**Supplementary Materials**

**Data S1. SAS code for collapsing categories in procedure types.**

PR1C=PRCCS1; /\*principal procedure grouping\*/

if PRCCS1 in (1:9, 199) then PR1C=1; /\*neuro surgery\*/

if PRCCS1 in (10:12) then PR1C=2; /\*endocrinology\*/

if PRCCS1 in (13:21, 220) then PR1C=3; /\*ophtalmology\*/

if PRCCS1 in (22:35) then PR1C=4; /\*otorhinoaryngology (ear, nose, throat)\*/

if PRCCS1 in (36:42) then PR1C=5; /\*thoracic surgery\*/

if PRCCS1 in (43:50, 201:203) then PR1C=6; /\*cardiac surgery\*/

if PRCCS1 in (51:63) then PR1C=7; /\*(cardio)vascular surgery\*/

if PRCCS1 in (64:67) then PR1C=8; /\*hematology\*/

if PRCCS1 in (68:99) then PR1C=9; /\*gastro intestinal procedures\*/

if PRCCS1 in (100:118, 200) then PR1C=10; /\*urology\*/

if PRCCS1 in (119:132) then PR1C=11; /\*gynecology\*/

if PRCCS1 in (133:141) then PR1C=12; /\*obstetrics\*/

if PRCCS1 in (142:164) then PR1C=13; /\*orthopedic/trauma surgery\*/

if PRCCS1 in (165:167) then PR1C=14; /\*gyn\*/

if PRCCS1 in (168:176) then PR1C=15; /\*dermatology\*/

if PRCCS1 in (177:198) then PR1C=16; /\*radiology\*/

if PRCCS1 in (204:205, 216:217) then PR1C=17; /\*anesthesia/ICU\*/

if PRCCS1 in (206) then PR1C=18; /\*lab\*/

if PRCCS1 in (207:211) then PR1C=19; /\*nuclear medicine\*/

if PRCCS1 in (212:215) then PR1C=20; /\*physical therapy\*/

if PRCCS1 in (218:219) then PR1C=21;

if PRCCS1 in (221) then PR1C=22;

if PRCCS1 in (222) then PR1C=23;

if PRCCS1 in (223) then PR1C=24;

if PRCCS1 in (224) then PR1C=25; Predictive variables used for the imputation included patient age, mortality, weekend admission status, disposition of patient, admission type, insurance type, length of hospital stay, comorbidities, and major diagnostic and procedure categories.

if PRCCS1 in (225) then PR1C=26;

if PRCCS1 in (226) then PR1C=27;

if PRCCS1 in (227) then PR1C=28;

if PRCCS1 in (228) then PR1C=29;

if PRCCS1 in (229) then PR1C=30;

if PRCCS1 in (230) then PR1C=31;

if PRCCS1 in (231) then PR1C=32;

if PRCCS1 not in (1:231) then PR1C=33; /\*no procedure\*/

**Data S2. Sample R code for performing imputation using rando draw, hotdeck, joint MI, conditional MI**

Required R packages: Amelia, HotDeckImputation, mice, mi

1. Random Draw

**sample**(complete.data$race, sum(as.numeric(is.na(data$race))))

# data is the entire dataset.

# complete.data$race is the subset that has no missing race data.

# sum(as.numeric(is.na(data$race))) is to calculate the number of subjects #that has missing race data

#’**sample’**is a function that can be used to randomly draw race data from the complete dataset to impute missing race.

Detailed information can be found here:

<https://stat.ethz.ch/R-manual/R-devel/library/base/html/sample.html>

2. Hotdeck:

**impute.NN\_HD**(data, distance = "man", weights = "range", attributes = "sim")

# impute.NN\_HD is a function in the HotDeckImputation package for performing hot deck imputation.

# distance: Distance type to use when searching for the nearest neighbor (e.g., “man’’, Manhattan distance)

#weights: Weights by which the variables should be scaled.

# attributes: Determines how attributes should be handled. Currently only "sim", meaning donor and recipient pools are disjoint, is implemented.

