Supplementary Materials for Manuscript of MADA Project

Supplement Materials

Shiwani Sapkota

5/4/23

1 Overview

In this supplement file, you can find details about the data file extraction and codes involved in the process of data extraction and data analysis.

2 Code and file information

```
# Skip this code chunk if you already have `mammogram_ses1.rds` raw data file
# NHIS provides .xml and .dat file while extracting the data from website
# Commenting these codes as they were used for data extraction and data preparation proces

# Preparing and loading the raw data file and saving as .csv and .rds files
# ddi <- read_ipums_ddi("nhis_00003.xml")
# data <- read_ipums_micro(ddi)

# Saving the raw data file as .csv file and commenting out this code so it does not get cr
# readr::write_csv(here("data", "raw_data", "mammogram_ses.csv")

# It seems the created csv file is very large (153 MB) and GitKraken won't like it so lets

# Saving the raw data file as .rds file and commenting out this code so it does not get cr
# saveRDS(here("data", "raw_data", "mammogram_ses1.rds")

# From here, you can follow general steps involved in data analysis and the
# respective codes are documented in the `code` folder with the given subfolders:
```

```
# 1. processing_cleaning_code folder
# 2. analysis_code folder
```

3 List of Tables

4 Additional results

5 Exploratory Analysis

6 Supplementary Table 1

Characteristic	non-hispanic, N = 4,883	hispanic, N = 839	non-hispanic, N = 4,953	hispanic, N = 819	non-hispanic, N = 4,819	hispanic, N = 979	non-hispanic, N = 6,401	hispanic, N = 1,155
mammogram_status	4,191 (86%)	646 (77%)	4,291 (87%)	633 (77%)	4,271 (89%)	821 (84%)	5,679 (89%)	934 (81%)
age								
40-44 years	1,101 (23%)	218 (26%)	969 (20%)	217 (26%)	798 (17%)	231 (24%)	871 (14%)	259 (22%)
45-49 years	959 (20%)	176 (21%)	959 (19%)	173 (21%)	848 (18%)	209 (21%)	953 (15%)	214 (19%)
50-54 years	857 (18%)	137 (16%)	846 (17%)	157 (19%)	829 (17%)	157 (16%)	967 (15%)	183 (16%)
55-59 years	624 (13%)	94 (11%)	763 (15%)	103 (13%)	749 (16%)	131 (13%)	1,049 (16%)	155 (13%)
60-64 years	507 (10%)	75 (8.9%)	579 (12%)	69 (8.4%)	707 (15%)	108 (11%)	981 (15%)	118 (10%)
65-69 years	419 (8.6%)	70 (8.3%)	423 (8.5%)	60 (7.3%)	503 (10%)	81 (8.3%)	892 (14%)	141 (12%)
70-74 years	416 (8.5%)	69 (8.2%)	414 (8.4%)	40 (4.9%)	385 (8.0%)	62 (6.3%)	688 (11%)	85 (7.4%)
income at or above poverty threshold	4,341 (89%)	594 (71%)	4,431 (89%)	614 (75%)	4,177 (87%)	700 (72%)	5,596 (87%)	842 (73%)

