Predictors of Medicare Provider Overcharge

Darya Gahramanova | Priyanka Thakur | Shaolin Pu | Anh Nguyen

1 EXECUTIVE SUMMARY

Medicare is a health insurance program created by the United States' federal government. It is for people who are 65 years old or older, some younger individuals with disabilities and people with End-Stage Renal Disease (Medicare.gov, n.d.). According to the Medicare website, Medicare consists of three parts:

- Medicare Part A (Hospital Insurance) which covers inpatient hospital stays, skilled nursing facility care, hospice care and various home health care.
- Medicare Part B (Medical Insurance) which covers qualified doctor's services, outpatient care, preventive services and medical supplies.
- Medicare Part D (Prescription Drug Coverage) covers the cost of prescription drugs.

Aside from the original Medicare, there are also Medicare Advantage (MA) plans which include Medicare benefits and are provided through private insurers (Medicare.gov, n.d.).

Over the years, there have been reports on how medical procedures provided to Medicare for the MA plans are charged at a higher price for various reasons such as fraud, exaggeration of sickness by medical providers or administrative errors.

The Centers for Medicare and Medicaid Services (CMS) published Medicare provider utilization and payment data from 2012 to 2017 on their website. This series of data outlines the utilization and payments for services, procedures and description drugs provided to the beneficiaries of Medicare by specific physicians, inpatient/outpatient hospitals and other suppliers. These datasets have information on utilization, payment and submitted charges. The payment information includes the allowed amount and Medicare payment. The said information is organized by the National Provider Identifier (NPI), Healthcare Common Procedure Coding System (HCPCS) code, and place of service, including multiple countries and states.

In this project, we looked at data provided by the CMS from 2012 to 2017 to determine if there are any trends of overcharging in the original Medicare, and what are the key predictors that can be looked at to predict overcharge, whether it be the state where the practice took place, certain health conditions or the type of procedure performed. Types of procedures performed per provider as well as the provider type turned out to be one of the most significant determinants of overcharge. Other factors including state, year, and beneficiary chronic conditions were also associated with the overcharge amount.

2 PROBLEM DEFINITION AND SIGNIFICANCE

According to the National Health Expenditures 2017 Highlights document provided by CMS, the United States healthcare spending was $3.5 trillion (an increase of 3.9 percent), or $10,739 per person (cms.gov).

However, in 2016 CMS found that $41.1 billion of Medicare payments were improper, which accounts for 11 percent of Medicare payments, as stated in the Medicare Fee-For-Service 2016 Improper Payments Report (cms.gov). This is not necessarily considered fraud and is rather an estimation of payments that did not meet Medicare coverage, coding, and billing rules. While CMS has been working on decreasing improper payments through Healthcare Fraud Prevention Partnership, Medical Review Strategies, Provider Education, and Policy Clarifications, there are factors that affect individual providers' tendency to overcharge. These include the size of the provider, area of specialty, state regulations, and which HCPCS code was assigned for a service. Kaiser Health News reported that it was known by officials for years that some Medicare plans overcharge the government by overstating their patients' sickness or by charging Medicare for treating serious medical conditions that are unprovable (Schulte & Weber, 2019).

Medicare is susceptible to mismanagement and improper payments, considering how complex and big the program is. If left unattended, the government might have tens of billions of improper payments per fiscal year. This can be detrimental to the overall funding for the program and cause misuse of tax money. Therefore, our project aims to pinpoint potential factors associated with the provider overcharge in the U.S. and propose corresponding recommendations to help the patients and Medicare.

3 PRIOR LITERATURE

Title	Data	Methods	Findings
The impact of payor/provider type on health care use and expenditures among the frail elderly. Experton, B., Li, Z., Branch, L. G., Ozminkowski, R. J., & Mellon-Lacey, D. M. (1997). The impact of payor/provider type on health care use and expenditures among the frail elderly. American journal of public health, 87(2), 210–216. https://doi.org/10.2105/ajph.87.2.210	- 700 patients aged 65+ of Sharp Home HealthCare, San Diego - 220 patients have Medicare fee-for-service, 310 with Medicare HMO and 170 with both Medicare and Medicaid - Analyses only have 450 people who satisfied all criteria	- Bivariate and multivariate analyses: estimates relationships between payor source and health care expenditures and usage - Ordinary least squares regression analyses to estimate relationships between payor type and log of (1) total health care expenditure, (2) total number of inpatient hospital days and (3) total number of subsequent home health visits - Logistic regression analyses: estimated relationships between payor status and having one or more inpatient admissions/ER visits, the use of home health care and the use of long-term skilled nursing care - Variance inflation factor tests: determine if heteroskedasticity existed in the form of nonconstant variance among the regression residuals associated with the three payor categories	- Total expenditures for HMO beneficiaries were 46.8% lower than those who were dually enrolled in Medicare and Medicaid. - Total expenditures for fee-for-service beneficiaries were, on average, 52.2% lower than those who were dually enrolled. - Poor or fair self-assessed health status results in 68.0% higher in total health expenditures than better status. - Participants who have 4+ prescribed medications had 254.8% more total expenditures than those with no prescribed medications. - Participants with 2+ inpatient hospitalizations during the previous year had 74.1% higher total expenditures than those who had fewer admissions.
Patient-Sharing Networks of Physicians and Health Care Utilization and Spending Among Medicare Beneficiaries Landon, B. E., Keating, N. L., Onnela, J. P., Zaslavsky, A. M., Christakis, N. A., & O'Malley, A. J. (2018). Patient-Sharing Networks of Physicians and Health Care Utilization and Spending Among Medicare Beneficiaries. JAMA internal medicine, 178(1), 66–73. https://doi.org/10.1001/jamainternmed.2017.5034	- Medicare administrative data (2006-2010) - An average of 3 761 223 Medicare beneficiaries per year; 40 241 physicians in 51 hospitals referral regions	- 2-sided t tests or $\chi 2$ tests at the 5% level: evaluate bivariate differences - Multivariable linear regression: examine the relationship between selected network measures versus outcomes of interested variables both individually and together - Hierarchical generalized linear model and fitting marginal regression models with adjusted variances: explore specifications for accounting for communities's clustering of observations - Generalized linear model: analyze each outcomes - Linear regression model and independent working correlation model: analyze log of cost and compare with results from ordinary least squares estimator - Poisson: analyze total community hospital admissions and days in hospital Logistic regression: analyze quality measures	- Major increases in spending were in inpatient services and hospital outpatient services - Patients of physicians with more connections to other physicians and more shared care outside of the community had higher total spending - Inpatient care had slightly lower spending for patients of physicians who belongs to communities with higher proportions of primary care physicians - Patients who were in the care of physicians linked to more physicians had more hospital admissions and days, emergency visits, specialist visits and primary care visits - Patients with physicians who were in networks that had more primary care physicians had more primary care visits
Associations Among Hospital Capacity, Utilization, and Mortality of U.S. Medicare Beneficiaries, Controlling for Sociodemographic Factors Fisher, E. S., Wennberg, J. E., Stukel, T. A., Skinner, J. S., Sharp, S. M., Freeman, J. L., & Gittelsohn, A. M. (2000). Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors. Health services research, 34(6), 1351–1362.	- 20% sample of Medicare enrollees aged 65+ who were eligible for Medicare Part A during 1989 or 1990 - Final study population: 5.53 million Medicare enrollees	- Poisson regression (weighting by the number of subjects in each age-sex-race and zip code stratum): estimate relationship between in-area bed supply and hospital discharge rate - Logistic regression: estimate relationship between three measures of hospital resources and Medicare mortality	- Medicare beneficiaries in areas with higher per capita numbers of hospitals beds were associated with higher hospitalization rates, mainly for nonsurgical causes of admission - Enrollees in areas with more beds were more likely to be admitted at least once during the analyzed year. - Apparent differences in utilization across income and racial groups - Each income and racial group saw an association between increased capacity and utilization
Healthcare costs and utilization for Medicare beneficiaries with Alzheimer's Zhao, Y., Kuo, T. C., Weir, S., Kramer, M. S., & Ash, A. S. (2008). Healthcare	- Data obtained from the MEDSTAT MarketScan Medicare Supplemental and Coordination of Benefits Database from 2003-2004.	- Comparisons of total health care cost using the mean, computed t-tests of the differences in means, standard deviation, cost dispersion measured by coefficient of variation and cost distribution by place of service	- The AD group had a higher overall illness burden, more comorbid medical conditions. It also had higher but less variable costs - Significant excess utilization were seen in AD group for inpatient services,

costs and utilization for Medicare beneficiaries with Alzheimer's. BMC health services research, 8, 108. https://doi.org/10.1186/1472-6963-8-108	- Medicare beneficiaries aged 65+ with employer-sponsored supplemental insurance. - 627 775 participants who had 12 months of enrollment in 2003 and at least 1 month in 2004. - Did not include cost and utilization data for skilled nursing facilities and nursing home	- Examination of health care utilization using percentages of users, mean cost per member per year, and the number of visits per member per year - Regression: estimate the independent effect on overall cost and utilization of Alzheimer's Disease (AD) - Weighted Least Squares regression: examine AD's marginal contribution to costs and health care utilization for 2004 and measure the extent of AD's contributions to excess healthcare cost or utilization - Logistic regression: examine AD's marginal contribution to ER visits and inpatient admissions	pharmacy, ER visits, and home health care - Patients who had AD had significantly higher healthcare costs and utilization than demographically-matched Medicare beneficiaries.
The Impact of Hospital-Acquired Conditions on Medicare Program Payments Kandilov, A., Coomer, N. M., & Dalton, K. The impact of hospital-acquired conditions on Medicare program payments. Medicare Medicaid Res Rev. 2014; 4 (4): E1-23.	- Data taken from Medicare public use claims files from 2009-2010, including the Medicare Provider Analysis and Review, the Medicare Standard Analytic Files and the Enrollment Database - 6 Hospital-Acquired Conditions (HAC) with the highest volume and surgical site infection with the highest volume (total of 24499 patients)	- Descriptive analyses: analyze the unadjusted differences between HACs and the matched non-HAC in per-episode payments - Multivariate model: examine total Medicare episode payments - Log-linear regression with provider fixed effects: estimate incremental payment effect of each HAC	- For all selected HACs, total Medicare episode payments higher significantly higher than the matched comparison non-HAC episodes - Payments for HAC episodes in almost all subsets of episode payments were almost always significantly higher except for outpatient payments - Medicare paid $146 million per year more across the chosen HAC care episodes compared to the matched non-HACs
Variation in the Ratio of Physician Charges to Medicare Payments by Specialty and Region Bai, G., & Anderson, G. F. (2017). Variation in the ratio of physician charges to Medicare payments by specialty and region. Jama, 317(3), 315-318.	- Data taken from the Medicare physician utilization and payment data in 2014 from the Centers for Medicare & Medicaid Services - 429 273 physicians across 54 medical specialties	- Calculation of overall median physician excess charge for each state, each medical specialty: identify systematic patterns of variation - Examination of the top 2.5% physicians in high excess charge for specialty and geographic distribution	- Physician charge-to-Medicare payment ratio varied across specialties, with anesthesiology as the highest median and general practice as the lowest - The ratio also varied across states, with state median ranging between 2.0 (IQR, 1.5-3.1 for Michigan) and 3.8 (IQR, 2.9-6.5 for Wisconsin) - Out of 10 730 physicians with high excess charges, 55% were anesthesiologists, 3% were in general practice, family practice or internal medicine. 32% of these physicians practiced in 10 out of 306 hospital referral regions in NY, TX, WI, GA, NJ, CA, and NC
Predicting Overcharge in Common Healthcare Procedure with High Payment Variation Sen, Sagnika and Deokar, Amit V., "Predicting Overcharge in Common Healthcare Procedure with High Payment Variation" (2018). Proceedings of the 2018 Pre-ICIS SIGDSA Symposium. 19. https://aisel.aisnet.org/sigdsa2018/19	- Data taken from the 2014 Provider Utilization and Payment Data: Physician and Other Supplier for Medicare Part B fee-for-service beneficiaries - Amongst four services with abnormally high degree of variations in payment per service, HCPCS 66984 - removal of cataract was chosen for having the highest service volume and total payment by Medicare	- Binomial classification: find key influential predictors for high procedures overcharge ratio - Machine learning methods including linear Naive Bayes classifier, non-linear k-nearest-neighbors, Elastic Net, Gradient Boosting, Random Forest. Neural Networks and Support Vector Machines were used with both linear and nonlinear kernels.	- 4 out of the top 5 predictors were related to practice characteristics (overcharge ratio, reimbursement ratio, total payment to the practice by Medicare and the number of unique beneficiaries). Procedure total Medicare payment amount is related to the procedure itself - Provider's gender, affiliation (with an organization or working as an individual), racial makeover of the patient base were not significant differentiating features. - Correlation between practice and procedure level overcharge was positive and significant. - Large providers overcharge more - Provider type is an important feature in predicting overcharge

4 DATA SOURCE AND PREPARATION

Medicare Provider Utilization and Payment Data used in this project was collected by Centers for Medicare and Medicaid Services (CMS) and includes information on services and procedures provided to Medicare beneficiaries by physicians and other healthcare professionals. This data contains information on utilization, payment, and submitted charges aggregated in several ways. Two types of this data were selected for the analysis: one by Healthcare Common Procedure Coding System (HCPCS) code, and another by National Provider Identifier (NPI) - a total of 12 datasets for every year from 2012 to 2017.

The first type of dataset is *Medicare Provider Utilization and Payment Data: Physician and Other Supplier PUF*. This dataset, aggregated by procedures, includes information of the number of beneficiaries, payment and submitted charges, as well as location by national provider identifier (NPI). There are 26 columns and about 9.85M rows.

The second type of dataset is *Medicare Physician and Other Supplier National Provider Identifier (NPI) Aggregate Report,* which work as a supplement to the first type of datasets and are grouped by NPI. The second type of dataset includes information of beneficiary demographic and health characteristics that provided the percentage number of patients' age, race, Medicare and Medicaid entitlement, health conditions and risk scores. There are 71 columns and 1.09M rows.

The reason for including both datasets is the smaller level of aggregation for the procedure dataset, but additional information on beneficiaries in the provider dataset. Although the most accurate way to analyze this data would be to merge the two datasets to have information on both individual procedures and beneficiaries, it would be impossible to do using selected tools (R software). Therefore' we had to approach the cleaning and merging process differently.

First, we loaded the datasets and changed their columns to match, as they were changing with time. For instance, in procedures, average_Medicare_standard_amt was added only in 2014, rural/urban was added in 2017, and the standard deviation variables were removed after 2013. Column names in 2012 and 2013 were different from others and were changed. Then, for every year, a table for procedures and a table for providers was loaded and the following transformations were made.

For providers, the following columns were selected: NPI, entity code, provider state, provider country, provider type, Medicare participation indicator, number of HCPCS, total services, total unique beneficiaries, total submitted charge amount, total Medicare allowed amount, total Medicare payment amount, number of medical (non-ASP) HCPCS, total medical (non-ASP) services, total medical unique beneficiaries, total med submitted charge amount, total med Medicare allowed amount, total medical (non-ASP) Medicare payment amount, beneficiary average age, beneficiary age less 65 count, beneficiary age 65-74 count, beneficiary age 75-84 count, beneficiary age greater than 84 count, beneficiary female count, beneficiary male count, beneficiary race white count, beneficiary race black count, beneficiary race Asian Pacific count, beneficiary race Hispanic count, beneficiary race American Indian or Alaska Native count, beneficiary race other count, beneficiary nondual (qualified to receive Medicare only benefits) count, beneficiary dual (qualified to receive Medicare and Medicaid benefits) count, beneficiary with atrial fibrillation %, beneficiary with Alzheimer's %, beneficiary with Asthma %, beneficiary with cancer %, beneficiary with heart failure %, beneficiary with kidney disease %, beneficiary with chronic obstructive pulmonary disease %, beneficiary with depression %, beneficiary with diabetes %, beneficiary with hyperlipidemia %, beneficiary with hypertension %, beneficiary with ischemic heart disease %, beneficiary with osteoporosis %, beneficiary with rheumatoid arthritis/osteoarthritis %, beneficiary with schizophrenia %, beneficiary with stroke %, beneficiary average risk score. We omitted some variables that we were not going to use: total drug services / unique beneficiaries / charges / payments, because they can be calculated as Total - Medical (non-ASP); med_suppress_indicator because of too many missing variable; provider's information like name, credentials (because of too many different values), gender (we assume that gender does not affect overcharge), and address fields (we are using the state level).

Furthermore, non-US and non-individual providers (entity code = organizations) were filtered out from the dataset. We noticed that some small regions of the US (AA, AE, AP, AS, GU, MP, PR, VI) only had a few values, so we filtered them out too. All dollar fields, incorrectly parsed as strings, were cleaned and converted to numeric variables. Charges equal to zero were treated as missing variables, and all rows with missing charges were dropped. Overcharge was

calculated as total submitted charge amount - Medicare payment amount. After that, country, entity code, charge and payment variables were dropped. Categorical/character variables were converted to lowercase, empty values were replaced with NAs, and converted to factor variables for the analysis.