Detailed information can be found here: <https://www.rdocumentation.org/packages/HotDeckImputation/versions/1.1.0/topics/impute.NN_HD>

3. Joint Multiple Imputation:

**amelia**(data, m=5 , noms = nomlist, ords = odslist)

#amerlia is the function for performing joint MI.

#m: the number of imputed datasets to create.

#noms: a vector of numbers or names indicating columns in the data that are nominal variables.

#ords: a vector of numbers or names indicating columns in the data that are ordinal variables.

Detailed information can be found here:

<https://www.rdocumentation.org/packages/Amelia/versions/1.7.4/topics/amelia>?

4. Conditional Multiple Imputation (mice):

**mice**(data, m = 5, predictorMatrix)

#mice is the function for performing multivariate imputations by chained equations.

# data: A data frame or a matrix containing the incomplete data. Missing values are coded as NA.

#m: the number of imputed datasets to create.

# predictorMatrix: A matrix containing 0/1 data specifying the set of predictors to be used for each target column.

#Variable types must be specified before implementing mice. For example, race is nominal: factor(race); income is ordinal: orderd(income).

Detailed information can be found here:

<https://www.rdocumentation.org/packages/mice/versions/2.25/topics/mice>

5. Conditional Multiple Imputation (mi):

**mi**(data, n.chains = 5, include)

#mi is the function for performing multivariate imputations by chained equations in a Bayesian framework.

#data: a numeric matrix, containing a list of missing variables. ‘data’ must be converted to a missing\_data.frame first (see online Details)

#Variable types must be specified. For example, race is nominal: factor(race); income is ordinal: orderd(income).

# n.chains: the number of imputed datasets to create.

#include: A vector of indicators that decide whether or not (Yes/No) to include a specific variable in an imputation process.

Detailed information can be found here:

<https://cran.r-project.org/web/packages/mi/mi.pdf>

**Data S3. Evaluation of post-imputation performance**

Figure S1. Root mean square difference (RMSD) of coefficient estimates for linear regression for length of stay.