Characteristic	non-hispanic, N = 4,883	hispanic, N = 839	non-hispanic, N = 4,953	hispanic, N = 819	non-hispanic, N = 4,819	hispanic, N = 979	non-hispanic, N = 6,401	hispanic, N = 1,155
below poverty threshold	542 (11%)	245 (29%)	522 (11%)	205 (25%)	642 (13%)	279 (28%)	805 (13%)	313 (27%)
education								
never attended/kindergarten only	7 (0.1%)	23 (2.7%)	12 (0.2%)	15 (1.8%)	14 (0.3%)	24 (2.5%)	12 (0.2%)	19 (1.6%)
less than high school	617 (13%)	386 (46%)	512 (10%)	331 (40%)	408 (8.5%)	357 (36%)	415 (6.5%)	402 (35%)
high school	1,686 (35%)	202 (24%)	1,549 (31%)	224 (27%)	1,378 (29%)	264 (27%)	1,653 (26%)	294 (25%)
some college or Associate degree	1,399 (29%)	147 (18%)	1,511 (31%)	153 (19%)	1,575 (33%)	213 (22%)	2,117 (33%)	285 (25%)
bachelor degree and higher	1,174 (24%)	81 (9.7%)	1,369 (28%)	96 (12%)	1,444 (30%)	121 (12%)	2,204 (34%)	155 (13%)
marital_status								
never married	461 (9.4%)	90 (11%)	480 (9.7%)	90 (11%)	550 (11%)	126 (13%)	735 (11%)	150 (13%)
married	2,446 (50%)	407 (49%)	2,486 (50%)	380 (46%)	2,366 (49%)	488 (50%)	3,086 (48%)	567 (49%)
single (separated/widowed/divorced)	1,839 (38%)	313 (37%)	1,846 (37%)	315 (38%)	1,724 (36%)	334 (34%)	2,337 (37%)	395 (34%)
living with partner	137 (2.8%)	29 (3.5%)	141 (2.8%)	34 (4.2%)	179 (3.7%)	31 (3.2%)	243 (3.8%)	43 (3.7%)
region_residence								
northeast	937 (19%)	159 (19%)	855 (17%)	117 (14%)	830 (17%)	150 (15%)	1,138 (18%)	175 (15%)
north central/midwest	1,235 (25%)	37 (4.4%)	1,318 (27%)	52 (6.3%)	1,166 (24%)	54 (5.5%)	1,455 (23%)	77 (6.7%)
south	1,822 (37%)	321 (38%)	1,829 (37%)	342 (42%)	1,775 (37%)	363 (37%)	2,252 (35%)	399 (35%)
west	889 (18%)	322 (38%)	951 (19%)	308 (38%)	1,048 (22%)	412 (42%)	1,556 (24%)	504 (44%)

Characteristic	non-hispanic, N = 4,883	hispanic, N = 839	non-hispanic, N = 4,953	hispanic, N = 819	non-hispanic, N = 4,819	hispanic, N = 979	non-hispanic, N = 6,401	hispanic, N = 1,155
insurance_status								
no	4,370	614	4,441	582	4,224	683	6,002	918
coverage	(89%)	(73%)	(90%)	(71%)	(88%)	(70%)	(94%)	(79%)
coverage	513	225	512	237	595	296	399	237
	(11%)	(27%)	(10%)	(29%)	(12%)	(30%)	(6.2%)	(21%)
health_status								
excellent/very	4,161	621	4,166	615	4,017	752	5,391	870
good/good	(85%)	(74%)	(84%)	(75%)	(83%)	(77%)	(84%)	(75%)
fair/poor	722	218	787	204	802	227	1,010	285
	(15%)	(26%)	(16%)	(25%)	(17%)	(23%)	(16%)	(25%)
usual_medicalcare_status								
no place	355	120	343	147	409	178	389	138
	(7.3%)	(14%)	(6.9%)	(18%)	(8.5%)	(18%)	(6.1%)	(12%)
usual	4,528	719	4,610	672	4,410	801	6,012	1,017
place	(93%)	(86%)	(93%)	(82%)	(92%)	(82%)	(94%)	(88%)
smoking_status								
never	2,511	579	2,740	607	2,775	736	3,758	876
	(51%)	(69%)	(55%)	(74%)	(58%)	(75%)	(59%)	(76%)
current	1,137	118	980	110	938	112	1,115	115
	(23%)	(14%)	(20%)	(13%)	(19%)	(11%)	(17%)	(10.0%)
former	1,235	142	1,233	102	1,106	131	1,528	164
	(25%)	(17%)	(25%)	(12%)	(23%)	(13%)	(24%)	(14%)
alcohol_status								
never	1,219	377	1,215	352	1,103	411	1,280	449
	(25%)	(45%)	(25%)	(43%)	(23%)	(42%)	(20%)	(39%)
current	2,834	335	2,901	342	2,906	420	4,005	532
	(58%)	(40%)	(59%)	(42%)	(60%)	(43%)	(63%)	(46%)
former	830	127	837	125	810	148	1,116	174
	(17%)	(15%)	(17%)	(15%)	(17%)	(15%)	(17%)	(15%)
diabetes_status	436	108	596	133	658	209	942	239
	(8.9%)	(13%)	(12%)	(16%)	(14%)	(21%)	(15%)	(21%)