Similar steps were performed on procedures dataset, which does not have any beneficiary information, but has HCPCS codes, descriptions, and number of services provided, number of unique beneficiaries, and average charges and payments per each code per each doctor. We did not use HCPCS descriptions or charges / payments. The dataset was cleaned and made smaller using the steps above. However, it was still unmanageable for the analysis because of thousands of different codes. Therefore, we aggregated the codes into groups based on the information retrieved from www.aapc.com. This was done manually using regular expressions because some codes contain letters as well as numbers. Resulting groups were more generalized than HCPCS descriptions, but still carried much more information than just the total number of codes per provider. We pivoted these groups from rows to columns, with every value representing the number of services provided by a particular doctor. At first, we performed rather general aggregations, but some of these groups ended up having too many values compared to others, so we split these groups into more specific categories. On the other side, those procedure groups that only had a few values, were aggregated into either more generic groups or “miscellaneous services”. Now that this dataset was aggregated by providers, we could merge it with the providers dataset on NPI. A respective year (e.g. 2012) was added to this merged dataset. Then the above steps were performed for every year including 2017. Finally, all years were merged.

When we checked the provider types, we noticed that some types were the same but treated as different factors because of minor differences in the way they were written, so we renamed them. Similar to the procedure groupings, we aggregated provider types whenever possible if they had too few values, and dropped several rows that had provider types that were unknown or only occurred once. Our data was reduced from more than 9,000,000 rows per one year to about 100,000 rows and 100 columns in total.

5 VARIABLE CHOICE

In the previous section we discussed which variables we excluded from the analysis right away. As we were analyzing data further, we excluded several more variables from our models: NPI (we only needed it for merging), Medicare participation indicator (all providers ended up participating in Medicare after data cleaning), medical services (this was already captured by procedure groups and was redundant), and beneficiary age groups (we preferred a single average age column to avoid too many variables).

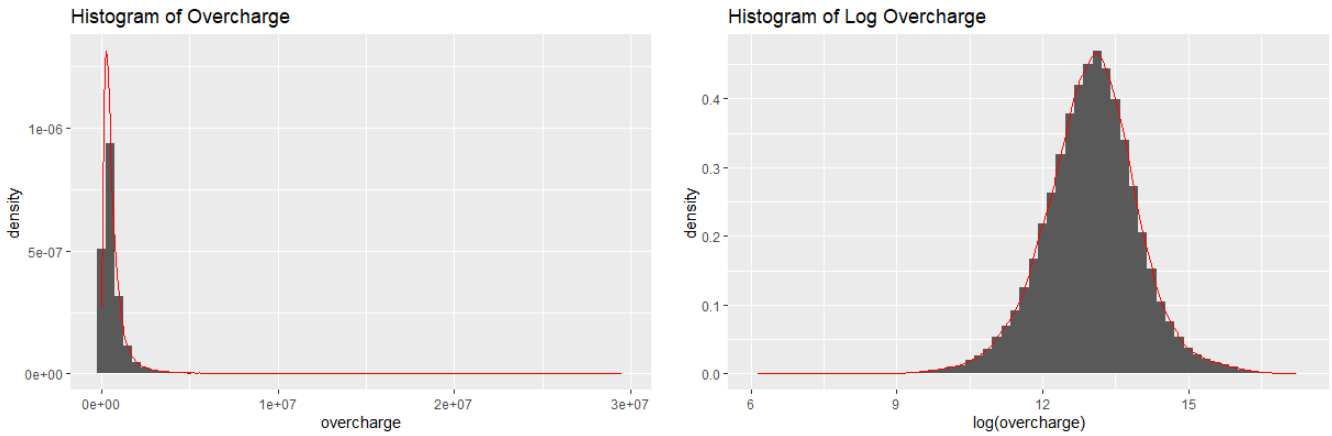
Originally chronic conditions were represented as percentages, and beneficiary demographic characteristics and procedures as counts. In order to make everything consistent, beneficiary demographic characteristics and procedures were converted to proportions to the total number of beneficiaries and the total number of procedures respectively. Totals were included as separate variables. For every group of percentage variables, one was dropped to avoid perfect collinearity. Year was converted to a factor variable to serve as a fixed effect. Final variables are summarized below:

Dependent Variable: Overcharge		
Predictor	Effect	Rationale
Year	?	Since there was improper Medicare Advantage payment for 2016, we want to look at different years to see if there is any patterns in our data
Nppes provider state	+/-	Regional policy affects hospital utilization and costs
Provider type	+/-	Provider type affects the cost and utilization since some procedures are more expensive or more demanded than others
Total unique beneficiaries	+/-	More unique beneficiaries result in more utilization; physicians with higher utilization might locate in populated cities and/ or larger in size compared to other providers, which could result in higher price
Number of HCPCS	+	The more procedures the doctor performs, the higher his prices and utilization are likely to be
Beneficiary average age	+	The greater the age, the higher the cost since older people are more prone to medical issues due to declining health
Beneficiary female %	+	Females tend to seek more medical help than males
Beneficiary race white %	+/-	Utilization is higher for white beneficiaries because cost of care is less of a barrier for them. Cost should be the same for all beneficiaries
Beneficiary race black %		
Beneficiary race Asian Pacific %		
Beneficiary race Hispanic %		
Beneficiary race American Indian or Alaska Native %		
Beneficiary dual (Medicare & Medicaid) %	-	People who are eligible for both Medicare and Medicaid may have lower prices than those who only have one or none.
Beneficiary chronic conditions (beneficiary with atrial fibrillation %, beneficiary with Alzheimer's %, beneficiary with Asthma %, beneficiary with cancer %, beneficiary with heart failure %, beneficiary with kidney disease %, beneficiary with chronic obstructive pulmonary disease %, beneficiary with depression %, beneficiary with diabetes %, beneficiary with hyperlipidemia %, beneficiary with hypertension %, beneficiary with ischemic heart disease %, beneficiary with osteoporosis %, beneficiary with rheumatoid arthritis/osteoarthritis %, beneficiary with schizophrenia %, beneficiary with stroke %)	+/-	Chronic conditions increase the cost for beneficiaries. Some diseases are more common or expensive than others, while some diseases have less specialists, hence less access to treatment and subsequently, more costs (from prolonged wait and/or multiple treatments & complications)
Beneficiary average risk score	+	The higher the risk score, the higher the cost since risk score reflects a person's predicted health costs compared to an average beneficiary
% of procedures per aggregation performed by the provider (Ambulatory services, Anesthesia, Category iii, Drugs administered other than oral method, Chemotherapy drugs, Vaccines or toxoids, Psychiatry services and procedures, Dialysis services and procedures, Gastroenterology procedures, Ophthalmology services and procedures, Special otorhinolaryngologic services and procedures, Cardiovascular procedures, Non invasive vascular	+/-	Some procedures are more expensive than others. Prior literature pointed out that anesthesiology and surgical procedures are amongst the most expensive ones

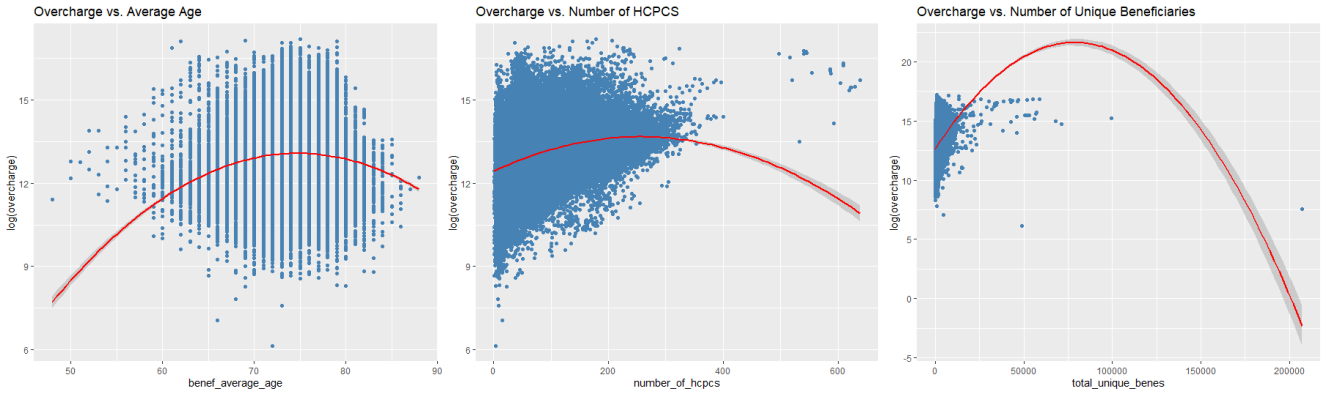
diagnostic studies, Pulmonary procedures, Allergy and clinical immunology procedures, Endocrinology services, Neurology and neuromuscular procedures, Highly complex drug or biologic agent administration, Therapeutic procedures, Physical medicine and rehabilitation evaluations, Moderate conscious sedation, Other medicine services and procedures, Office or other outpatient services, Hospital observation services, Hospital inpatient services, Emergency department services, Critical care services, Nursing facility services, Home services, Prolonged services, Other care evaluation and management services, Organ or disease-oriented panels, Drug assay, Urinalysis procedures, Chemistry procedures, Hematology and coagulation procedures, Immunology procedures, Microbiology procedures, Cytopathology procedures, Surgical pathology procedures, Diagnostic radiology procedures, Diagnostic ultrasound procedures, Radiologic guidance, Breast and mammography, Bone or joint studies, Radiation oncology treatment, Nuclear medicine procedures, Fine needle aspiration biopsy procedures, Surgical procedures on the auditory system, Surgical procedures on the cardiovascular system, Surgical procedures on the digestive system, Surgical procedures on the eye and ocular adnexa, Surgical procedures on genital system, Surgical procedures on the hemic and lymphatic systems, Surgical procedures on the integumentary system, Surgical procedures on the musculoskeletal system, Surgical procedures on the nervous system, Surgical procedures on the respiratory system, Surgical procedures on the urinary system, Other surgical procedures, Screening examinations and disease management training, Miscellaneous diagnostic and therapeutic services, Initial services for Medicare enrollment, Gross and microscopic examinations prostate biopsy, Counseling screening and prevention services, Miscellaneous services)		

6 DESCRIPTIVE ANALYSIS AND DATA VISUALIZATIONS

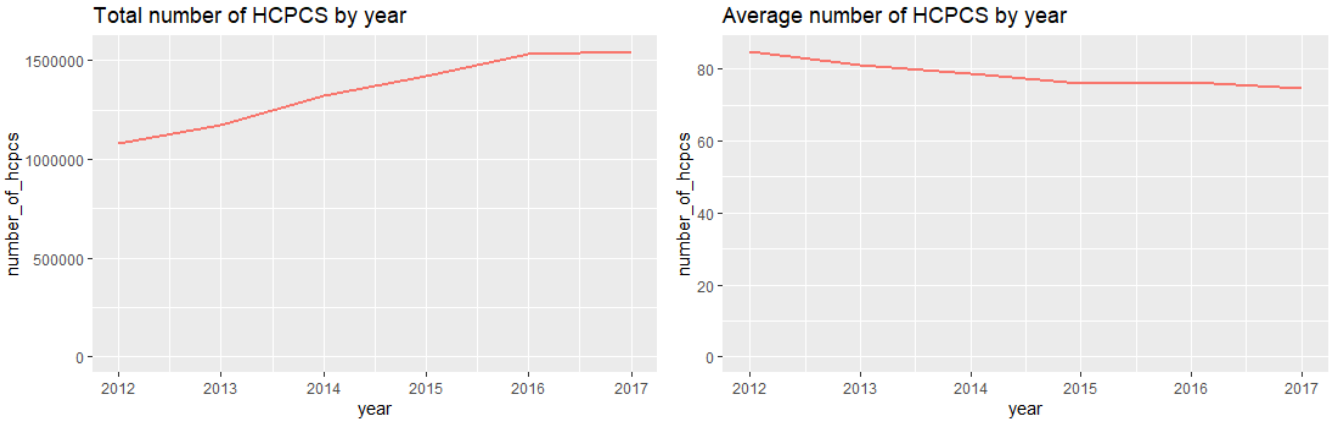
To know whether the numerical variables normally distributed, histograms are applied on each numerical variable. Those histograms of count variables are highly right skewed, and all count variables have a large amount lower value count and are not normally distributed. For example, the histogram of overcharge shows a right skewed distribution. After applying log transformation on overcharge, the histogram of log(overcharge) shows log(overcharge)'s distribution much closer to the normal distribution.



Considering average age could have a positive impact on overcharge, older patients more likely have serious diseases and require more complex treatments. To explore whether average age has impacts on overcharge, a plot of log(overcharge) and average was created. The plot doesn't clearly show any pattern or linear relationship between average age and log(overage):

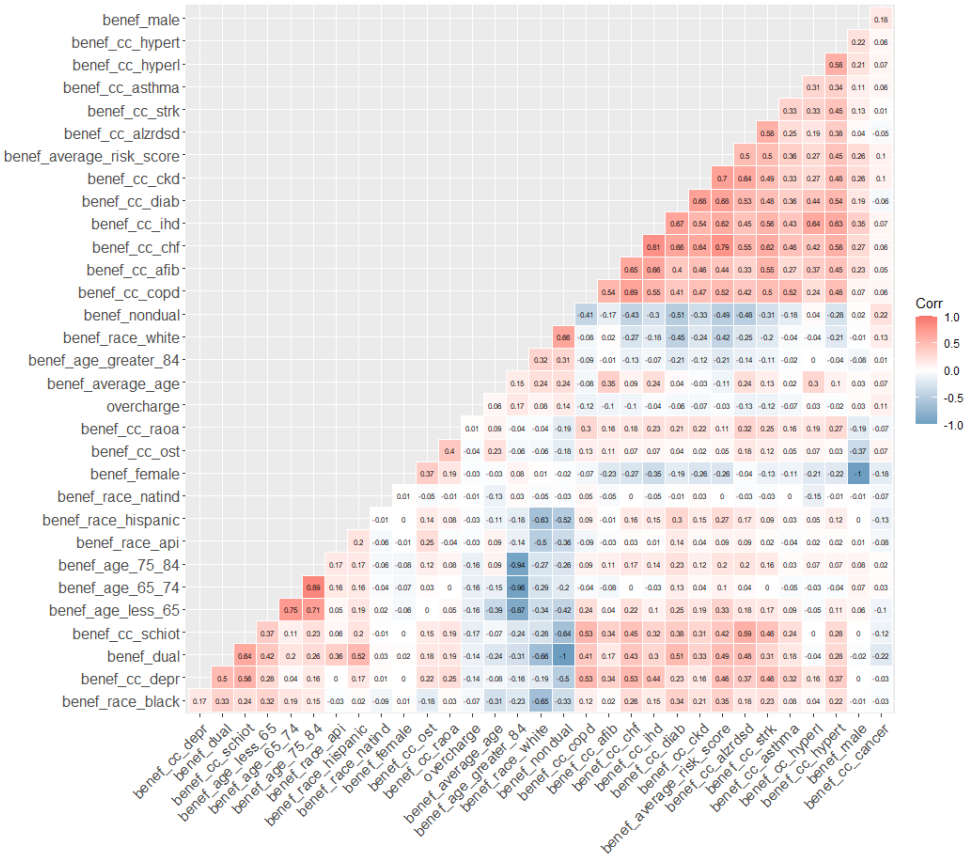


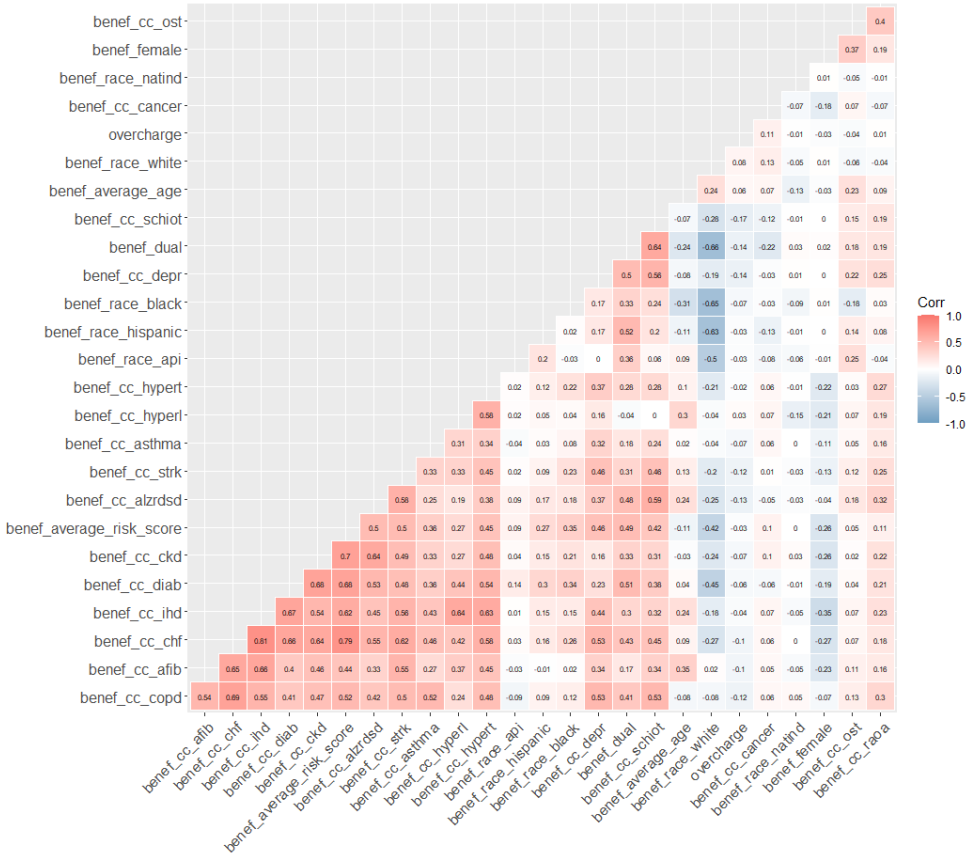
The chart below demonstrates how the number of HCPCS changed with time. Interestingly, the total number of procedures stopped increasing after 2016, while the average number of procedures was decreasing with time. This goes in line with CMS's report indicating that stricter regulations reduced overcharge after 2016.