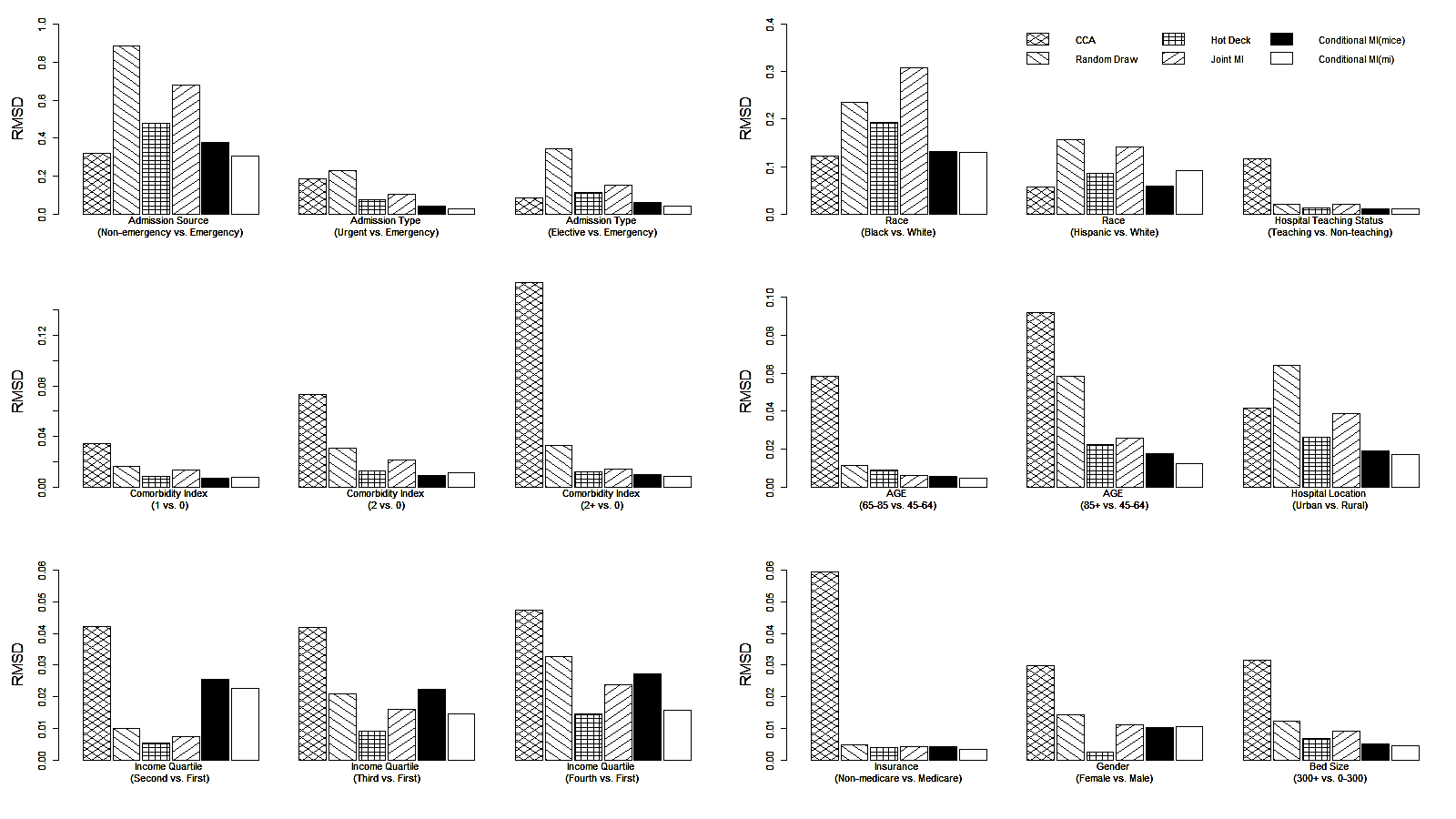


Figure S2. Root mean square difference (RMSD) of coefficient estimates for logistic regression for any complications.