7 Supplementary Table 2

Characteristic	**non-hispanic**, N = 2,624	**hispanic**, N = 758	**non-hispanic**, N = 18,432	**hispanic**, N = 3,034
year				
2000	692 (26%)	193 (25%)	4,191 (23%)	646 (21%)
2005	662 (25%)	186 (25%)	4,291 (23%)	633 (21%)
2010	548 (21%)	158 (21%)	4,271 (23%)	821 (27%)
2015	722 (28%)	221 (29%)	5,679 (31%)	934 (31%)
age				
40-44 years	1,076 (41%)	344 (45%)	2,663 (14%)	581 (19%)
45-49 years	483 (18%)	156 (21%)	3,236 (18%)	616 (20%)
50-54 years	314 (12%)	88 (12%)	3,185 (17%)	546 (18%)
55-59 years	242 (9.2%)	62 (8.2%)	2,943 (16%)	421 (14%)
60-64 years	179 (6.8%)	46 (6.1%)	2,595 (14%)	324 (11%)
65-69 years	173 (6.6%)	36 (4.7%)	2,064 (11%)	316 (10%)
70-74 years	157 (6.0%)	26 (3.4%)	1,746 (9.5%)	230 (7.6%)
income				
at or above poverty threshold	2,075 (79%)	500 (66%)	16,470 (89%)	2,250 (74%)
below poverty threshold	549 (21%)	258 (34%)	1,962 (11%)	784 (26%)
education				
never attended/kindergarten only	16 (0.6%)	30 (4.0%)	29 (0.2%)	51 (1.7%)
less than high school	391 (15%)	346 (46%)	1,561 (8.5%)	1,130 (37%)
high school	928 (35%)	193 (25%)	5,338 (29%)	791 (26%)
some college or Associate degree	726 (28%)	121 (16%)	5,876 (32%)	677 (22%)
bachelor degree and higher	563 (21%)	68 (9.0%)	5,628 (31%)	385 (13%)
marital_status				
never married	458 (17%)	113 (15%)	1,768 (9.6%)	343 (11%)
married	1,119 (43%)	366 (48%)	9,265 (50%)	1,476 (49%)
single (separated/widowed/divorced)	932 (36%)	238 (31%)	6,814 (37%)	1,119 (37%)
living with partner	115 (4.4%)	41 (5.4%)	585 (3.2%)	96 (3.2%)
region_residence				
northeast	393 (15%)	84 (11%)	3,367 (18%)	517 (17%)
north central/midwest	618 (24%)	60 (7.9%)	4,556 (25%)	160 (5.3%)
south	1,037 (40%)	318 (42%)	6,641 (36%)	1,107 (36%)
west	576 (22%)	296 (39%)	3,868 (21%)	1,250 (41%)

Characteristic	**non-hispanic**, N = 2,624	**hispanic**, N = 758	**non-hispanic**, N = 18,432	**hispanic**, N = 3,034
insurance_status				
no coverage	2,004 (76%)	405 (53%)	17,033 (92%)	2,392 (79%)
coverage	620 (24%)	353 (47%)	1,399 (7.6%)	642 (21%)
health_status				
excellent/very good/good	2,187 (83%)	600 (79%)	15,548 (84%)	2,258 (74%)
fair/poor	437 (17%)	158 (21%)	2,884 (16%)	776 (26%)
usual_medicalcare_status				
no place	510 (19%)	242 (32%)	986 (5.3%)	341 (11%)
usual place	2,114 (81%)	516 (68%)	17,446 (95%)	2,693 (89%)
smoking_status				
never	1,384 (53%)	590 (78%)	10,400 (56%)	2,208 (73%)
current	793 (30%)	103 (14%)	3,377 (18%)	352 (12%)
former	447 (17%)	65 (8.6%)	4,655 (25%)	474 (16%)
alcohol_status				
never	787 (30%)	396 (52%)	4,030 (22%)	1,193 (39%)
current	1,397 (53%)	276 (36%)	11,249 (61%)	1,353 (45%)
former	440 (17%)	86 (11%)	3,153 (17%)	488 (16%)
diabetes_status	246 (9.4%)	90 (12%)	2,386 (13%)	599 (20%)