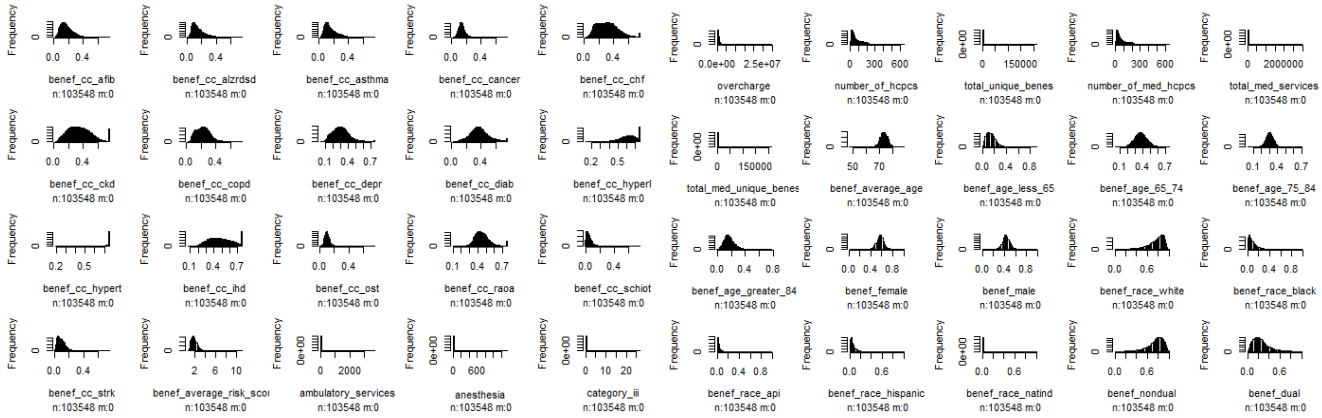
Furthermore, we are interested in correlations between predictor variables. These variables are already converted to proportions as described earlier. First correlation plot includes all percentage variables for beneficiary demographic characteristics, even though we expect that mutually exclusive variables will be correlated. The plot proves our guess, as variables indicating male/female gender or dual/nondual eligibility are almost perfectly correlated.

Second correlation plot variables that we want to include in our models. There are still correlations between some variables, like average risk score with chronic conditions. However, we are not omitting these variables, because some chronic conditions might not necessarily result in a higher risk score, and the risk score does not capture chronic conditions only. Same goes to the correlations between dual eligibility with chronic conditions and between conditions themselves. While they might be related, they do not necessarily replace each other: dual eligibility may relate to conditions but varies by states and other factors. Thus, we did not remove all those variables, but were cautious with the analysis and estimated variance inflation factors for our models.





Finally, we looked at the distributions of all numeric variables to assess if further transformations were needed. We can see that most of the variables are right-skewed and need a log transformation just like Overcharge. Although there are a few exceptions that are normal or left-skewed, we decided to keep similar variables on the same scale and apply log transformations to everything but beneficiary average age.




Frequency
0e+00
0 20000
_administered_other_than_o
n:103548 m:0
0 600
chemotherapy_drugs
n:103548 m:0
0 1500
vaccines_or_toxoids
n:103548 m:0
0 1500
sychiatry_services_and_proc
n:103548 m:0
0 200 500
dialysis_services_and_proce
n:103548 m:0
0
0 150 300
hospital_observation_servi
n:103548 m:0
0 1000
hospital_inpatient_service
n:103548 m:0
0 200
emergency_department_ser
n:103548 m:0
0 100 200
critical_care_services
n:103548 m:0
0 2000
nursing_facility_services
n:103548 m:0
0 5 15
gastroenterology_proceduthalmology_services_and_pr
n:103548 m:0
0 1000
rhinolaryngologic_services_
n:103548 m:0
0 1000
cardiovascular_procedure
n:103548 m:0
0 3000
invasive_vascular_diagnosti
n:103548 m:0
0 400 1000
home_services
n:103548 m:0
0 400 800
prolonged_services
n:103548 m:0
0 50 150
re_evaluation_and_manage
n:103548 m:0
0 100
rgan_or_disease_oriented_p
n:103548 m:0
0 300
drug_assay
n:103548 m:0
0 100
0 300 600
pulmonary_procedures
n:103548 m:0
0 1000
y_and_clinical_immunology_p
n:103548 m:0
0 40 80
endocrinology_services
n:103548 m:0
0 1500
ology_and_neuromuscular_p
n:103548 m:0
0 100 200
nplex_drug_or_biologic_agen
n:103548 m:0
0 200 400
urinalysis_procedures
n:103548 m:0
0 1000
chemistry_procedures
n:103548 m:0
0 300 600
atology_and_coagulation_pr
n:103548 m:0
0 400
immunology_procedures
n:103548 m:0
0 600
microbiology_procedures
n:103548 m:0
0 2000
therapeutic_proceduresal_medicine_and_rehabilitation
n:103548 m:0
0 1000
moderate_conscious_sedar
n:103548 m:0
0 10 20
_medicine_services_and_pr
n:103548 m:0
0 20 50
ffice_or_other_outpatient_se
n:103548 m:0
0 400 1000
cytopathology_procedure
n:103548 m:0
0 2000
surgical_pathology_proced
n:103548 m:0
0 6000
diagnostic_radiology_proced
n:103548 m:0
0 600
diagnostic_ultrasound_proce
n:103548 m:0
0 400
0 100 250
radiologic_guidance
n:103548 m:0
0 400 800
breast_and_mammograph
n:103548 m:0
0 300
bone_or_joint_studies
n:103548 m:0
0 150 300
radiation_oncology_treatm
n:103548 m:0
0 200 500
nuclear_medicine_procedu
n:103548 m:0
0 60 120
needle_aspiration_biopsy_pr
n:103548 m:0
0 30 60
microscopic_examinations_
n:103548 m:0
0 300 600
ling_screening_and_preventi
n:103548 m:0
0 200000
miscellaneous_services
n:103548 m:0
0 1500
al_procedures_on_the_audit
n:103548 m:0
0 1000
rocedures_on_the_cardiova
n:103548 m:0
0 40
al_procedures_on_the_diges
n:103548 m:0
0 200 500
rocedures_on_the_eye_and_
n:103548 m:0
0 5 10
gical_procedures_on_genital
n:103548 m:0
0 2 4 6
edures_on_the_hemic_and_
n:103548 m:0
0 1000
procedures_on_the_integum
n:103548 m:0
0 100 250
procedures_on_the_musculos
n:103548 m:0
0 200 400
al_procedures_on_the_nervo
n:103548 m:0
0 200
d_procedures_on_the_respira
n:103548 m:0
0 100
cal_procedures_on_the_urina
n:103548 m:0
0.0 1.5 3.0
other_surgical_procedure
n:103548 m:0
0 600
aminations_and_disease_mæ
n:103548 m:0
0 400 800
ous_diagnostic_and_therapa
n:103548 m:0
0 20 40
l_services_for_medicare_e
n:103548 m:0

7 DATA MODELING AND QUALITY CHECKS

Three models have been tested for this analysis. All of them are linear multilevel models using either fixed or mixed effects of the year, state, and provider type. Numeric variables include beneficiary average age, number of beneficiaries by gender, race, Medicaid eligibility, chronic conditions, risk score, and, except for the first model, procedures by their groups. All numeric variables, including overcharge, were log-transformed, and 0.01 is added to those variables that contain zeros to avoid the log-of-zero error. We used AIC scores to compare the models, as two of them include random effects and are built using LMER package, which does not provide R^2. In order to verify if the model with the lowest AIC is indeed the best, we also ran the same models on the 75% training set and validated on the remaining 25% by computing root mean squared errors (RMSE). The training and validation split was stratified by the provider type to be representative of the whole dataset, as the number of providers in each type varied significantly.

In the first model, HCPCS groups were not included in order to increase predictive powers of other variables and see whether the results are consistent with other models. Provider type is assumed to have random intercept, while years and states are fixed effects.

Second model includes year, state, and provider type as fixed effects. Additionally, HCPCS groups were included. Again, all numeric variables are log-transformed.

Third model has the same variables as the second, but it treats the provider type as a random intercept variable instead of fixed effects. This was done in case if having too many fixed effects would cause collinearity and overfitting, as well as eating up too many degrees of freedom. This model also uses weights to correct for heteroscedasticity - a weighted least squares approach. These weights were estimated by creating a linear model from Model 2's residuals vs. fitted and squaring the inverted value of the resulting fitted values vector. This is done to reduce the residual variance as the fitted values of overcharge increase.

For the full model output, please refer to the Appendix. The models were compared using AIC scores as well as validation on test data. The following table summarizes the scores:

	Model 1	Model 2	Model 3
Variable summary	Year, state, provider type, number of beneficiaries, number of HCPCS, beneficiary age, gender, race, Medicaid eligibility, chronic condition, risk score	Year, state, provider type, number of beneficiaries, number of HCPCS, beneficiary age, gender, race, Medicaid eligibility, chronic condition, risk score, HCPCS groups	Year, state, provider type, number of beneficiaries, number of HCPCS, beneficiary age, gender, race, Medicaid eligibility, chronic condition, risk score, HCPCS groups
AIC	217693.2	192771.2	193118.6
RMSE	877635.1	769823.8	766854.2

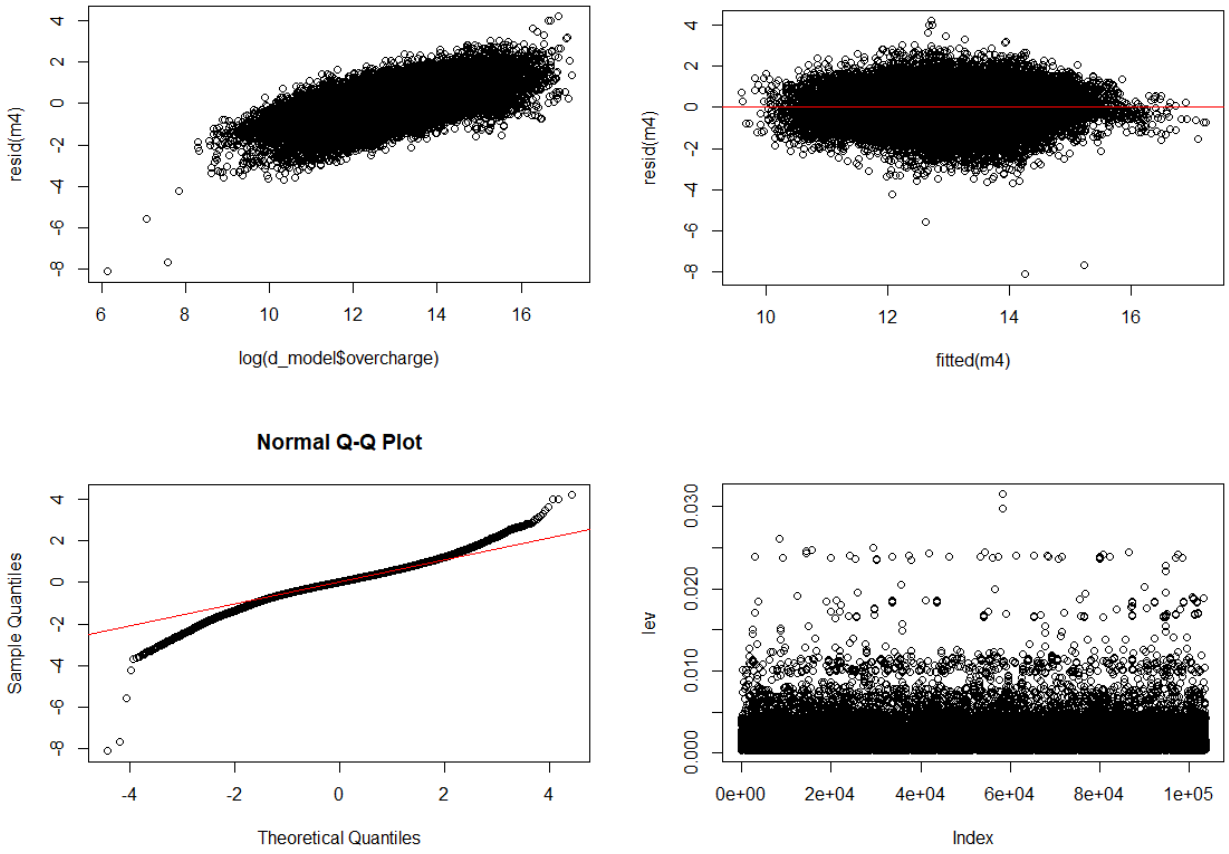
According to the results, Model 2 performed best based on the AIC score. However, Model 3 did better in terms of the root mean square error computed on the 25% validation set, because it corrects for heteroscedasticity. That means that the dataset indeed suffered from the extreme values, and Model 3 helps with that. It is thus selected as the best model because it is the most robust yet accurate model in this case.

To know whether the model is better than the random guesswork, two baseline RMSEs were estimated. One is on the predicted results of Model 3 compared to shuffled values of overcharge - and resulted in an RMSE of 1152428. Another one is the values of overcharge compared to the mean overcharge - with an RMSE of 1057204. In both cases, our models help predicting overcharge better than the baseline models.

Selected model has to pass the following assumptions to be considered reliable: linearity, normality, homoscedasticity, independence, and no multicollinearity. The top left plot is residuals vs. overcharge, intending to identify the existence of a linear relationship between Y value and its residuals. The relationship looks linear, so the

linearity assumption holds. The top right plot is residuals vs. fitted values, which shows that there is no autocorrelation pattern, meaning that the independence assumption holds. There is still some extent of heteroscedasticity although it was already lessened by Model 3's weights. This extent is still mostly acceptable, and we can assume that the model passes the homoscedasticity test. We do notice a few outliers on all plots, including the bottom-left Q-Q plot. With the exception of these outliers, the model looks mostly normal and can pass the normality test. The last plot is the model's leverage that proves that some extreme instances exist in the dataset which make it hard to have a very accurate model.

The last test is on multicollinearity between the predictor variables. Although we removed the variables that can be calculated from other variables from the models, some extent of relationship between the variables is still expected because a) there are many predictors in the dataset and b) each row represents each doctor's beneficiaries and procedures. The VIF test indicates that most of the variables are within acceptable limit, with the exception of percent of Medicare beneficiaries qualified to receive Medicare and Medicaid (5.6), percent of beneficiaries with Alzheimer's (5.0), heart failure (6.4), chronic kidney disease (7.1), and ischemic heart disease (6.2). We have already seen these correlation patterns in the previous section, so it is recommended to separate those variables or use other predictors in a way that they are unrelated in future studies.



8 RESULTS AND RECOMMENDATIONS

We used hierarchical linear modeling techniques to identify key predictors for Medicare excess charges by providers and found several important insights. According to both the first model, which puts more emphasis on variables other than procedures, and the selected third one, states that have the lowest overcharge are Wyoming (WY), Idaho (ID), South Dakota (SD), and Hawaii (HI). The ones with the highest overcharge are Alaska (AK) and Wisconsin (WI), which surpassed even such expensive states as California and New York. It appears that these two states do not have any state balance-billing protection (www.commonwealthfund.org), which controls the billing process to prevent providers from "accidentally" overcharging. Although some of the low-overcharge states do not have billing protections either, they have a lower cost index in general. Hawaii is an exception in this case, and the low overcharge there is associated with strong state laws (www.civilbeat.org). Our recommendation is thus to force stronger regulations in states with high overcharge.

According to both models, the year with the lowest overcharge turned out to be 2012, and with the highest – 2016. Prices are generally expected to go up with time, which explains the lowest amount in 2012. The reason why overcharge was lower in 2017 than in 2016 is related to the regulations put in place by CMS: Healthcare Fraud Prevention Partnership, Medical Review Strategies, Provider Education, and Policy Clarifications. This proves that interventions work and are necessary to control the overcharge rate.

The size of the provider also plays a role in how much they overcharge. As the total number of procedures increases by 1%, overcharge increases by 0.4%, and as the total number of beneficiaries that provider has increases by 1%, overcharge increases by 0.6%. This means that the more patients the provider has and the more procedures they perform, the more likely they are going to overcharge. Our recommendation for Medicare would be to look closer into the largest providers in order to identify if potential restrictions could result in savings.

Age was not proven to be as important to predict overcharge as we expected. As the patient's age increases by 1, overcharge increases by 1%. This is probably because the beneficiary age range was similar for most of the providers, as people have to be of a certain age to be eligible for Medicare.