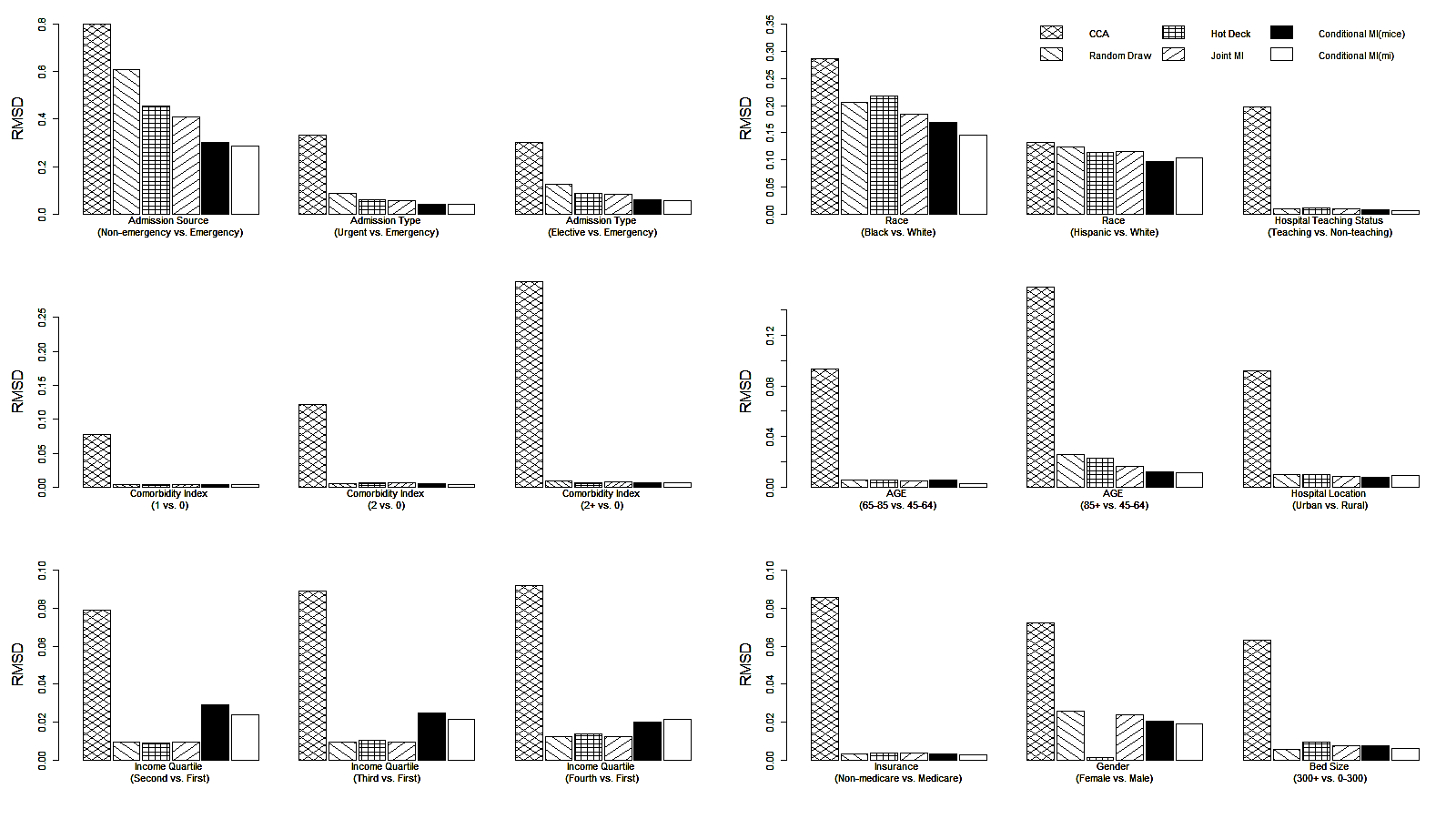
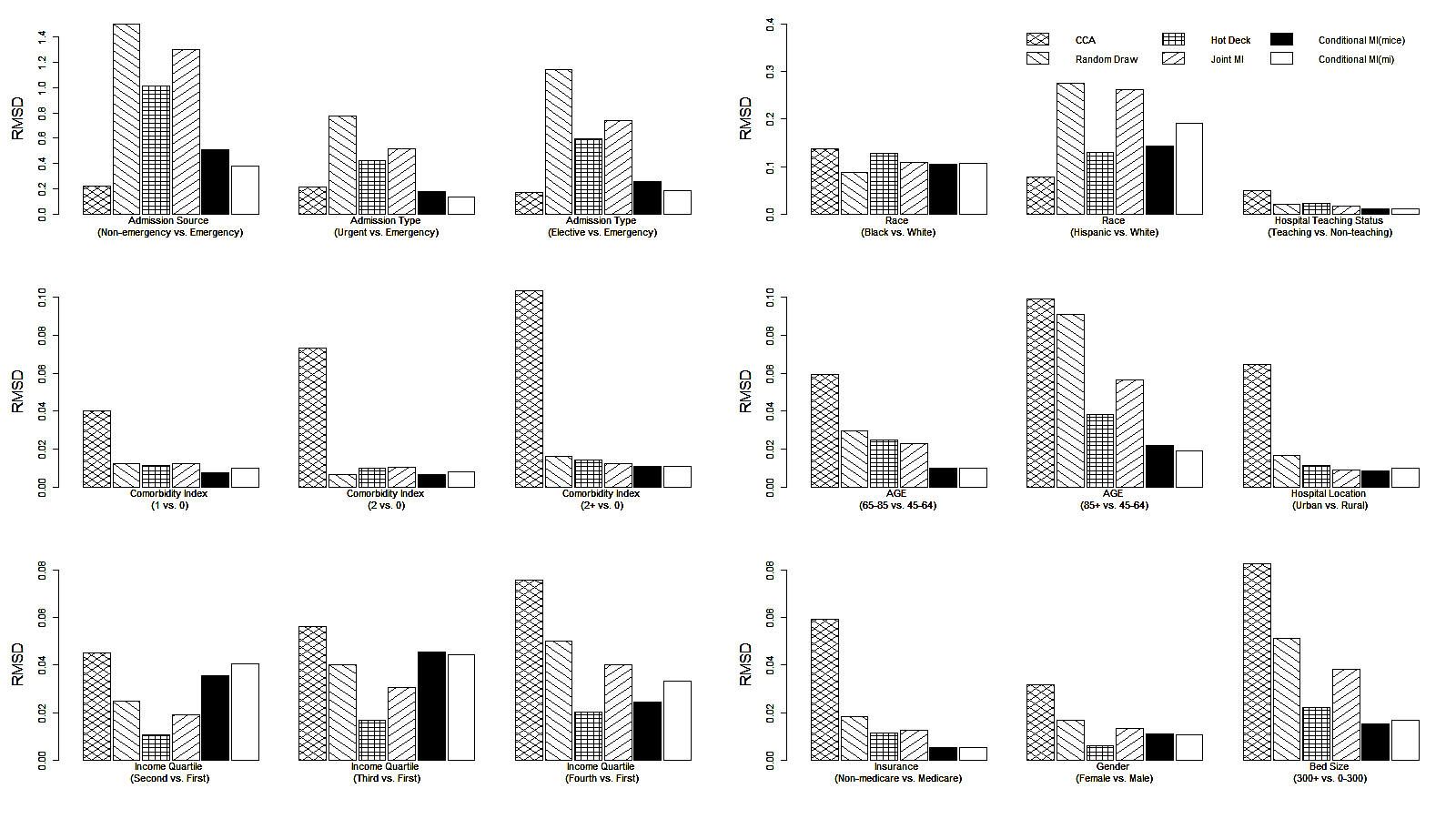
****