8 Model Fitting

9 Bivariate Logistic Regression

10 Supplementary Table 3

```
# A tibble: 4 x 3
  term                `odds ratio`     p.value
  <chr>                      <dbl>       <dbl>
1 (Intercept)                 5.47 0
2 as.factor(year)2005         1.06 0.246
3 as.factor(year)2010         1.32 0.000000328
4 as.factor(year)2015         1.28 0.000000790
```

11 Supplementary Table 4

```
# A tibble: 2 x 3
  term                                `odds ratio`  p.value
  <chr>                                      <dbl>    <dbl>
1 (Intercept)                                7.02  0
2 as.factor(hispanic_status)Hispanic         0.570 7.09e-35
```

12 Supplementary Table 5

```
# A tibble: 7 x 3
  term                        `odds ratio`   p.value
  <chr>                              <dbl>     <dbl>
1 (Intercept)                         2.28 1.28e-148
2 as.factor(age)45-49 years           2.64 3.80e- 74
3 as.factor(age)50-54 years           4.06 1.98e-115
4 as.factor(age)55-59 years           4.84 1.03e-119
5 as.factor(age)60-64 years           5.68 4.04e-115
6 as.factor(age)65-69 years           4.98 2.94e- 92
7 as.factor(age)70-74 years           4.73 4.11e- 77
```

13 Supplementary Table 6

```
# A tibble: 2 x 3
  term                                      `odds ratio`  p.value
  <chr>                                            <dbl>    <dbl>
1 (Intercept)                                      7.27  0
2 as.factor(income)Below poverty threshold         0.468 3.02e-63
```

14 Supplementary Table 7

```
# A tibble: 5 x 3
  term                                                `odds ratio`  p.value
  <chr>                                                      <dbl>    <dbl>
1 (Intercept)                                                 1.74 2.78e- 3
2 as.factor(education)Less than high school                   2.10 9.20e- 5
3 as.factor(education)High school                             3.14 1.08e- 9
4 as.factor(education)Some college or Associate degree        4.45 2.50e-15
```

```
5 as.factor(education)Bachelor degree and higher               5.48 3.07e-19
```

15 Supplementary Table 8

```
# A tibble: 4 x 3
  term                                                    odds ~1   p.value
  <chr>                                                     <dbl>     <dbl>
1 (Intercept)                                               3.70 4.08e-169
2 as.factor(marital_status)Married                          1.96 1.31e- 34
3 as.factor(marital_status)Single (separated/widowed/divorced)  1.83 8.96e- 27
4 as.factor(marital_status)Living with partner              1.18 9.83e-  2
# ... with abbreviated variable name 1: `odds ratio`
```

16 Supplementary Table 9

```
# A tibble: 4 x 3
  term                                          `odds ratio`  p.value
  <chr>                                                <dbl>    <dbl>
1 (Intercept)                                          8.14  0
2 as.factor(region_residence)North Central/Midwest     0.854 1.32e- 2
3 as.factor(region_residence)South                     0.702 4.72e-10
4 as.factor(region_residence)West                      0.721 7.25e- 8
```

17 Supplementary Table 10

```
# A tibble: 2 x 3
  term                                `odds ratio`   p.value
  <chr>                                      <dbl>     <dbl>
1 (Intercept)                                8.06  0
2 as.factor(insurance_status)Coverage        0.260 9.97e-201
```

18 Supplementary Table 11

```
# A tibble: 2 x 3
  term                               `odds ratio` p.value
  <chr>                                     <dbl>   <dbl>
1 (Intercept)                               6.39    0
```

```
2 as.factor(health_status)Fair/poor        0.963   0.436
```

19 Supplementary Table 12

```
# A tibble: 2 x 3
  term                                          `odds ratio`   p.value
  <chr>                                                <dbl>     <dbl>
1 (Intercept)                                           1.76 1.53e- 35
2 as.factor(usual_medicalcare_status)Usual place        4.34 2.02e-188
```

20 Supplementary Table 13

```
# A tibble: 3 x 3
  term                              `odds ratio`  p.value
  <chr>                                    <dbl>    <dbl>
1 (Intercept)                              6.39  0
2 as.factor(smoking_status)Current         0.652 4.93e-22
3 as.factor(smoking_status)Former          1.57  7.37e-18
```