What we found surprising is that the providers with more female patients had lower overcharge that those with more male patients. We expected females to have more health problems as thus overcharge because they have a longer life expectancy as well as female-specific problems like breast cancer. Further investigation is needed to determine if providers are more likely to overcharge male beneficiaries because of the policies or conditions that occur more often in males.

Regarding beneficiaries' races/ethnicities, providers turned out to overcharge white patients most: 18% for white patients versus close to 0% for other races. We cannot derive conclusions on whether this is because white patients have worse health because there were almost 10 times more white beneficiaries than those of any other race. This is much more than the actual difference in population by those races. This means that Medicare benefits should be promoted more among minorities.

Providers with more patients who are eligible for dual Medicare and Medicaid benefits are overcharging more than those with more Medicare-only patients by 10%. Dual eligibility is available to beneficiaries of certain age with certain conditions. Thus, this overcharge might be related do worse health of those patients. However, it might also be related to the providers trying to get more from those patients as they do not bring much profit. According to www.kff.org, there were widespread reports of dual eligibles being overcharged before. We suggest that stricter regulations are put on providers regarding dual beneficiaries.

Some chronic conditions resulted a positive effect on overcharge, while others, interestingly, negative. Beneficiaries with hypertension and hyperlipidemia had the highest overcharge, and those with heart failure and atrial fibrillation – the lowest. Because people with chronic conditions rely on medications for their entire lives, we would recommend that a higher emphasis is placed on covering medication costs for those with the highest overcharge.

As we expected, average risk score significantly affects the overcharge amount. As the average risk score increases by 1, overcharge increases by 25%. Of course, this is because the severity of sickness would increase total charges.

However, this also means that the overcharge on top of what Medicare covers makes it even harder to pay out of pocket for the patients with severe conditions, making them more likely to die rather than receive treatment. This might be very hard to change, but the government might investigate how to financially help people with severe conditions in future, and possibly restrict the overcharge on those conditions.

Provider type was another important determinant of the overcharge. The amount was the highest for surgical oncologists, neurosurgeons, and occupational therapists, and the lowest for optometrists, maxillofacial surgeons, and ophthalmologists. This has two explanations: first, doctors treating larger organs tend to overcharge because they need very expensive tools for their treatments. Second, they overcharge because of the importance and popularity of their field. As a result, these doctors tend to make more money, which incentivizes future doctors to go to those fields. This would cause too many doctors to be specialized in one field, and very few in other, more rare fields. Therefore, when a patient does seek care for those less-demanded conditions, it would be hard to find a doctor. We would recommend to look further into the doctors with the lowest overcharge, and either raise the salaries for those who are underpaid, or incentivize medical students to choose those specializations in some other way to balance the number of doctors in each specialty.

Finally, procedure types performed by the providers were also significantly associated with overcharges. Surgical procedures on the eye and ocular adnexa, radiation oncology treatment, and chemotherapy drugs were some of the highest-overcharged procedures. On the other hand, vaccines or toxoids, urinalysis procedures, and microbiology procedures were the lowest. This is definitely related to the complexity of the procedure, that is, surgeries and cancer were the most overcharged due to being the most complex procedures. Just like before, we would recommend to look into regulations on pricing these procedures, because not only the cost of the procedure itself is high, but also the overcharge on top of it, making these treatments unavailable for those who cannot afford it. Because a large portion of these costs for the providers is the cost of necessary tools, the prices on these tools may be reconsidered. Finally, we would recommend looking into procedures on a more granular level in future studies, as they are the most important determinants of the overcharge and are very likely to be exploited. For instance, one could include both HCPCS codes and the more general groups that we created to identify if some providers assign more expensive codes to cheaper procedures that belong in the same group in order to charge more.

To address the issues identified during this study, further exploration is needed to find better ways to rescale data columns and group provider type to reduce the amount that may help improve our model and enable us to understand how the prominent providers differ in terms of their overcharge behavior.

9 REFERENCES

Bai, G., & Anderson, G. F. (2017). *Variation in the ratio of physician charges to Medicare payments by specialty and region*. Jama, 317(3), 315-318.

Centers for Medicare & Medicaid Services. (n.d.). *National Health Expenditures 2017 Highlights*. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/downloads/highlights.pdf

Experton, B., Li, Z., Branch, L. G., Ozminkowski, R. J., & Mellon-Lacey, D. M. (1997). *The impact of payor/provider type on health care use and expenditures among the frail elderly*. American journal of public health, 87(2), 210–216. https://doi.org/10.2105/ajph.87.2.210

Fisher, E. S., Wennberg, J. E., Stukel, T. A., Skinner, J. S., Sharp, S. M., Freeman, J. L., & Gittelsohn, A. M. (2000). *Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors*. Health services research, 34(6), 1351–1362.

Kandilov, A., Coomer, N. M., & Dalton, K. *The impact of hospital-acquired conditions on Medicare program payments*. Medicare Medicaid Res Rev. 2014; 4 (4): E1-23.

Landon, B. E., Keating, N. L., Onnela, J. P., Zaslavsky, A. M., Christakis, N. A., & O'Malley, A. J. (2018). *Patient-Sharing Networks of Physicians and Health Care Utilization and Spending Among Medicare Beneficiaries*. JAMA internal medicine, 178(1), 66–73. https://doi.org/10.1001/jamainternmed.2017.5034

Medicare.gov. (n.d.). *What's Medicare?* https://www.Medicare.gov/what-Medicare-covers/your-Medicare-coverage-choices/whats-Medicare

Medicare.gov. (n.d.). *Medicare Advantage Plans*. https://www.Medicare.gov/sign-up-change-plans/types-of-Medicare-health-plans/Medicare-advantage-plans

Schulte, F., & Weber, L. (2019, 7 16). *Insurers Running Medicare Advantage Plans Overbill Taxpayers By Billions As Feds Struggle To Stop It*. Kaiser Health News. https://khn.org/news/Medicare-advantage-overbills-taxpayers-by-billions-a-year-as-feds-struggle-to-stop-it/

Sen, Sagnika and Deokar, Amit V., *"Predicting Overcharge in Common Healthcare Procedure with High Payment Variation"* (2018). Proceedings of the 2018 Pre-ICIS SIGDSA Symposium. 19. https://aisel.aisnet.org/sigdsa2018/19

U.S. Government Accountability Office. (2017, July 19). *Medicare Advantage Program Integrity: CMS's Efforts to Ensure Proper Payments and Identify and REcover Improper Payments.* (Publication No. GAO-17-761T). Retrieved from https://www.gao.gov/assets/690/685934.pdf

Zhao, Y., Kuo, T. C., Weir, S., Kramer, M. S., & Ash, A. S. (2008). *Healthcare costs and utilization for Medicare beneficiaries with Alzheimer's*. BMC health services research, 8, 108. https://doi.org/10.1186/1472-6963-8-108

10 APPENDIX

```
library(plyr)
library(dplyr)
library(tidyr)
library(PerformanceAnalytics)
library(Hmisc)
library(lme4)
library(ggplot2)
library(gridExtra)
library(caret)
library(car)
library(plm)
library(robustlmm)
library(stargazer)
```

Read the data

```
procedure_12 <-
read.delim("Medicare_Provider_Util_Payment_PUF_CY2012/Medicare_Provider_Util_Payment_PUF_CY2012.txt")
procedure_13 <-
read.delim("Medicare_Provider_Util_Payment_PUF_CY2013/Medicare_Provider_Util_Payment_PUF_CY2013.txt")
procedure_14 <-
read.delim("Medicare_Provider_Util_Payment_PUF_CY2014/Medicare_Provider_Util_Payment_PUF_CY2014.txt")
procedure_15 <-
read.delim("Medicare_Provider_Util_Payment_PUF_CY2015/Medicare_Provider_Util_Payment_PUF_CY2015.txt")
procedure_16 <-
read.delim("Medicare_Provider_Util_Payment_PUF_CY2016/Medicare_Provider_Util_Payment_PUF_CY2016.txt")
procedure_17 <-
read.delim("Medicare_Provider_Util_Payment_PUF_CY2017/Medicare_Provider_Util_Payment_PUF_CY2017.txt")
provider_12 <-
read.delim("Medicare_Physician_and_Other_Supplier_NPI_Aggregate_CY2012/Medicare_Physician_and_Other_Supplier_NP
I_Aggregate_CY2012.txt")
provider_13 <-
read.delim("Medicare_Physician_and_Other_Supplier_NPI_Aggregate_CY2013/Medicare_Physician_and_Other_Supplier_NP
I_Aggregate_CY2013.txt")
provider_14 <-
read.delim("Medicare_Physician_and_Other_Supplier_NPI_Aggregate_CY2014/Medicare_Physician_and_Other_Supplier_NP
I_Aggregate_CY2014.txt")
provider_15 <-
read.delim("Medicare_Physician_and_Other_Supplier_NPI_Aggregate_CY2015/Medicare_Physician_and_Other_Supplier_NP
I_Aggregate_CY2015.txt")
provider_16 <-
read.delim("Medicare_Physician_and_Other_Supplier_NPI_Aggregate_CY2016/Medicare_Physician_and_Other_Supplier_NP
I_Aggregate_CY2016.txt")
provider_17 <-
read.delim("Medicare_Physician_and_Other_Supplier_NPI_Aggregate_CY2017/Medicare_Physician_and_Other_Supplier_NP
I_Aggregate_CY2017.txt")
```

Make column names lowercase

```
colnames(procedure_12) = tolower(make.names(colnames(procedure_12)))
colnames(procedure_13) = tolower(make.names(colnames(procedure_13)))
colnames(procedure_14) = tolower(make.names(colnames(procedure_14)))
colnames(procedure_15) = tolower(make.names(colnames(procedure_15)))
colnames(procedure_16) = tolower(make.names(colnames(procedure_16)))
colnames(procedure_17) = tolower(make.names(colnames(procedure_17)))
colnames(provider_12) = tolower(make.names(colnames(provider_12)))
colnames(provider_13) = tolower(make.names(colnames(provider_13)))
colnames(provider_14) = tolower(make.names(colnames(provider_14)))
colnames(provider_15) = tolower(make.names(colnames(provider_15)))
colnames(provider_16) = tolower(make.names(colnames(provider_16)))
colnames(provider_17) = tolower(make.names(colnames(provider_17)))
```

Make all columns the same

```
provider_12 <- cbind(provider_12[1:67], provider_12[70]) # dropping 2 empty columns

colnames(provider_12) <- c(colnames(provider_16[1:21]), colnames(provider_16[23:29]),
colnames(provider_16[31:37]), colnames(provider_16[39:71]))
colnames(provider_13) <- colnames(provider_12)
colnames(provider_14) <- colnames(provider_16)
colnames(provider_15) <- colnames(provider_16)
```

Transformations below are demonstrated on Providers 2012 and Procedures 2012 tables only, but are applied to

all tables 2012-2017

Providers

```
# Select relevant columns:
provs <- c("npi", "nppes_entity_code", "nppes_provider_state", "nppes_provider_country", "provider_type",
"medicare_participation_indicator", "number_of_hcpcs", "total_services", "total_unique_benes",
"total_submitted_chrg_amt", "total_medicare_allowed_amt", "total_medicare_payment_amt", "number_of_med_hcpcs",
"total_med_services", "total_med_unique_benes", "total_med_submitted_chrg_amt",
"total_med_medicare_allowed_amt", "total_med_medicare_payment_amt", "beneficiary_average_age",
"beneficiary_age_less_65_count", "beneficiary_age_65_74_count", "beneficiary_age_75_84_count",
"beneficiary_age_greater_84_count", "beneficiary_female_count", "beneficiary_male_count",
"beneficiary_race_white_count", "beneficiary_race_black_count", "beneficiary_race_api_count",
"beneficiary_race_hispanic_count", "beneficiary_race_natind_count", "beneficiary_race_other_count",
"beneficiary_nondual_count", "beneficiary_dual_count", "beneficiary_cc_afib_percent",
"beneficiary_cc_alzrdsd_percent", "beneficiary_cc_asthma_percent", "beneficiary_cc_cancer_percent",
"beneficiary_cc_chf_percent", "beneficiary_cc_ckd_percent", "beneficiary_cc_copd_percent",
"beneficiary_cc_depr_percent", "beneficiary_cc_diab_percent", "beneficiary_cc_hyperl_percent",
"beneficiary_cc_hypert_percent", "beneficiary_cc_ihd_percent", "beneficiary_cc_ost_percent",
"beneficiary_cc_raoa_percent", "beneficiary_cc_schiot_percent", "beneficiary_cc_strk_percent",
"beneficiary_average_risk_score")
prov_12 <- subset(provider_12, select=provs)

# Select US only
prov_12 <- subset(prov_12, nppes_provider_country=="US")
# Some regions (with only a few values - drop them
prov_12 %>% group_by(nppes_provider_state) %>% summarise(no_rows = length(nppes_provider_state))
'%notin%' <- Negate('%in%')
prov_12 <- prov_12[which(prov_12$nppes_provider_state %notin% c("AA", "AE", "AP", "AS", "GU", "MP", "PR", "VI",
"XX")), ]
# Select only individuals (not organizations)
prov_12 <- subset(prov_12, nppes_entity_code=="I")

# Numbers in providers are characters. Need to get rid of special characters and convert them to numbers
providers_num = c("number_of_hcpcs", "total_services", "total_unique_benes", "total_submitted_chrg_amt",
"total_medicare_allowed_amt", "total_medicare_payment_amt", "total_med_services", "total_med_unique_benes",
"total_med_submitted_chrg_amt", "total_med_medicare_allowed_amt", "total_med_medicare_payment_amt",
"beneficiary_age_less_65_count", "beneficiary_age_65_74_count", "beneficiary_age_75_84_count",
"beneficiary_age_greater_84_count", "beneficiary_female_count", "beneficiary_male_count",
"beneficiary_race_white_count", "beneficiary_race_black_count", "beneficiary_race_api_count",
"beneficiary_race_hispanic_count", "beneficiary_race_natind_count", "beneficiary_race_other_count",
"beneficiary_nondual_count", "beneficiary_dual_count")
prov_12[providers_num] <- lapply(prov_12[providers_num], function(x) as.numeric(gsub('[$,]', '', x)))

# Drop rows where charge = 0 or NA
prov_12$total_med_submitted_chrg_amt <- lapply(prov_12$total_med_submitted_chrg_amt, function(x) ifelse(x==0,
NA, x))
prov_12 <- prov_12[complete.cases(prov_12$total_med_submitted_chrg_amt),]

# Calculate Overcharge
prov_12$overcharge <- prov_12$total_submitted_chrg_amt - prov_12$total_medicare_payment_amt

# Drop no longer needed columns
prov_12 <- prov_12[ , !(names(prov_12) %in% c("nppes_provider_country", "nppes_entity_code",
"total_medicare_allowed_amt", "total_submitted_chrg_amt", "total_medicare_payment_amt"))]

providers_fac = c("nppes_provider_state", "provider_type", "medicare_participation_indicator")
# Lowercase all factor variables to make them the same
prov_12[providers_fac] <- lapply(prov_12[providers_fac], tolower)
# Empty strings in factor variables are not recognized as NAs. We need to get them manually
prov_12[providers_fac] <- lapply(prov_12[providers_fac], as.character)
prov_12[providers_fac] <- lapply(prov_12[providers_fac], function(x) ifelse(x!="", x, NA))
prov_12[providers_fac] <- lapply(prov_12[providers_fac], as.factor)
prov_12 <- droplevels(prov_12)
```