Figure S3. Root mean square difference (RMSD) of coefficient estimates for multinomial logistic regression for hospital TKA volume.



**Data S4. Real data analysis**

We conducted three regression analyses using real data from 2005 SID-CO to study racial disparities in total knee arthroplasty (TKA). We chose three outcomes of interest in TKA racial disparities research: length of hospital stay (a continuous outcome), any surgical complications (a binary outcome), and utilization of high TKA volume hospitals (an ordinal outcome: low [<200], medium [200-400], high annual TKA volume [400+]). Linear regression, logistic regression, and multinomial logistic regression were conducted for length of hospital stay, complications, and TKA volume, respectively. In addition to race, regression model covariates included age, gender, comorbidity index, median household income, admission type, admission source, insurance type, hospital bed size, teaching status and location.

Missing data on several variables (race, total charges, gender, income, admission source) were imputed using different statistical methods (random draw, hot deck, joint MI, conditional MI (mice), conditional MI (mi)). Sample R code for all imputation strategies can be found in Supplemental materials part A. Predictive variables used for the imputation included patient age, mortality, weekend admission status, disposition of patient, admission type, insurance type, length of hospital stay, comorbidities, and major diagnostic and procedure categories.

Regression models were run on complete cases (CCA) and imputed datasets from random draw (RD), hot deck (HD), joint MI, conditional MI (mice), and conditional MI (MI). Results including coefficient estimates, standard errors and p-values can be found in Tables 1-3. Since results from conditional MI (mice) and conditional MI (mi) were very similar, we only presented those from mi.