21 Supplementary Table 14

```
# A tibble: 3 x 3
  term                              `odds ratio`  p.value
  <chr>                                    <dbl>    <dbl>
1 (Intercept)                               4.42 0
2 as.factor(alcohol_status)Current          1.71 4.26e-38
3 as.factor(alcohol_status)Former           1.57 2.12e-15
```

22 Supplementary Table 15

```
# A tibble: 2 x 3
  term                            `odds ratio`  p.value
  <chr>                                  <dbl>    <dbl>
1 (Intercept)                             6.07 0
2 as.factor(diabetes_status)Yes           1.46 3.50e-10
```

23 Supplementary Table 16

```
# A tibble: 11 x 3
   term                                `odds ratio`   p.value
   <chr>                                      <dbl>     <dbl>
 1 (Intercept)                                2.36  4.02e- 78
 2 as.factor(year)2005                        1.02  7.65e-  1
 3 as.factor(year)2010                        1.22  4.30e-  4
 4 as.factor(year)2015                        1.09  9.95e-  2
 5 as.factor(age)45-49 years                  2.61  4.66e- 72
 6 as.factor(age)50-54 years                  3.98  1.74e-111
 7 as.factor(age)55-59 years                  4.67  2.31e-113
 8 as.factor(age)60-64 years                  5.41  7.59e-108
 9 as.factor(age)65-69 years                  4.81  3.33e- 87
10 as.factor(age)70-74 years                  4.55  8.09e- 73
11 as.factor(hispanic_status)Hispanic         0.627 9.86e- 23
```

24 Supplementary Table 17

```
# A tibble: 19 x 3
   term                                                  odds ~1   p.value
   <chr>                                                   <dbl>     <dbl>
 1 (Intercept)                                             0.548 4.65e-  3
 2 as.factor(age)45-49 years                               2.79  2.74e- 78
 3 as.factor(age)50-54 years                               4.34  3.62e-120
 4 as.factor(age)55-59 years                               5.35  8.27e-128
 5 as.factor(age)60-64 years                               6.37  1.54e-123
 6 as.factor(age)65-69 years                               5.85  1.20e-103
 7 as.factor(age)70-74 years                               5.88  2.66e- 91
 8 as.factor(hispanic_status)Hispanic                      0.935 2.10e-  1
 9 as.factor(income)Below poverty threshold                0.663 4.86e- 15
10 as.factor(education)Less than high school               2.18  9.19e-  5
11 as.factor(education)High school                         3.03  2.59e-  8
12 as.factor(education)Some college or Associate degree    4.73  9.92e- 15
13 as.factor(education)Bachelor degree and higher          5.95  1.56e- 18
14 as.factor(marital_status)Married                        1.69  7.29e- 19
15 as.factor(marital_status)Single (separated/widowed/divorce~ 1.43  5.63e-  9
16 as.factor(marital_status)Living with partner            1.24  4.17e-  2
17 as.factor(region_residence)North Central/Midwest        0.778 1.76e-  4
18 as.factor(region_residence)South                        0.704 3.70e-  9
19 as.factor(region_residence)West                         0.708 8.30e-  8
```

```
# ... with abbreviated variable name 1: `odds ratio`
```

25 Supplementary Table 18

```
# A tibble: 5 x 3
  term                                            `odds ratio`  p.value
  <chr>                                                  <dbl>    <dbl>
1 (Intercept)                                            3.10  5.24e-90
2 as.factor(insurance_status)Coverage                    0.371 7.19e-87
3 as.factor(health_status)Fair/poor                      0.911 7.56e- 2
4 as.factor(usual_medicalcare_status)Usual place         2.70  4.27e-69
5 as.factor(diabetes_status)Yes                          1.37  1.16e- 6
```

26 Supplementary Table 19

```
# A tibble: 5 x 3
  term                                `odds ratio`  p.value
  <chr>                                      <dbl>    <dbl>
1 (Intercept)                                4.60  0
2 as.factor(smoking_status)Current           0.575 1.56e-33
3 as.factor(smoking_status)Former            1.36  8.04e- 9
4 as.factor(alcohol_status)Current           1.76  1.99e-39
5 as.factor(alcohol_status)Former            1.61  5.12e-16
```

27 List of Figures

28 Machine Learning (Model Selection)

29 Supplementary Figure 1