Procedures

```
# Select relevant columns:
procs <- c("npi", "nppes_entity_code", "nppes_provider_state", "nppes_provider_country", "provider_type",
"medicare_participation_indicator", "place_of_service", "hcpcs_code", "hcpcs_drug_indicator", "line_srvc_cnt",
"bene_unique_cnt", "average_medicare_allowed_amt", "average_submitted_chrg_amt",
"average_medicare_payment_amt")
proc_12 <- subset(procedure_12, select=procs)

# Select US only
proc_12 <- subset(proc_12, nppes_provider_country=="US")
# Some regions (with only a few values - drop them
proc_12 %>% group_by(nppes_provider_state) %>% summarise(no_rows = length(nppes_provider_state))
'%notin%' <- Negate('%in%')
```

```
proc_12 <- proc_12[which(proc_12$nppes_provider_state %notin% c("AA", "AE", "AP", "AS", "GU", "MP", "PR", "VI",
"XX")), ]
# Select only individuals (not organizations)
proc_12 <- subset(proc_12, nppes_entity_code=="I")

# Ged rid of special characters and convert them to numbers
procedures_num = c("average_medicare_allowed_amt", "average_submitted_chrg_amt",
"average_medicare_payment_amt")
proc_12[procedures_num] <- lapply(proc_12[procedures_num], function(x) as.numeric(gsub('[$,]', '', x)))

# Drop no longer needed columns
proc_12 <- proc_12[ , !(names(proc_12) %in% c("nppes_provider_country", "nppes_entity_code",
"average_medicare_allowed_amt", "average_submitted_chrg_amt", "average_medicare_payment_amt"))]

procedures_fac = c("npi", "nppes_provider_state", "provider_type", "medicare_participation_indicator",
"place_of_service", "hcpcs_code", "hcpcs_drug_indicator")
# Lowercase all factor variables to make them the same
proc_12[procedures_fac] <- lapply(proc_12[procedures_fac], tolower)
# Empty strings in factor variables are not recognized as NAs. We need to get them manually
proc_12[procedures_fac] <- lapply(proc_12[procedures_fac], as.character)
proc_12[procedures_fac] <- lapply(proc_12[procedures_fac], function(x) ifelse(x!="", x, NA))

# Group HCPCS codes
proc_12$hcpcs_desc <- gsub('^(0010[0-9]|001[1-9][0-9]|00[2-9][0-9]{2}|01[0-9]{3})$', 'Anesthesia',
proc_12$hcpcs_code)
proc_12$hcpcs_desc <- gsub('^(1000[4-9]|100[12][0-9])$', 'Fine Needle Aspiration Biopsy Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(77031|77032|1003[0-9]|100[4-9][0-9]|10[1-9][0-9]{2}|1[1-9][0-9]{3}|200[0-
9]{2})$', 'Surgical Procedures on the Integumentary System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(2010[0-9]|201[1-9][0-9]|20[2-9][0-9]{2}|2[1-9][0-9]{3})$', 'Surgical Procedures
on the Musculoskeletal System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(3000[0-9]|300[1-9][0-9]|30[1-9][0-9]{2}|3[12][0-9]{3})$', 'Surgical Procedures on
the Respiratory System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(3300[0-9]|330[1-9][0-9]|33[1-9][0-9]{2}|3[4-6][0-9]{3}|37[0-7][0-9]{2})$',
'Surgical Procedures on the Cardiovascular System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(3810[0-9]|381[1-9][0-9]|38[2-9][0-9]{2})$', 'Surgical Procedures on the Hemic and
Lymphatic Systems', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(3900[0-9]|390[1-9][0-9]|39[1-5][0-9]{2})$', 'Surgical Procedures on the
Mediastinum and Diaphragm', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(4049[0-9]|40[5-9][0-9]{2}|4[1-9][0-9]{3})$', 'Surgical Procedures on the
Digestive System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(5001[0-9]|500[2-9][0-9]|50[1-9][0-9]{2}|5[12][0-9]{3}|53[0-8][0-9]{2})$',
'Surgical Procedures on the Urinary System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(5400[0-9]|540[1-9][0-9]|54[1-9][0-9]{2}|55[0-8][0-9]{2})$', 'Surgical Procedures
on the Male Genital System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(55920)$', 'Reproductive System Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(5597[0-9]|55980)$', 'Intersex Surgery', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(5640[5-9]|564[1-9][0-9]|56[5-9][0-9]{2}|5[78][0-9]{3})$', 'Surgical Procedures on
the Female Genital System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(5900[0-9]|590[1-9][0-9]|59[1-8][0-9]{2})$', 'Surgical Procedures for Maternity
Care and Delivery', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(6000[0-9]|600[1-9][0-9]|60[1-6][0-9]{2})$', 'Surgical Procedures on the Endocrine
System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(6100[0-9]|610[1-9][0-9]|61[1-9][0-9]{2}|6[2-4][0-9]{3})$', 'Surgical Procedures
on the Nervous System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(6509[1-9]|65[1-9][0-9]{2}|6[67][0-9]{3}|68[0-8][0-9]{2})$', 'Surgical Procedures
on the Eye and Ocular Adnexa', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(6900[0-9]|690[1-9][0-9]|69[1-8][0-9]{2}|699[0-7][0-9])$', 'Surgical Procedures on
the Auditory System', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(69990)$', 'Operating Microscope Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(7001[0-9]|700[2-9][0-9]|70[1-9][0-9]{2}|7[1-5][0-9]{3}|76[0-4][0-9]{2})$',
'Diagnostic Radiology Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(7650[6-9]|765[1-9][0-9]|76[6-9][0-9]{2})$', 'Diagnostic Ultrasound Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(7700[1-9]|7701[0-9]|7702[0-2])$', 'Radiologic Guidance', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(7704[6-9]|7705[0-9]|7706[0-7])$', 'Breast and Mammography', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(7707[1-9]|7708[0-6])$', 'Bone or Joint Studies', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(7726[1-9]|772[7-9][0-9]|77[3-7][0-9]{2})$', 'Radiation Oncology Treatment',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(7800[0-9]|780[1-9][0-9]|78[1-9][0-9]{2}|79[0-9]{3})$', 'Nuclear Medicine
Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8004[7-9]|800[5-7][0-9]|8008[01])$', 'Organ or Disease Oriented Panels',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8010[0-9]|801[1-9][0-9]|802[0-9]{2}|803[0-6][0-9]|8037[0-7])$', 'Therapeutic Drug
Assays', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8030[5-9]|803[1-6][0-9]|8037[0-7])$', 'Drug Assay Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8040[0-9]|804[1-3][0-9])$', 'Evocative or Suppression Testing Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8050[0-2])$', 'Clinical Pathology Consultations', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8100[0-9]|810[1-9][0-9])$', 'Urinalysis Procedures', proc_12$hcpcs_desc)
```

```
proc_12$hcpcs_desc <- gsub('^(8110[5-9]|811[1-9][0-9]|81[23][0-9]{2}|8140[0-8])$', 'Molecular Pathology
Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8141[0-9]|814[2-7][0-9])$', 'Genomic Sequencing Procedures and Other Molecular
Multianalyte Assays', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8149[0-9]|815[0-9]{2})$', 'Multianalyte Assays with Algorithmic Analyses',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8200[0-9]|820[1-9][0-9]|82[1-9][0-9]{2}|8[34][0-9]{3})$', 'Chemistry Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8500[2-9]|850[1-9][0-9]|85[1-9][0-9]{2})$', 'Hematology and Coagulation
Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8600[0-9]|860[1-9][0-9]|86[1-7][0-9]{2}|868[0-4][0-9])$', 'Immunology
Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8685[0-9]|868[6-9][0-9]|869[0-9]{2})$', 'Transfusion Medicine Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8700[3-9]|870[1-9][0-9]|87[1-9][0-9]{2})$', 'Microbiology Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8800[0-9]|880[1-9][0-9])$', 'Postmortem Examination Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8810[4-9]|881[1-9][0-9])$', 'Cytopathology Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8823[0-9]|882[4-9][0-9])$', 'Cytogenetic Studies', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8830[0-9]|883[1-9][0-9])$', 'Surgical Pathology Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8872[0-9]|887[34][0-9])$', 'Transcutaneous Laboratory Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(89049|890[5-9][0-9]|891[0-9]{2}|892[0-3][0-9]|89240)$', 'Other Pathology and
Laboratory Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(8925[0-9]|892[6-9][0-9]|893[0-8][0-9]|8939[0-8])$', 'Reproductive Medicine
Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9028[1-9]|9029[0-9]|903[0-9]{2})$', 'Immune Globulins Serum or Recombinant
Products', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9046[0-9]|9047[0-4])$', 'Immunization Administration for Vaccines or Toxoids',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9047[6-9]|904[89][0-9]|90[56][0-9]{2}|907[0-4][0-9]|9075[0-6])$', 'Vaccines or
Toxoids', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9078[5-9]|9079[0-9]|908[0-9]{2})$', 'Psychiatry Services and Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9090[1-9]|9091[0-3])$', 'Biofeedback Services and Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9093[5-9]|909[4-9][0-9])$', 'Dialysis Services and Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9101[0-9]|910[2-9][0-9]|91[12][0-9]{2})$', 'Gastroenterology Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9200[2-9]|920[1-9][0-9]|92[1-4][0-9]{2})$', 'Ophthalmology Services and
Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9250[2-9]|925[1-9][0-9]|926[0-9]{2}|92700)$', 'Special Otorhinolaryngologic
Services and Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9292[0-9]|929[3-9][0-9]|93[0-7][0-9]{2})$', 'Cardiovascular Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9388[0-9]|9389[0-9]|939[0-8][0-9]|9399[0-8])$', 'Non-Invasive Vascular Diagnostic
Studies', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9400[2-9]|940[1-9][0-9]|94[1-7][0-9]{2})$', 'Pulmonary Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9500[4-9]|950[1-9][0-9]|951[0-9]{2})$', 'Allergy and Clinical Immunology
Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(95249|9525[01])$', 'Endocrinology Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9570[0-9]|957[1-9][0-9]|95[89][0-9]{2}|960[01][0-9]|96020)$', 'Neurology and
Neuromuscular Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(96040)$', 'Medical Genetics and Genetic Counseling Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9610[0-9]|961[1-3][0-9]|9614[0-6])$', 'Central Nervous System Assessments',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9615[0-9]|9616[0-9]|9617[01])$', 'Health and Behavior Assessment or Intervention
Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9636[0-9]|963[7-9][0-9]|964[0-9]{2}|965[0-4][0-9])$', 'Highly Complex Drug or
Biologic Agent Administration', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9656[7-9]|9657[0-4])$', 'Photodynamic Therapy Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9690[0-9]|969[1-9][0-9])$', 'Special Dermatological Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9700[0-9]|970[1-3][0-9])$', 'Application of a Modality', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9711[0-9]|971[2-4][0-9]|97150)$', 'Therapeutic Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9715[1-8])$', 'Adaptive Behavior Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9716[1-9]|971[7-9][0-9]|97[2-7][0-9]{2})$', 'Physical Medicine and Rehabilitation
Evaluations', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9780[2-4])$', 'Medical Nutrition Therapy Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9781[0-4])$', 'Acupuncture Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9892[5-9])$', 'Osteopathic Manipulative Treatment Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9894[0-3])$', 'Chiropractic Manipulative Treatment Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9896[0-2])$', 'Education and Training for Patient Self-Management',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9896[6-9]|9897[0-2])$', 'Non-Face-to-Face Nonphysician Services',
proc_12$hcpcs_desc)
```

```
proc_12$hcpcs_desc <- gsub('^(9900[0-9]|990[1-8][0-9]|9909[01])$', 'Special Services, Procedures and Reports',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9910[0-9]|991[1-3][0-9]|99140)$', 'Qualifying Circumstances for Anesthesia',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9914[1-9]|9915[0-7])$', 'Moderate (Conscious) Sedation', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9917[0-9]|991[89][0-9])$', 'Other Medicine Services and Procedures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9950[0-9]|995[1-9][0-9]|9960[0-2])$', 'Home Health Procedures and Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9960[5-7])$', 'Medication Therapy Management Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9920[1-9]|9921[0-5])$', 'Office or Other Outpatient Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9921[7-9]|9922[0-6])$', 'Hospital Observation Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9922[1-9]|9923[0-9])$', 'Hospital Inpatient Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9924[1-9]|9925[0-5])$', 'Consultation Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9928[1-8])$', 'Emergency Department Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(99291|99292)$', 'Critical Care Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9930[4-9]|9931[0-8])$', 'Nursing Facility Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9932[4-9]|993[34][0-9]|99350)$', 'Home Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9935[4-9]|993[6-9][0-9]|9940[0-9]|9941[0-6])$', 'Prolonged Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9936[6-8])$', 'Case Management Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9937[4-9]|99380)$', 'Care Plan Oversight Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9938[1-9]|9939[0-9]|994[0-2][0-9])$', 'Preventive Medicine Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9944[1-9]|9944[0-9])$', 'Non-Face-to-Face Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9945[0-8])$', 'Special Evaluation and Management Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9945[0-9]|9946[0-5])$', 'Newborn Care Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9946[6-9]|9947[0-9]|9948[0-6])$', 'Inpatient Neonatal Intensive Care Services and
Pediatric and Neonatal Critical Care Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9948[3-6])$', 'Cognitive Assessment and Care Plan Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9948[7-9]|9949[01])$', 'Care Management Evaluation and Management Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(9949[2-4])$', 'Psychiatric Collaborative Care Management Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(99495|99496)$', 'Transitional Care Evaluation and Management Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(99497|99498)$', 'Advance Care Planning Evaluation and Management Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(99499)$', 'Other Evaluation and Management Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(0{3}[1-9]|00[1-9][0-9]|0[1-9][0-9]{2}|[1-8][0-9]{3}|900[0-7])F$', 'Category II',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(0{3}[1-9])M$', 'Multianalyte Assay', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(0{3}[0-9]|00[1-9][0-9]|0[1-5][0-9]{2}|06[01][0-9])T$', 'Category III',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(0{3}[1-3])U$', 'Laboratory Analyses', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^A[0-9][0-9][0-9][0-9]$', 'Ambulatory Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^B[0-9][0-9][0-9][0-9]$', 'Enteral and Parenteral Therapy', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^C[0-9][0-9][0-9][0-9]$', 'Outpatient PPS', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^D[0-9][0-9][0-9][0-9]$', 'Dental Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^E[0-9][0-9][0-9][0-9]$', 'Durable Medical Equipment', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(9141|0{3}[89]|0010)$', 'Vaccine Administration', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G0027$', 'Analysis of Semen Specimen', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(006[89]|0070)$', 'Professional Services for Drug Infusion', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G0071$', 'Telemed Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(007[6-9]|008[0-7])$', 'Home Care Management Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(010[1-9]|011[0-9]|012[0-4])$', 'Screening Examinations and Disease Management
Training', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(038[5-9]|012[7-9]|01[3-9][0-9]|02[0-9]{2}|03[0-6][0-9]|037[0-2])$',
'Miscellaneous Diagnostic and Therapeutic Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(037[89]|038[0-4])$', 'Hospital Observation and Emergency Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G0390$', 'Other Emergency Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(0396|0397)$', 'Alcohol and Substance Abuse Assessments', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(039[89]|0400)$', 'Sleep Studies, In Home', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(040[2-5])$', 'Initial Services for Medicare Enrollment', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(040[6-8])$', 'Followup Telehealth Consultations', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(0409|041[01])$', 'Psychological Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(041[2-5])$', 'Fracture Treatment', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(041[6-9])$', 'Gross and Microscopic Examinations Prostate Biopsy',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(0420|0421)$', 'Face-to-Face Educational Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(042[2-4])$', 'Cardiac and Pulmonary Rehabilitation Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(042[5-7])$', 'Initial Telehealth Consultations', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(0428|0429)$', 'Filler Procedures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(043[1-5])$', 'Laboratory Screening Tests', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(043[6-9]|044[0-9]|045[01])$', 'Counseling, Screening, and Prevention Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(045[2-9]|046[0-3])$', 'Miscellaneous Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(046[6-9]|0470)$', 'Federally Qualified Health Center Visits',
```

```
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(047[1-9]|04[89][0-9]|05[0-9]{2}|06[0-5][0-9])$', 'Other Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(091[3-8])$', 'Quality Measures for Cataract Surgery', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(100[1-9]|101[0-9]|102[0-3])$', 'Clinical Decision Support Mechanism',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G2000$', 'Convulsive Therapy Procedure', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(2170|2171)$', 'Arteriovenous Fistula Procedure', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(2166)$', 'Physical and Cognitive Status', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(202[1-9]|20[34][0-9]|205[0-8])$', 'Care Management Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(2168|2169)$', 'Therapy Maintenance Program in Home Health Setting',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(2167)$', 'Performance Measures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(206[1-3])$', 'Online Assessment by Qualified Nonphysician Healthcare
Professional', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(2064|2065)$', 'Comprehensive Care Management Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(2066)$', 'Cardiac Device Evaluation', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(206[7-9]|207[0-5])$', 'Medication Assisted Treatment Programme',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(207[6-9]|208[01])$', 'Opioid Use Disorder - Evaluation and Treatment',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(2082|2083)$', 'Evaluation and Management Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(208[6-8])$', 'Opioid Use Disorder - Treatment', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(2089|209[0-9]|21[0-4][0-9]|215[0-3])$', 'Functional Status', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(215[4-9]|216[0-5])$', 'Immunization Status', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(200[1-9]|201[0-5])$', 'Other Evaluation and Management Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(600[1-9]|601[0-7])$', 'Radiation Therapy Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(6019|602[0-9])$', 'Colonoscopy or Anoscopy', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(603[0-9]|604[0-9]|605[0-8])$', 'Drug Assay', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(839[5-9]|8[45][0-9]{2}|86[0-2][0-9]|863[0-5])$', 'Additional Quality Measures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(864[7-9]|86[56][0-9]|867[0-4])$', 'Quality Measures Related for Risk-adjusted
Functional Status Scoring', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(869[4-9]|8[78][0-9]{2}|89[0-6][0-9]|897[0-6])$', 'More Quality Measures',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(897[89]|89[89][0-9])$', 'Functional Limitation Reporting', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(900[1-9]|901[0-2])$', 'Medicare Coordinated Care Demonstration',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(901[3-9]|90[2-9][0-9]|91[0-3][0-9]|9140)$', 'Medicare Demonstration Projects',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G9143$', 'Warfarin Responsiveness Testing', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G9147$', 'Outpatient Intravenous Insulin Treatment', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(914[89]|915[0-3])$', 'Primary Care Quality Measures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G9156$', 'Provider Assessment for Wheelchair', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G9157$', 'Diagnostic Cardiac Doppler Ultrasound', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(915[89]|91[67][0-9]|918[0-6])$', 'Functional Limitation Reporting',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G9187$', 'Bundled Payment Care', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(918[89]|919[0-9]|9[2-7][0-9]{2}|98[0-8][0-9]|989[0-3])$', 'Additional Assorted
Quality Measures', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(989[4-7])$', 'Radiology services Prostate', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^(G9898|G9910)$', 'Geriatric care Management', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(990[2-9])$', 'Tobacco Screening', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(991[2-5])$', 'Anti TNF Diagnostics for HBV status', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(991[6-8])$', 'Functional Status codes', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(9911|9919|992[0-9]|993[0-7])$', 'Screening', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(993[89]|9940)$', 'Geriatric Care Management and Other Services',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(994[1-9])$', 'Pain assessment', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(995[4-9]|996[01])$', 'Medications - Antiemetics and Antimicrobials',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(9962|9963)$', 'Embolization', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(996[4-9]|9970)$', 'Screening, Wellness and Physician visits',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(997[4-7])$', 'Vision Assessment', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^G(997[89]|998[0-7])$', 'Remote In-House Evaluation And Management Assessment',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^H[0-9][0-9][0-9][0-9]$', 'Alcohol and Drug Abuse Treatment', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^J(900[0-9]|90[1-9][0-9]|9[1-9][0-9]{2})$', 'Chemotherapy Drugs',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^J(012[0-9]|01[3-9][0-9]|0[2-9][0-9]{2}|[1-8][0-9]{3})$', 'Drugs Administered Other
than Oral Method', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^K[0-9][0-9][0-9][0-9]$', 'Durable Medical Equipment Offered Only by Regional
Carriers', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^L[0-9][0-9][0-9][0-9]$', 'Orthotics and Prosthetics', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^M[0-9][0-9][0-9][0-9]$', 'Miscellaneous Medical Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^P[0-9][0-9][0-9][0-9]$', 'Pathology and Laboratory Services', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^Q[0-9][0-9][0-9][0-9]$', 'Temporary Codes', proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^R[0-9][0-9][0-9][0-9]$', 'Diagnostic Radiology Services', proc_12$hcpcs_desc)
```

```
proc_12$hcpcs_desc <- gsub('^S[0-9][0-9][0-9][0-9]$', 'Temporary National Codes Non-Medicare',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^T[0-9][0-9][0-9][0-9]$', 'National Codes Established for State Medicaid Agencies',
proc_12$hcpcs_desc)
proc_12$hcpcs_desc <- gsub('^V[0-9][0-9][0-9][0-9]$', 'Vision and Hearing Services', proc_12$hcpcs_desc)

# Factorize character variables
proc_12[procedures_fac] <- lapply(proc_12[procedures_fac], as.factor)
proc_12 <- droplevels(proc_12)

# Group and pivot the number of procedures by provider
proc_12_group <- proc_12 %>% group_by(npi, hcpcs_desc) %>% summarise(overcharge=mean(overcharge),
line_srvc_cnt=sum(line_srvc_cnt), bene_unique_cnt=sum(bene_unique_cnt))
proc_12_pivot <- proc_12_group %>% pivot_wider(id_cols='npi', names_from='hcpcs_desc',
values_from='line_srvc_cnt', values_fill=0)
```