Table S1. Regression results for length of stay

Table S2. Regression results for any complications

Table S3. Regression results for hospital volume

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1** |  | CCA | | | RD | | | HD | | | Joint MI | | | Conditional MI | | |
| Variable |  | Est | SE | pval | Est | SE | pval | Est | SE | pval | Est | SE | pval | Est | SE | pval |
| Age | 45-64 (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 65-85 | 0.176 | 0.086 | 0.040 | 0.218 | 0.066 | 0.001 | 0.200 | 0.066 | 0.002 | 0.196 | 0.065 | 0.003 | 0.202 | 0.066 | 0.002 |
| 85+ | 0.208 | 0.156 | 0.185 | 0.227 | 0.131 | 0.084 | 0.206 | 0.132 | 0.118 | 0.205 | 0.135 | 0.129 | 0.213 | 0.132 | 0.107 |
| Admission  Source | Emergency (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Non-Emergency | -0.492 | 0.621 | 0.428 | -0.131 | 0.379 | 0.730 | -0.501 | 0.488 | 0.304 | -0.347 | 0.427 | 0.416 | -0.519 | 0.485 | 0.285 |
| Race | White (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Black | 0.590 | 0.170 | 0.001 | 0.410 | 0.131 | 0.002 | 0.429 | 0.139 | 0.002 | 0.222 | 0.120 | 0.064 | 0.428 | 0.174 | 0.014 |
| Hispanic | 0.083 | 0.096 | 0.388 | 0.128 | 0.075 | 0.087 | 0.072 | 0.091 | 0.427 | 0.021 | 0.082 | 0.796 | 0.065 | 0.078 | 0.409 |
| Gender | Male (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | 0.017 | 0.055 | 0.760 | 0.063 | 0.048 | 0.190 | 0.060 | 0.048 | 0.215 | 0.053 | 0.048 | 0.274 | 0.062 | 0.048 | 0.194 |
| Income  Quartile | 1st quartile (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2nd quartile | -0.097 | 0.062 | 0.117 | -0.042 | 0.057 | 0.458 | -0.042 | 0.057 | 0.462 | -0.059 | 0.059 | 0.315 | -0.037 | 0.052 | 0.481 |
| 3rd quartile | -0.133 | 0.071 | 0.061 | -0.174 | 0.059 | 0.003 | -0.155 | 0.059 | 0.009 | -0.179 | 0.064 | 0.005 | -0.140 | 0.055 | 0.012 |
| 4th quartile | -0.208 | 0.085 | 0.015 | -0.191 | 0.062 | 0.002 | -0.166 | 0.065 | 0.011 | -0.181 | 0.063 | 0.004 | -0.110 | 0.056 | 0.048 |
| Insurance | Medicare (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Non-Medicare | -0.047 | 0.095 | 0.621 | -0.004 | 0.072 | 0.951 | -0.029 | 0.070 | 0.681 | -0.031 | 0.072 | 0.668 | -0.036 | 0.070 | 0.605 |
| Comorbidity  Index | 0 (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | 0.253 | 0.052 | <.0001 | 0.248 | 0.043 | <.0001 | 0.242 | 0.043 | <.0001 | 0.251 | 0.044 | <.0001 | 0.245 | 0.044 | <.0001 |
| 2 | 0.879 | 0.174 | <.0001 | 0.826 | 0.138 | <.0001 | 0.832 | 0.137 | <.0001 | 0.852 | 0.141 | <.0001 | 0.841 | 0.139 | <.0001 |
| 2+ | 0.447 | 0.239 | 0.062 | 0.420 | 0.174 | 0.015 | 0.550 | 0.239 | 0.022 | 0.521 | 0.255 | 0.042 | 0.526 | 0.244 | 0.031 |
| Bed Size | 0-300 (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 300+ | -0.160 | 0.172 | 0.355 | -0.108 | 0.143 | 0.450 | -0.096 | 0.143 | 0.503 | -0.093 | 0.147 | 0.529 | -0.102 | 0.144 | 0.481 |
