Minority men | 0.0449▼ | 0.0423┼ | 0.4432 | 0.1459 | 0.3092\* | 0.0008┼ | 0.0843 | 0.0824 | 0.0156┼ |
| Minority women | 0.0722 | 0.7276\* | 0.1886 | 0.3048 | 0.9421\* | 0.1004 | 0.1252 | 0.0854\* | 0.0010▼ |

\* = significant in original results only

┼ = significant in replicated results only

▼ = significant in both original and replicated results

### References

Wong JJ, Hood KK, Breland JY. Correlates of health care use among White and minority men and women with diabetes: An NHANES study. Diabetes Res Clin Pract. 2019;150:122-128. <doi:10.1016/j.diabres.2019.03.001>

Cistola AS. HSR/NHANES. *GitHub.* https://github.com/andrewcistola/HSR/NHANES