Combine providers with procedures

```
d_12 <- merge(prov_12, proc_12_pivot, on="npi", all = FALSE, all.x = TRUE)
d_12 <- complete.cases(d_12, )
```

After ALL years were combined, add a "year" column and combine all years together

```
d_12$year = 2012
d_13$year = 2013
d_14$year = 2014
d_15$year = 2015
d_16$year = 2016
d_17$year = 2017
# Combine tables
d <- rbind.fill(d_12, d_13, d_14, d_15, d_16, d_17)
# Transform column names
colnames(d) <- tolower(colnames(d))
colnames(d) <- gsub("\\.\\.+", ".", colnames(d))
colnames(d) <- gsub("\\.", "_", colnames(d))
str(d, list.len=ncol(d))
```

```
'data.frame':       103599 obs. of  195 variables:
 $ npi_year                                                    : Factor w/ 103599 levels "10030001342013",..: 1 2 3 4 5 6 7 8 9 10
...
 $ npi                                                         : Factor w/ 52766 levels "1003000134","1003002072",..: 1 1 2 2 2 3
4 4 4 4 ...
 $ year                                                        : int  2013 2014 2015 2016 2017 2014 2012 2013 2015 2016 ...
 $ overcharge                                                  : num  993634 1111775 1298455 2286168 2091698 ...
 $ nppes_provider_state                                        : Factor w/ 59 levels "AA","AE","AK",..: 19 19 37 37 37 44 40 40 40
40 ...
 $ provider_type                                               : Factor w/ 98 levels "addiction medicine",..: 72 72 97 97 97 26 26
26 26 26 ...
 $ number_of_hcpcs                                             : int  13 13 84 97 84 98 118 115 148 147 ...
 $ total_unique_benes                                          : int  3940 4276 701 863 845 1909 1704 1987 3321 2750 ...
 $ number_of_med_hcpcs                                         : int  13 13 79 91 79 98 117 114 148 147 ...
 $ total_med_services                                          : num  7518 8643 2675 4245 3702 ...
 $ total_med_unique_benes                                      : int  3940 4276 701 863 845 1909 1704 1987 3321 2750 ...
 $ total_med_submitted_chrg_amt                                : num  1211595 1357139 1492179 2523770 2290842 ...
 $ total_med_medicare_allowed_amt                              : num  285504 321642 283387 450678 403728 ...
 $ total_med_medicare_payment_amt                              : num  217961 245364 219805 346535 310574 ...
 $ beneficiary_average_age                                     : int  76 76 70 71 71 71 72 72 72 73 ...
 $ beneficiary_age_less_65_count                               : int  100 102 198 222 231 483 194 209 386 297 ...
 $ beneficiary_age_65_74_count                                 : int  1669 1856 217 276 280 612 869 1029 1584 1368 ...
 $ beneficiary_age_75_84_count                                 : int  1472 1500 183 206 193 453 480 566 992 795 ...
 $ beneficiary_age_greater_84_count                            : int  699 818 103 159 141 361 161 183 359 290 ...
 $ beneficiary_female_count                                    : int  2025 2252 197 260 285 1049 1033 1217 2019 1685 ...
 $ beneficiary_male_count                                      : int  1915 2024 504 603 560 860 671 770 1302 1065 ...
 $ beneficiary_race_white_count                                : int  3736 4057 504 660 670 1596 1531 1767 2647 2221 ...
 $ beneficiary_race_black_count                                : int  53 36 103 108 85 261 28 33 136 99 ...
 $ beneficiary_race_api_count                                  : int  41 47 14 13 15 19 69 67 113 111 ...
 $ beneficiary_race_hispanic_count                             : int  71 69 67 68 53 17 55 80 320 223 ...
 $ beneficiary_race_natind_count                               : int  0 0 0 0 0 0 0 0 0 0 ...
 $ beneficiary_nondual_count                                   : int  3781 4132 376 486 481 1299 1547 1835 2721 2323 ...
 $ beneficiary_dual_count                                      : int  159 144 325 377 364 610 157 152 600 427 ...
 $ beneficiary_cc_afib_percent                                 : int  8 12 16 19 18 25 9 9 11 9 ...
 $ beneficiary_cc_alzrdsd_percent                              : int  4 7 19 26 26 20 7 7 11 11 ...
 $ beneficiary_cc_asthma_percent                               : int  12 4 20 33 11 12 9 10 11 15 ...
 $ beneficiary_cc_cancer_percent                               : int  13 13 23 23 18 16 10 11 11 9 ...
 $ beneficiary_cc_chf_percent                                  : int  14 11 31 32 32 42 14 13 14 14 ...
 $ beneficiary_cc_ckd_percent                                  : int  8 14 48 54 56 49 10 10 17 23 ...
 $ beneficiary_cc_copd_percent                                 : int  11 7 29 30 27 34 17 17 11 11 ...
 $ beneficiary_cc_depr_percent                                 : int  20 12 29 32 37 38 32 30 19 20 ...
 $ beneficiary_cc_diab_percent                                 : int  11 20 47 47 47 45 12 14 35 33 ...
 $ beneficiary_cc_hyperl_percent                               : int  57 53 70 71 63 62 63 64 67 65 ...
 $ beneficiary_cc_hypert_percent                               : int  60 57 75 75 75 75 66 67 71 72 ...
 $ beneficiary_cc_ihd_percent                                  : int  29 28 55 58 57 59 32 33 35 33 ...
 $ beneficiary_cc_ost_percent                                  : int  8 9 6 6 7 12 9 7 11 12 ...
 $ beneficiary_cc_raoa_percent                                 : int  36 37 45 49 46 49 39 42 52 53 ...
 $ beneficiary_cc_schiot_percent                               : int  1 1 10 10 8 10 2 2 2 2 ...
 $ beneficiary_cc_strk_percent                                 : int  4 3 11 13 12 19 4 5 7 5 ...
 $ beneficiary_average_risk_score                              : num  1.02 1 1.98 1.96 1.97 ...
 $ additional_assorted_quality_measures                        : int  0 0 0 0 0 0 0 0 0 0 ...
 $ administrative_miscellaneous_and_investigational            : num  0 0 0 0 0 0 0 0 0 0 ...
 $ advance_care_planning_evaluation_and_management_services    : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
$ alcohol_and_substance_abuse_assessments                                : int  0 0 0 0 0 0 0 0 0 0 ...
$ allergy_and_clinical_immunology_procedures                             : int  0 0 0 0 0 0 0 0 0 0 ...
$ ambulance_and_other_transport_services_and_supplies                    : int  0 0 0 0 0 0 0 0 0 0 ...
$ analysis_of_semen_specimen                                             : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_burn_excisions_or_debridement_procedures                : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_intrathoracic_or_procedures_on_the_thorax               : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_obstetric_procedures                                    : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_other_procedures                                        : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_procedures_on_the_abdomen                               : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_procedures_on_the_head_or_neck                          : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_procedures_on_the_leg                                   : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_procedures_on_the_perineum_or_pelvis                    : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_procedures_on_the_shoulder_or_arm                       : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_procedures_on_the_spine_and_spinal_cord                 : int  0 0 0 0 0 0 0 0 0 0 ...
$ anesthesia_for_radiological_procedures                                 : int  0 0 0 0 0 0 0 0 0 0 ...
$ application_of_a_modality                                              : int  0 0 0 0 0 0 0 0 0 0 ...
$ biofeedback_services_and_procedures                                    : int  0 0 0 0 0 0 0 0 0 0 ...
$ bone_or_joint_studies                                                  : num  0 0 0 0 0 0 64 87 182 205 ...
$ breast_and_mammography                                                 : num  0 0 0 0 0 0 60 106 19 0 ...
$ bundled_payment_care                                                   : int  0 0 0 0 0 0 0 0 0 0 ...
$ cardiac_and_pulmonary_rehabilitation_services                          : int  0 0 0 0 0 0 0 0 0 0 ...
$ cardiovascular_procedures                                              : num  0 0 0 0 0 0 0 0 0 0 ...
$ care_management_evaluation_and_management_services                     : int  0 0 0 0 0 0 0 0 0 0 ...
$ category_ii                                                            : int  0 0 0 0 0 0 0 0 0 0 ...
$ category_iii                                                           : num  0 0 0 0 0 0 0 0 0 0 ...
$ central_nervous_system_assessments                                     : num  0 0 0 0 0 0 0 0 0 0 ...
$ chemistry_procedures                                                   : int  0 0 0 0 0 0 0 0 0 0 ...
$ chemotherapy_drugs                                                     : num  0 0 24 52 81 0 0 0 0 0 ...
$ chiropractic_manipulative_treatment_procedures                         : int  0 0 0 0 0 0 0 0 0 0 ...
$ clinical_pathology_consultations                                       : int  0 0 0 0 0 0 0 0 0 0 ...
$ colonoscopy_or_anoscopy                                                : int  0 0 0 0 0 0 0 0 0 0 ...
$ counseling_screening_and_prevention_services                           : int  0 0 0 53 0 0 0 0 0 0 ...
$ critical_care_services                                                 : num  0 0 0 0 0 0 0 0 0 0 ...
$ cytogenetic_studies                                                    : int  0 0 0 0 0 0 0 0 0 0 ...
$ cytopathology_procedures                                               : num  0 0 0 0 0 0 0 0 0 0 ...
$ dental_services                                                        : int  0 0 0 0 0 0 0 0 0 0 ...
$ diagnostic_cardiac_doppler_ultrasound                                  : int  0 0 0 0 0 0 0 0 0 0 ...
$ diagnostic_radiology_procedures                                        : num  0 0 0 15 15 ...
$ diagnostic_radiology_services                                          : int  0 0 0 0 0 0 0 0 0 0 ...
$ diagnostic_ultrasound_procedures                                       : num  0 0 0 44 0 364 321 405 474 274 ...
$ dialysis_services_and_procedures                                       : int  0 0 0 0 0 0 0 0 0 0 ...
$ drugs_administered_other_than_oral_method                              : num  0 0 0 247 285 0 0 0 0 0 ...
$ drug_assay                                                             : int  0 0 0 0 0 0 0 0 0 0 ...
$ drug_assay_procedures                                                  : int  0 0 0 0 0 0 0 0 0 0 ...
$ durable_medical_equipment                                              : int  0 0 0 0 0 0 0 0 0 0 ...
$ durable_medical_equipment_offered_only_by_regional_carriers            : int  0 0 0 0 0 0 0 0 0 0 ...
$ emergency_department_services                                          : int  0 0 12 16 0 0 0 0 0 0 ...
$ endocrinology_services                                                 : int  0 0 0 0 0 0 0 0 0 0 ...
$ evocative_or_suppression_testing_procedures                            : int  0 0 0 0 0 0 0 0 0 0 ...
$ face_to_face_educational_services                                      : int  0 0 0 0 0 0 0 0 0 0 ...
$ federally_qualified_health_center_visits                               : int  0 0 0 0 0 0 0 0 0 0 ...
$ filler_procedures                                                      : int  0 0 0 0 0 0 0 0 0 0 ...
$ fine_needle_aspiration_biopsy_procedures                               : int  0 0 0 0 0 0 0 0 0 0 ...
$ followup_telehealth_consultations                                      : int  0 0 0 0 0 0 0 0 0 0 ...
$ fracture_treatment                                                     : int  0 0 0 0 0 0 0 0 0 0 ...
$ gastroenterology_procedures                                            : int  0 0 0 0 0 0 0 0 0 0 ...
$ genomic_sequencing_procedures_and_other_molecular_multianalyte_assays: int  0 0 0 0 0 0 0 0 0 0 ...
$ gross_and_microscopic_examinations_prostate_biopsy                     : int  0 0 0 0 0 0 0 0 0 0 ...
$ health_and_behavior_assessment_or_intervention_procedures              : int  0 0 0 0 0 0 0 0 0 0 ...
$ hematology_and_coagulation_procedures                                  : int  0 0 0 0 0 0 0 0 0 0 ...
$ highly_complex_drug_or_biologic_agent_administration                   : num  0 0 47 83 94 0 0 0 0 0 ...
$ home_services                                                          : int  0 0 0 0 0 0 0 0 0 0 ...
$ hospital_inpatient_services                                            : num  0 0 64 133 247 0 0 0 0 0 ...
$ hospital_observation_services                                          : num  0 0 174 282 216 0 0 0 0 0 ...
$ immune_globulins_serum_or_recombinant_products                         : int  0 0 0 0 0 0 0 0 0 0 ...
$ immunization_administration_for_vaccines_or_toxoids                    : int  0 0 0 0 0 0 0 0 0 0 ...
$ immunology_procedures                                                  : num  0 0 0 0 0 0 0 0 0 0 ...
$ initial_services_for_medicare_enrollment                               : int  0 0 0 0 0 0 0 0 0 0 ...
$ initial_telehealth_consultations                                       : int  0 0 0 0 0 0 0 0 0 0 ...
$ laboratory_screening_tests                                             : int  0 0 0 0 0 0 0 0 0 0 ...
$ medical_and_surgical_supplies                                          : int  0 0 0 0 0 0 0 0 0 0 ...
$ medical_nutrition_therapy_procedures                                   : int  0 0 0 0 0 0 0 0 0 0 ...
$ medicare_coordinated_care_demonstration                                : int  0 0 0 0 0 0 0 0 0 0 ...
$ microbiology_procedures                                                : int  0 0 0 0 0 0 0 0 0 0 ...
$ miscellaneous_diagnostic_and_therapeutic_services                      : num  0 0 0 0 0 0 58 109 21 0 ...
$ miscellaneous_medical_services                                         : int  0 0 0 0 0 0 0 0 0 0 ...
$ miscellaneous_services                                                 : num  0 757 0 0 0 0 0 0 0 0 ...
$ moderate_conscious_sedation                                            : num  0 0 0 0 0 0 0 0 0 0 ...
$ molecular_pathology_procedures                                         : int  0 0 0 0 0 0 0 0 0 0 ...
$ multianalyte_assay                                                     : int  0 0 0 0 0 0 0 0 0 0 ...
$ neurology_and_neuromuscular_procedures                                 : num  0 0 18 51 57 0 0 0 0 0 ...
$ non_invasive_vascular_diagnostic_studies                               : num  0 0 0 0 0 324 90 87 211 130 ...
$ nuclear_medicine_procedures                                            : num  0 0 0 0 0 0 0 0 0 0 ...
$ nursing_facility_services                                              : int  0 0 0 0 0 0 0 0 0 0 ...
$ office_or_other_outpatient_services                                    : num  0 0 1078 1504 1486 ...
$ operating_microscope_procedures                                        : int  0 0 0 0 0 0 0 0 0 0 ...
$ ophthalmology_services_and_procedures                                  : num  0 0 0 0 0 0 0 0 0 0 ...
$ organ_or_disease_oriented_panels                                       : int  0 0 0 0 0 0 0 0 0 0 ...
$ orthotics_and_prosthetics                                              : int  0 0 0 0 0 0 0 0 0 0 ...
$ osteopathic_manipulative_treatment_procedures                          : int  0 0 0 0 0 0 0 0 0 0 ...
$ other_evaluation_and_management_services                               : int  0 0 0 0 0 0 0 0 0 0 ...
$ other_medicine_services_and_procedures                                 : int  0 0 0 0 0 0 0 0 0 0 ...
$ other_pathology_and_laboratory_procedures                              : int  0 0 0 0 0 0 0 0 0 0 ...
$ other_services                                                         : int  0 0 0 0 0 0 0 0 0 0 ...
$ pathology_and_laboratory_services                                      : num  0 0 0 0 0 0 0 0 0 0 ...
$ photodynamic_therapy_procedures                                        : int  0 0 0 0 0 0 0 0 0 0 ...
$ physical_medicine_and_rehabilitation_evaluations                       : num  0 0 0 0 0 0 0 0 0 0 ...
$ primary_care_quality_measures                                          : int  0 0 0 0 0 0 0 0 0 0 ...
$ prolonged_services                                                     : int  0 0 29 0 14 0 0 0 0 0 ...
$ psychiatry_services_and_procedures                                     : int  0 0 0 0 0 0 0 0 0 0 ...
$ psychological_services                                                 : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
 $ pulmonary_procedures                                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ radiation_oncology_treatment                                  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ radiation_therapy_services                                    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ radiologic_guidance                                           : int  0 0 0 0 0 0 0 0 0 0 ...
 $ reproductive_system_procedures                                : int  0 0 0 0 0 0 0 0 0 0 ...
 $ screening_examinations_and_disease_management_training        : num  0 0 0 0 0 0 0 0 0 0 ...
 $ sleep_studies_in_home                                         : int  0 0 0 0 0 0 0 0 0 0 ...
 $ special_dermatological_procedures                             : int  0 0 0 0 0 0 0 0 0 0 ...
 $ special_otorhinolaryngologic_services_and_procedures          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ special_services_procedures_and_reports                       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_pathology_procedures                                 : num  7500 7867 0 0 0 ...
 $ surgical_procedures_for_maternity_care_and_delivery           : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_auditory_system                    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_cardiovascular_system              : num  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_digestive_system                   : int  0 0 0 0 16 24 0 0 0 0 ...
 $ surgical_procedures_on_the_endocrine_system                   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_eye_and_ocular_adnexa              : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_female_genital_system              : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_hemic_and_lymphatic_systems        : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_integumentary_system               : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_male_genital_system                : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_mediastinum_and_diaphragm          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_musculoskeletal_system             : int  0 0 0 0 0 0 0 0 16 0 ...
 $ surgical_procedures_on_the_nervous_system                     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_respiratory_system                 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ surgical_procedures_on_the_urinary_system                     : int  0 0 754 1211 937 0 0 0 0 0 ...
 $ temporary_codes                                               : num  0 0 0 0 0 ...
 $ temporary_national_codes_non_medicare                         : int  0 0 0 0 0 0 0 0 0 0 ...
 $ therapeutic_drug_assays                                       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ therapeutic_procedures                                        : int  0 0 0 0 0 0 0 0 0 0 ...
 $ transcutaneous_laboratory_procedures                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ transfusion_medicine_procedures                               : int  0 0 0 0 0 0 0 0 0 0 ...
 $ transitional_care_evaluation_and_management_services          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ urinalysis_procedures                                         : int  0 0 298 561 432 0 0 0 0 0 ...
 $ vaccine_administration                                        : num  0 0 0 0 0 0 0 0 0 0 ...
 $ vaccines_or_toxoids                                           : int  0 0 0 0 0 0 0 0 0 0 ...
 $ vision_and_hearing_services                                   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ warfarin_responsiveness_testing                               : int  0 0 0 0 0 0 0 0 0 0 ...
 $ ambulatory_services                                           : num  0 0 0 0 0 0 0 0 0 0 ...
 $ anesthesia                                                    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ chemotherapy_drugs_and_drugs_administered_other_than_oral_method : num  0 0 24 299 366 0 0 0 0 0 ...
 $ medicine_services_and_procedures                              : num  0 0 1422 2069 2114 ...
 $ pathology_and_laboratory_procedures                           : num  7500 7867 298 561 432 ...
 $ radiology_procedures                                          : num  0 0 0 59 15 ...
 $ surgery                                                       : num  0 0 754 1211 953 ...
 $ temporary_procedures_and_professional_services                : num  0 757 0 53 0 0 58 109 21 0 ...
```