| Hospital  Teaching  Status | Non-Teaching (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Teaching | 0.055 | 0.152 | 0.717 | 0.082 | 0.120 | 0.497 | 0.075 | 0.122 | 0.542 | 0.060 | 0.125 | 0.633 | 0.081 | 0.124 | 0.516 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2** |  | CCA | | | RD | | | HD | | | Joint MI | | | Conditional MI | | |
| Variable |  | Est | SE | pval | Est | SE | pval | Est | SE | pval | Est | SE | pval | Est | SE | pval |
| Age | 45-64 (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 65-85 | 0.334 | 0.226 | 0.139 | 0.440 | 0.159 | 0.006 | 0.410 | 0.160 | 0.011 | 0.423 | 0.167 | 0.011 | 0.417 | 0.159 | 0.009 |
| 85+ | 0.698 | 0.493 | 0.157 | 0.847 | 0.382 | 0.027 | 0.804 | 0.377 | 0.033 | 0.792 | 0.404 | 0.050 | 0.827 | 0.373 | 0.027 |
| Admission  Source | Emergency (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Non-Emergency | -0.925 | 1.266 | 0.465 | -0.949 | 0.605 | 0.117 | -0.889 | 1.163 | 0.445 | -0.850 | 1.191 | 0.476 | -0.886 | 1.161 | 0.445 |
| Race | White (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Black | 0.237 | 0.414 | 0.567 | 0.235 | 0.393 | 0.551 | 0.138 | 0.328 | 0.674 | -0.071 | 0.366 | 0.847 | 0.332 | 0.397 | 0.406 |
| Hispanic | 0.162 | 0.312 | 0.605 | 0.188 | 0.261 | 0.473 | 0.051 | 0.263 | 0.845 | -0.032 | 0.254 | 0.901 | 0.148 | 0.266 | 0.578 |
| Gender | Male (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | -0.127 | 0.146 | 0.382 | -0.118 | 0.123 | 0.339 | -0.122 | 0.123 | 0.322 | -0.110 | 0.130 | 0.396 | -0.102 | 0.123 | 0.408 |
| Income  Quartile | 1st quartile (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2nd quartile | 0.019 | 0.148 | 0.901 | 0.041 | 0.139 | 0.769 | 0.091 | 0.142 | 0.519 | 0.064 | 0.144 | 0.656 | 0.217 | 0.161 | 0.178 |
| 3rd quartile | -0.079 | 0.193 | 0.681 | -0.134 | 0.174 | 0.440 | -0.096 | 0.175 | 0.584 | -0.143 | 0.185 | 0.440 | 0.150 | 0.190 | 0.429 |
| 4th quartile | -0.163 | 0.252 | 0.517 | -0.006 | 0.224 | 0.980 | -0.015 | 0.239 | 0.949 | -0.016 | 0.227 | 0.945 | 0.110 | 0.231 | 0.634 |
| Insurance | Medicare (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Non-Medicare | -0.182 | 0.193 | 0.346 | -0.166 | 0.146 | 0.254 | -0.182 | 0.144 | 0.207 | -0.155 | 0.154 | 0.314 | -0.175 | 0.143 | 0.221 |
| Comorbidity  Index | 0 (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | 0.441 | 0.181 | 0.015 | 0.503 | 0.142 | 0.000 | 0.495 | 0.143 | 0.001 | 0.496 | 0.147 | 0.001 | 0.503 | 0.143 | 0.000 |
| 2 | 0.876 | 0.241 | 0.000 | 0.771 | 0.217 | 0.000 | 0.769 | 0.219 | 0.000 | 0.783 | 0.222 | 0.000 | 0.769 | 0.218 | 0.000 |
| 2+ | 0.510 | 0.507 | 0.314 | 0.955 | 0.451 | 0.034 | 0.958 | 0.451 | 0.034 | 0.927 | 0.457 | 0.043 | 0.972 | 0.453 | 0.032 |
| Bed Size | 0-300 (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 300+ | -0.095 | 0.306 | 0.756 | -0.176 | 0.248 | 0.477 | -0.201 | 0.248 | 0.417 | -0.201 | 0.252 | 0.425 | -0.187 | 0.248 | 0.451 |
| Hospital  Teaching  Status | Non-Teaching (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Teaching | 0.357 | 0.222 | 0.108 | 0.571 | 0.178 | 0.001 | 0.583 | 0.181 | 0.001 | 0.489 | 0.199 | 0.015 | 0.571 | 0.192 | 0.003 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 3** |  | CCA | | | RD | | | HD | | | Joint MI | | | Conditional MI | | |
| Variable |  | Est | SE | pval | Est | SE | pval | Est | SE | pval | Est | SE | pval | Est | SE | pval |
| Age | 45-64 (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 65-85 | 0.228 | 0.260 | 0.380 | 0.419 | 0.311 | 0.177 | 0.408 | 0.310 | 0.188 | 0.422 | 0.307 | 0.170 | 0.403 | 0.307 | 0.189 |
| 85+ | -0.058 | 0.348 | 0.868 | 0.093 | 0.384 | 0.809 | 0.102 | 0.397 | 0.797 | 0.101 | 0.387 | 0.794 | 0.085 | 0.386 | 0.825 |
| Admission  Source | Emergency (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Non-Emergency | 0.952 | 0.707 | 0.178 | 1.644 | 0.786 | 0.036 | 0.616 | 0.540 | 0.255 | 0.791 | 0.624 | 0.211 | 0.696 | 0.518 | 0.179 |
| Race | White (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Black | 0.258 | 0.569 | 0.650 | 0.347 | 0.378 | 0.359 | 0.347 | 0.415 | 0.403 | 0.216 | 0.354 | 0.542 | 0.360 | 0.408 | 0.378 |
| Hispanic | -0.547 | 0.365 | 0.134 | -0.411 | 0.250 | 0.100 | -0.487 | 0.281 | 0.083 | -0.305 | 0.293 | 0.298 | -0.430 | 0.298 | 0.149 |
| Gender | Male (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | 0.042 | 0.058 | 0.466 | 0.049 | 0.051 | 0.338 | 0.040 | 0.049 | 0.418 | 0.049 | 0.051 | 0.340 | 0.039 | 0.049 | 0.430 |
| Income  Quartile | 1st quartile (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2nd quartile | 0.225 | 0.211 | 0.286 | 0.146 | 0.191 | 0.445 | 0.148 | 0.186 | 0.428 | 0.146 | 0.192 | 0.447 | 0.295 | 0.208 | 0.156 |
| 3rd quartile | 0.554 | 0.419 | 0.186 | 0.308 | 0.306 | 0.313 | 0.287 | 0.301 | 0.342 | 0.329 | 0.310 | 0.288 | 0.314 | 0.332 | 0.345 |
| 4th quartile | 0.284 | 0.438 | 0.517 | 0.161 | 0.376 | 0.668 | 0.191 | 0.376 | 0.612 | 0.191 | 0.383 | 0.618 | 0.270 | 0.402 | 0.502 |
| Insurance | Medicare (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Non-Medicare | 0.246 | 0.317 | 0.439 | 0.567 | 0.420 | 0.177 | 0.537 | 0.419 | 0.201 | 0.545 | 0.402 | 0.175 | 0.541 | 0.417 | 0.195 |
| Comorbidity  Index | 0 (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | 0.070 | 0.062 | 0.254 | -0.002 | 0.054 | 0.968 | 0.009 | 0.052 | 0.857 | 0.013 | 0.053 | 0.807 | 0.009 | 0.053 | 0.865 |
| 2 | 0.045 | 0.212 | 0.832 | 0.120 | 0.171 | 0.484 | 0.129 | 0.169 | 0.446 | 0.128 | 0.177 | 0.470 | 0.126 | 0.173 | 0.469 |
| 2+ | 0.140 | 0.265 | 0.598 | 0.330 | 0.261 | 0.206 | 0.357 | 0.250 | 0.153 | 0.309 | 0.263 | 0.241 | 0.322 | 0.260 | 0.215 |
| Bed Size | 0-300 (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 300+ | 0.838 | 0.836 | 0.316 | 0.780 | 0.770 | 0.311 | 0.764 | 0.770 | 0.321 | 0.803 | 0.769 | 0.296 | 0.768 | 0.771 | 0.319 |
| Hospital  Teaching  Status | Non-Teaching (reference) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Teaching | 0.529 | 0.560 | 0.345 | 0.490 | 0.506 | 0.333 | 0.472 | 0.509 | 0.354 | 0.485 | 0.506 | 0.338 | 0.507 | 0.513 | 0.323 |