Some of the HCPCS groups have too few values. They can be aggregated into larger groups.

```
# Number of non-zero rows per column
colSums(d != 0)
# New groups
d$vaccines_or_toxoids <- d$vaccines_or_toxoids + d$immunization_administration_for_vaccines_or_toxoids +
d$vaccine_administration
d$psychiatry_services_and_procedures <- d$psychiatry_services_and_procedures +
d$biofeedback_services_and_procedures
d$neurology_and_neuromuscular_procedures <- d$neurology_and_neuromuscular_procedures +
d$central_nervous_system_assessments + d$health_and_behavior_assessment_or_intervention_procedures
d$therapeutic_procedures <- d$therapeutic_procedures + d$photodynamic_therapy_procedures +
d$application_of_a_modality + d$medical_nutrition_therapy_procedures +
d$osteopathic_manipulative_treatment_procedures + d$chiropractic_manipulative_treatment_procedures
d$other_medicine_services_and_procedures <- d$other_medicine_services_and_procedures +
d$special_dermatological_procedures
d$other_care_evaluation_and_management_services <- d$care_management_evaluation_and_management_services +
d$transitional_care_evaluation_and_management_services +
d$advance_care_planning_evaluation_and_management_services + d$other_evaluation_and_management_services
d$drug_assay <- d$drug_assay + d$drug_assay_procedures + d$therapeutic_drug_assays +
d$genomic_sequencing_procedures_and_other_molecular_multianalyte_assays
d$cytopathology_procedures <- d$cytopathology_procedures + d$cytogenetic_studies
d$other_pathology_and_laboratory_procedures <- d$other_pathology_and_laboratory_procedures +
d$evocative_or_suppression_testing_procedures + d$clinical_pathology_consultations +
d$transfusion_medicine_procedures + d$transcutaneous_laboratory_procedures +
d$pathology_and_laboratory_services
d$diagnostic_radiology_procedures <- d$diagnostic_radiology_procedures + d$diagnostic_radiology_services
d$radiation_oncology_treatment <- d$radiation_oncology_treatment + d$radiation_therapy_services
d$surgical_procedures_on_genital_system <- d$surgical_procedures_on_the_female_genital_system +
d$surgical_procedures_on_the_male_genital_system
d$other_surgical_procedures <- d$surgical_procedures_on_the_endocrine_system +
d$surgical_procedures_on_the_mediastinum_and_diaphragm + d$operating_microscope_procedures
d$miscellaneous_services <- d$miscellaneous_services + d$alcohol_and_substance_abuse_assessments +
d$sleep_studies_in_home + d$followup_telehealth_consultations + d$face_to_face_educational_services +
d$cardiac_and_pulmonary_rehabilitation_services + d$initial_telehealth_consultations + d$filler_procedures +
d$laboratory_screening_tests + d$other_services + d$colonoscopy_or_anoscopy +
d$medicare_coordinated_care_demonstration + d$primary_care_quality_measures +
d$additional_assorted_quality_measures + d$vision_and_hearing_services + d$miscellaneous_medical_services +
d$temporary_codes + d$orthotics_and_prosthetics + d$durable_medical_equipment
d <- d[c("npi_year", "npi", "year", "overcharge", "nppes_provider_state", "provider_type", "number_of_hcpcs",
"total_unique_benes", "number_of_med_hcpcs", "total_med_services", "total_med_unique_benes",
"total_med_submitted_chrg_amt", "total_med_medicare_allowed_amt", "total_med_medicare_payment_amt",
```

```
"beneficiary_average_age", "beneficiary_age_less_65_count", "beneficiary_age_65_74_count",
"beneficiary_age_75_84_count", "beneficiary_age_greater_84_count", "beneficiary_female_count",
"beneficiary_male_count", "beneficiary_race_white_count", "beneficiary_race_black_count",
"beneficiary_race_api_count", "beneficiary_race_hispanic_count", "beneficiary_race_natind_count",
"beneficiary_nondual_count", "beneficiary_dual_count", "beneficiary_cc_afib_percent",
"beneficiary_cc_alzrdsd_percent", "beneficiary_cc_asthma_percent", "beneficiary_cc_cancer_percent",
"beneficiary_cc_chf_percent", "beneficiary_cc_ckd_percent", "beneficiary_cc_copd_percent",
"beneficiary_cc_depr_percent", "beneficiary_cc_diab_percent", "beneficiary_cc_hyperl_percent",
"beneficiary_cc_hypert_percent", "beneficiary_cc_ihd_percent", "beneficiary_cc_ost_percent",
"beneficiary_cc_raoa_percent", "beneficiary_cc_schiot_percent", "beneficiary_cc_strk_percent",
"beneficiary_average_risk_score", "ambulatory_services", "anesthesia", "category_iii",
"drugs_administered_other_than_oral_method", "chemotherapy_drugs", "vaccines_or_toxoids",
"psychiatry_services_and_procedures", "dialysis_services_and_procedures", "gastroenterology_procedures",
"ophthalmology_services_and_procedures", "special_otorhinolaryngologic_services_and_procedures",
"cardiovascular_procedures", "non_invasive_vascular_diagnostic_studies", "pulmonary_procedures",
"allergy_and_clinical_immunology_procedures", "endocrinology_services",
"neurology_and_neuromuscular_procedures", "highly_complex_drug_or_biologic_agent_administration",
"therapeutic_procedures", "physical_medicine_and_rehabilitation_evaluations", "moderate_conscious_sedation",
"other_medicine_services_and_procedures", "office_or_other_outpatient_services",
"hospital_observation_services", "hospital_inpatient_services", "emergency_department_services",
"critical_care_services", "nursing_facility_services", "home_services", "prolonged_services",
"other_care_evaluation_and_management_services", "organ_or_disease_oriented_panels", "drug_assay",
"urinalysis_procedures", "chemistry_procedures", "hematology_and_coagulation_procedures",
"immunology_procedures", "microbiology_procedures", "cytopathology_procedures",
"surgical_pathology_procedures", "diagnostic_radiology_procedures", "diagnostic_ultrasound_procedures",
"radiologic_guidance", "breast_and_mammography", "bone_or_joint_studies", "radiation_oncology_treatment",
"nuclear_medicine_procedures", "fine_needle_aspiration_biopsy_procedures",
"surgical_procedures_on_the_auditory_system", "surgical_procedures_on_the_cardiovascular_system",
"surgical_procedures_on_the_digestive_system", "surgical_procedures_on_the_eye_and_ocular_adnexa",
"surgical_procedures_on_genital_system", "surgical_procedures_on_the_hemic_and_lymphatic_systems",
"surgical_procedures_on_the_integumentary_system", "surgical_procedures_on_the_musculoskeletal_system",
"surgical_procedures_on_the_nervous_system", "surgical_procedures_on_the_respiratory_system",
"surgical_procedures_on_the_urinary_system", "other_surgical_procedures",
"screening_examinations_and_disease_management_training", "miscellaneous_diagnostic_and_therapeutic_services",
"initial_services_for_medicare_enrollment", "gross_and_microscopic_examinations_prostate_biopsy",
"counseling_screening_and_prevention_services", "miscellaneous_services")]
colSums(d != 0)
```

```
                                                 npi_year                                                      npi
                                                   103599                                                   103599
                                                      year                                                overcharge
                                                   103599                                                   103599
                                       nppes_provider_state                                        provider_type
                                                   103599                                                   103599
                                           number_of_hcpcs                                  total_unique_benes
                                                   103599                                                   103599
                                        number_of_med_hcpcs                                   total_med_services
                                                   103599                                                   103599
                                      total_med_unique_benes                         total_med_submitted_chrg_amt
                                                   103599                                                   103599
                                total_med_medicare_allowed_amt                        total_med_medicare_payment_amt
                                                   103599                                                   103599
                                       beneficiary_average_age                        beneficiary_age_less_65_count
                                                   103599                                                   103583
                                  beneficiary_age_65_74_count                          beneficiary_age_75_84_count
                                                   103599                                                   103599
                              beneficiary_age_greater_84_count                             beneficiary_female_count
                                                   103590                                                   103599
                                       beneficiary_male_count                        beneficiary_race_white_count
                                                   103547                                                   103561
                                  beneficiary_race_black_count                          beneficiary_race_api_count
                                                   102814                                                    99068
                                 beneficiary_race_hispanic_count                        beneficiary_race_natind_count
                                                   101180                                                    23937
                                      beneficiary_nondual_count                               beneficiary_dual_count
                                                   103599                                                   103597
                                   beneficiary_cc_afib_percent                       beneficiary_cc_alzrdsd_percent
                                                   103599                                                   103599
                                  beneficiary_cc_asthma_percent                        beneficiary_cc_cancer_percent
                                                   103599                                                   103599
                                     beneficiary_cc_chf_percent                           beneficiary_cc_ckd_percent
                                                   103599                                                   103599
                                   beneficiary_cc_copd_percent                          beneficiary_cc_depr_percent
                                                   103599                                                   103599
                                   beneficiary_cc_diab_percent                        beneficiary_cc_hyperl_percent
                                                   103599                                                   103599
                                  beneficiary_cc_hypert_percent                           beneficiary_cc_ihd_percent
                                                   103599                                                   103599
                                     beneficiary_cc_ost_percent                          beneficiary_cc_raoa_percent
                                                   103599                                                   103599
                                  beneficiary_cc_schiot_percent                          beneficiary_cc_strk_percent
                                                   103534                                                   103599
                                 beneficiary_average_risk_score                  additional_assorted_quality_measures
                                                   103599                                                      263
          administrative_miscellaneous_and_investigational      advance_care_planning_evaluation_and_management_services
                                                    16199                                                      471
                   alcohol_and_substance_abuse_assessments                   allergy_and_clinical_immunology_procedures
                                                        9                                                      693
      ambulance_and_other_transport_services_and_supplies                               analysis_of_semen_specimen
                                                        0                                                        0
 anesthesia_for_burn_excisions_or_debridement_procedures      anesthesia_for_intrathoracic_or_procedures_on_the_thorax
                                                        0                                                      697
                        anesthesia_for_obstetric_procedures                        anesthesia_for_other_procedures
                                                        0                                                      362
                  anesthesia_for_procedures_on_the_abdomen             anesthesia_for_procedures_on_the_head_or_neck
                                                     1780                                                     1361
                   anesthesia_for_procedures_on_the_leg          anesthesia_for_procedures_on_the_perineum_or_pelvis
```

633
anesthesia_for_procedures_on_the_shoulder_or_arm
403
anesthesia_for_radiological_procedures
502
biofeedback_services_and_procedures
87
breast_and_mammography
14319
cardiac_and_pulmonary_rehabilitation_services
31
care_management_evaluation_and_management_services
596
category_iii
1852
chemistry_procedures
11194
chiropractic_manipulative_treatment_procedures
2
colonoscopy_or_anoscopy
18
critical_care_services
11868
cytopathology_procedures
4280
diagnostic_cardiac_doppler_ultrasound
0
diagnostic_radiology_services
1
dialysis_services_and_procedures
2313
drug_assay
6
durable_medical_equipment
8
emergency_department_services
7069
evocative_or_suppression_testing_procedures
21
federally_qualified_health_center_visits
0
fine_needle_aspiration_biopsy_procedures
2972
fracture_treatment
0
genomic_sequencing_procedures_and_other_molecular_multianalyte_assays
5
health_and_behavior_assessment_or_intervention_procedures
28
highly_complex_drug_or_biologic_agent_administration
10014
hospital_inpatient_services
36378
immune_globulins_serum_or_recombinant_products
0
immunology_procedures
2662
initial_telehealth_consultations
4
medical_and_surgical_supplies
221
medicare_coordinated_care_demonstration
15
miscellaneous_diagnostic_and_therapeutic_services
24255
miscellaneous_services
1122
molecular_pathology_procedures
36
neurology_and_neuromuscular_procedures
3749
nuclear_medicine_procedures
23937
office_or_other_outpatient_services
51189
ophthalmology_services_and_procedures
8643
orthotics_and_prosthetics
15
other_evaluation_and_management_services
7
other_pathology_and_laboratory_procedures
340
pathology_and_laboratory_services
221
physical_medicine_and_rehabilitation_evaluations
1215
prolonged_services
2328
psychological_services
0
radiation_oncology_treatment
160
radiologic_guidance
10852
screening_examinations_and_disease_management_training
2811
special_dermatological_procedures
331
special_services_procedures_and_reports
0
surgical_procedures_for_maternity_care_and_delivery
0
surgical_procedures_on_the_cardiovascular_system
21142
surgical_procedures_on_the_endocrine_system
190
surgical_procedures_on_the_female_genital_system
158
surgical_procedures_on_the_integumentary_system
9023
476
anesthesia_for_procedures_on_the_spine_and_spinal_cord
192
application_of_a_modality
623
bone_or_joint_studies
16985
bundled_payment_care
0
cardiovascular_procedures
26790
category_ii
0
central_nervous_system_assessments
488
chemotherapy_drugs
4337
clinical_pathology_consultations
147
counseling_screening_and_prevention_services
4291
cytogenetic_studies
120
dental_services
0
diagnostic_radiology_procedures
42192
diagnostic_ultrasound_procedures
41580
drugs_administered_other_than_oral_method
21904
drug_assay_procedures
89
durable_medical_equipment_offered_only_by_regional_carriers
0
endocrinology_services
280
face_to_face_educational_services
21
filler_procedures
1
followup_telehealth_consultations
1
gastroenterology_procedures
580
gross_and_microscopic_examinations_prostate_biopsy
740
hematology_and_coagulation_procedures
11796
home_services
1903
hospital_observation_services
37591
immunization_administration_for_vaccines_or_toxoids
781
initial_services_for_medicare_enrollment
1168
laboratory_screening_tests
194
medical_nutrition_therapy_procedures
8
microbiology_procedures
2459
miscellaneous_medical_services
4
moderate_conscious_sedation
3238
multianalyte_assay
0
non_invasive_vascular_diagnostic_studies
32400
nursing_facility_services
7078
operating_microscope_procedures
18
organ_or_disease_oriented_panels
4100
osteopathic_manipulative_treatment_procedures
50
other_medicine_services_and_procedures
476
other_services
368
photodynamic_therapy_procedures
285
primary_care_quality_measures
4
psychiatry_services_and_procedures
366
pulmonary_procedures
4616
radiation_therapy_services
56
reproductive_system_procedures
0
sleep_studies_in_home
197
special_otorhinolaryngologic_services_and_procedures
1383
surgical_pathology_procedures
6020
surgical_procedures_on_the_auditory_system
2995
surgical_procedures_on_the_digestive_system
7191
surgical_procedures_on_the_eye_and_ocular_adnexa
7854
surgical_procedures_on_the_hemic_and_lymphatic_systems
2095
surgical_procedures_on_the_male_genital_system
2124

surgical_procedures_on_the_mediastinum_and_diaphragm	14
surgical_procedures_on_the_nervous_system	4679
surgical_procedures_on_the_urinary_system	3934
temporary_national_codes_non_medicare	0
therapeutic_procedures	681
transfusion_medicine_procedures	41
urinalysis_procedures	7904
vaccines_or_toxoids	6440
warfarin_responsiveness_testing	0
anesthesia	2203
medicine_services_and_procedures	87511
radiology_procedures	61536
temporary_procedures_and_professional_services	34612

surgical_procedures_on_the_musculoskeletal_system	7815
surgical_procedures_on_the_respiratory_system	7853
temporary_codes	19136
therapeutic_drug_assays	175
transcutaneous_laboratory_procedures	51
transitional_care_evaluation_and_management_services	1606
vaccine_administration	9856
vision_and_hearing_services	2
ambulatory_services	16393
chemotherapy_drugs_and_drugs_administered_other_than_oral_method	23046
pathology_and_laboratory_procedures	25442
surgery	51357

npi_year	103599
year	103599
nppes_provider_state	103599
number_of_hcpcs	103599
number_of_med_hcpcs	103599
total_med_unique_benes	103599
total_med_medicare_allowed_amt	103599
beneficiary_average_age	103599
beneficiary_age_65_74_count	103599
beneficiary_age_greater_84_count	103590
beneficiary_male_count	103547
beneficiary_race_black_count	102814
beneficiary_race_hispanic_count	101180
beneficiary_nondual_count	103599
beneficiary_cc_afib_percent	103599
beneficiary_cc_asthma_percent	103599
beneficiary_cc_chf_percent	103599
beneficiary_cc_copd_percent	103599
beneficiary_cc_diab_percent	103599
beneficiary_cc_hypert_percent	103599
beneficiary_cc_ost_percent	103599
beneficiary_cc_schiot_percent	103534
beneficiary_average_risk_score	103599
anesthesia	2203
drugs_administered_other_than_oral_method	21904
vaccines_or_toxoids	10137
dialysis_services_and_procedures	2313
ophthalmology_services_and_procedures	8643
cardiovascular_procedures	26790
pulmonary_procedures	4616
endocrinology_services	280
highly_complex_drug_or_biologic_agent_administration	10014
physical_medicine_and_rehabilitation_evaluations	1215
other_medicine_services_and_procedures	807
hospital_observation_services	37591
emergency_department_services	7069
nursing_facility_services	7078
prolonged_services	2328
organ_or_disease_oriented_panels	4100
urinalysis_procedures	7904
hematology_and_coagulation_procedures	11796
microbiology_procedures	2459
surgical_pathology_procedures	

npi	103599
overcharge	103599
provider_type	103599
total_unique_benes	103599
total_med_services	103599
total_med_submitted_chrg_amt	103599
total_med_medicare_payment_amt	103599
beneficiary_age_less_65_count	103583
beneficiary_age_75_84_count	103599
beneficiary_female_count	103599
beneficiary_race_white_count	103561
beneficiary_race_api_count	99068
beneficiary_race_natind_count	23937
beneficiary_dual_count	103597
beneficiary_cc_alzrdsd_percent	103599
beneficiary_cc_cancer_percent	103599
beneficiary_cc_ckd_percent	103599
beneficiary_cc_depr_percent	103599
beneficiary_cc_hyperl_percent	103599
beneficiary_cc_ihd_percent	103599
beneficiary_cc_raoa_percent	103599
beneficiary_cc_strk_percent	103599
ambulatory_services	16393
category_iii	1852
chemotherapy_drugs	4337
psychiatry_services_and_procedures	453
gastroenterology_procedures	580
special_otorhinolaryngologic_services_and_procedures	1383
non_invasive_vascular_diagnostic_studies	32400
allergy_and_clinical_immunology_procedures	693
neurology_and_neuromuscular_procedures	3965
therapeutic_procedures	1137
moderate_conscious_sedation	3238
office_or_other_outpatient_services	51189
hospital_inpatient_services	36378
critical_care_services	11868
home_services	1903
other_care_evaluation_and_management_services	2246
drug_assay	268
chemistry_procedures	11194
immunology_procedures	2662
cytopathology_procedures	4299
diagnostic_radiology_procedures	

```
                                              6020                                               42192
                   diagnostic_ultrasound_procedures                                  radiologic_guidance
                                             41580                                               10852
                          breast_and_mammography                                  bone_or_joint_studies
                                             14319                                               16985
                      radiation_oncology_treatment                           nuclear_medicine_procedures
                                               176                                               23937
            fine_needle_aspiration_biopsy_procedures            surgical_procedures_on_the_auditory_system
                                              2972                                                2995
      surgical_procedures_on_the_cardiovascular_system        surgical_procedures_on_the_digestive_system
                                             21142                                                7191
      surgical_procedures_on_the_eye_and_ocular_adnexa              surgical_procedures_on_genital_system
                                              7854                                                2213
surgical_procedures_on_the_hemic_and_lymphatic_systems    surgical_procedures_on_the_integumentary_system
                                              2095                                                9023
     surgical_procedures_on_the_musculoskeletal_system          surgical_procedures_on_the_nervous_system
                                              7815                                                4679
        surgical_procedures_on_the_respiratory_system          surgical_procedures_on_the_urinary_system
                                              7853                                                3934
                        other_surgical_procedures screening_examinations_and_disease_management_training
                                               219                                                2811
     miscellaneous_diagnostic_and_therapeutic_services           initial_services_for_medicare_enrollment
                                             24255                                                1168
    gross_and_microscopic_examinations_prostate_biopsy         counseling_screening_and_prevention_services
                                               740                                                4291
                           miscellaneous_services
                                             20967
```

Some provider types are identical and can be renamed to have fewer categories. Some provider types are unknown and can be dropped from the dataset. Some only contain 1 observation and can be dropped too.

```
levels(d$provider_type)
d$provider_type <- gsub('allergy/ immunology', 'allergy/immunology', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('audiologist (billing independently)', 'audiologist', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('cardiovascular disease (cardiology)', 'cardiology', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('clinical cardiatric electrophysiology', 'cardiac electrophysiology', d$provider_type,
fixed = TRUE)
d$provider_type <- gsub('clinical cardiac electrophysiology', 'cardiac electrophysiology', d$provider_type,
fixed = TRUE)
d$provider_type <- gsub('colorectal surgery (formerly proctology)', 'colorectal surgery (proctology)',
d$provider_type, fixed = TRUE)
d$provider_type <- gsub('gynecological oncology', 'gynecology/oncology', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('hematology-oncology', 'hematology/oncology', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('independent diagnostic testing facility (idtf)', 'independent diagnostic testing
facility', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('obstetrics & gynecology', 'obstetrics/gynecology', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('occupational therapist in private practice', 'occupational therapist',
d$provider_type, fixed = TRUE)
d$provider_type <- gsub('oral surgery (dentists only)', 'oral surgery (dentist only)', d$provider_type, fixed =
TRUE)
d$provider_type <- gsub('physical therapist in private practice', 'physical therapist', d$provider_type, fixed
= TRUE)
d$provider_type <- gsub('psychologist, clinical', 'clinical psychologist', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('interventional pain management', 'pain management', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('interventional cardiology', 'cardiology', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('registered dietitian or nutrition professional', 'registered dietician/nutrition
professional', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('undefined physician type', '', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('unknown physician specialty code', '', d$provider_type, fixed = TRUE)
d$provider_type <- gsub('unknown supplier/provider', '', d$provider_type, fixed = TRUE)
# Remove unknown provider types
d$provider_type <- as.character(d$provider_type)
d$provider_type[d$provider_type==''] <- NA
d <- d[complete.cases(d),]
d$provider_type <- as.factor(d$provider_type)
# Remove provider types that only appear once
gr <- d %>% group_by(provider_type) %>% summarise(no_rows = length(provider_type))
gr[which(gr$no_rows == 1),]
d <- d[which(d$provider_type %notin% c('addiction medicine', 'certified nurse midwife', 'clinic or group
practice', 'clinical laboratory', 'gynecology/oncology', 'multispecialty clinic/group practice', 'portable x-
ray')), ]
d <- droplevels(d)
```

```
[1] "addiction medicine"                           "allergy/ immunology"
 [3] "allergy/immunology"                           "anesthesiologist assistants"
 [5] "anesthesiology"                               "audiologist"
 [7] "audiologist (billing independently)"          "cardiac electrophysiology"
 [9] "cardiac surgery"                              "cardiology"
[11] "cardiovascular disease (cardiology)"          "certified clinical nurse specialist"
[13] "certified nurse midwife"                      "certified registered nurse anesthetist (crna)"
[15] "chiropractic"                                 "clinic or group practice"
[17] "clinical cardiac electrophysiology"           "clinical cardiatric electrophysiology"
[19] "clinical laboratory"                          "clinical psychologist"
[21] "colorectal surgery (formerly proctology)"     "colorectal surgery (proctology)"
[23] "critical care (intensivists)"                 "crna"
[25] "dermatology"                                  "diagnostic radiology"
[27] "emergency medicine"                           "endocrinology"
[29] "family practice"                              "gastroenterology"
[31] "general practice"                             "general surgery"
[33] "geriatric medicine"                           "geriatric psychiatry"
[35] "gynecological oncology"                       "gynecological/oncology"
```

```
[37] "hand surgery"                                  "hematology"
[39] "hematology-oncology"                           "hematology/oncology"
[41] "hospice and palliative care"                   "hospitalist"
[43] "independent diagnostic testing facility"       "independent diagnostic testing facility (idtf)"
[45] "infectious disease"                            "internal medicine"
[47] "interventional cardiology"                     "interventional pain management"
[49] "interventional radiology"                      "licensed clinical social worker"
[51] "maxillofacial surgery"                         "medical oncology"
[53] "multispecialty clinic/group practice"          "nephrology"
[55] "neurology"                                     "neuropsychiatry"
[57] "neurosurgery"                                  "nuclear medicine"
[59] "nurse practitioner"                            "obstetrics & gynecology"
[61] "obstetrics/gynecology"                         "occupational therapist"
[63] "occupational therapist in private practice"    "ophthalmology"
[65] "optometry"                                     "oral surgery (dentist only)"
[67] "oral surgery (dentists only)"                  "orthopedic surgery"
[69] "osteopathic manipulative medicine"             "otolaryngology"
[71] "pain management"                               "pathology"
[73] "pediatric medicine"                            "peripheral vascular disease"
[75] "physical medicine and rehabilitation"          "physical therapist"
[77] "physical therapist in private practice"        "physician assistant"
[79] "plastic and reconstructive surgery"            "podiatry"
[81] "portable x-ray"                                "preventive medicine"
[83] "psychiatry"                                    "psychologist, clinical"
[85] "pulmonary disease"                             "radiation oncology"
[87] "registered dietician/nutrition professional"   "registered dietitian or nutrition professional"
[89] "rheumatology"                                  "sleep medicine"
[91] "sports medicine"                               "surgical oncology"
[93] "thoracic surgery"                              "undefined physician type"
[95] "unknown physician specialty code"              "unknown supplier/provider"
[97] "urology"                                       "vascular surgery"
```

provider_type	No_rows
<fctr>	<int>
addiction medicine	1
certified nurse midwife	1
clinic or group practice	1
clinical laboratory	1
gynecology/oncology	1
multispecialty clinic/group practice	1
portable x-ray	1

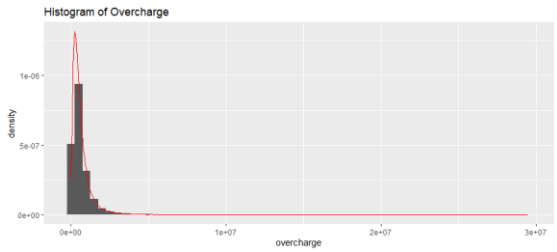
Check NAs, remove the row with negative overcharge

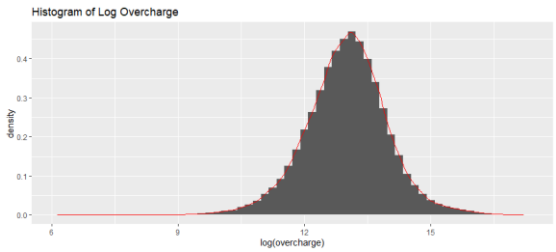
```
colSums(is.na(d))
d <- d[complete.cases(d),]
d[which(d$overcharge < 0),]
d <- d[which(d$overcharge > 0),]
colSums(d != 0)

ggplot(d, aes(x = overcharge)) +
  ggtitle("Histogram of Overcharge") +
  geom_histogram(aes(y = ..density..), bins = 60) +
  geom_density(color="red")

ggplot(d, aes(x = log(overcharge))) +
  ggtitle("Histogram of Log Overcharge") +
  geom_histogram(aes(y = ..density..), bins = 60) +
  geom_density(color="red")
```

```
                                    npi_year                                          npi
                                           0                                            0
                                        year                                   overcharge
                                           0                                            0
                           nppes_provider_state                                provider_type
                                           0                                            0
                                number_of_hcpcs                           total_unique_benes
                                           0                                            0
                             number_of_med_hcpcs                           total_med_services
                                           0                                            0
                          total_med_unique_benes                  total_med_submitted_chrg_amt
                                           0                                            0
                  total_med_medicare_allowed_amt                total_med_medicare_payment_amt
                                           0                                            0
                         beneficiary_average_age                 beneficiary_age_less_65_count
                                           0                                            0
                    beneficiary_age_65_74_count                    beneficiary_age_75_84_count
                                           0                                            0
                beneficiary_age_greater_84_count                       beneficiary_female_count
                                           0                                            0
                         beneficiary_male_count                  beneficiary_race_white_count
                                           0                                            0
                    beneficiary_race_black_count                     beneficiary_race_api_count
                                           0                                            0
                  beneficiary_race_hispanic_count                 beneficiary_race_natind_count
                                           0                                            0
                       beneficiary_nondual_count                        beneficiary_dual_count
                                           0                                            0
                    beneficiary_cc_afib_percent                beneficiary_cc_alzrdsd_percent
                                           0                                            0
                   beneficiary_cc_asthma_percent                beneficiary_cc_cancer_percent
                                           0                                            0
                      beneficiary_cc_chf_percent                   beneficiary_cc_ckd_percent
                                           0                                            0
                     beneficiary_cc_copd_percent                  beneficiary_cc_depr_percent
```

0	0
beneficiary_cc_diab_percent	beneficiary_cc_hyperl_percent
0	0
beneficiary_cc_hypert_percent	beneficiary_cc_ihd_percent
0	0
beneficiary_cc_ost_percent	beneficiary_cc_raoa_percent
0	0
beneficiary_cc_schiot_percent	beneficiary_cc_strk_percent
0	0
beneficiary_average_risk_score	ambulatory_services
0	0
anesthesia	category_iii
0	0
drugs_administered_other_than_oral_method	chemotherapy_drugs
0	0
vaccines_or_toxoids	psychiatry_services_and_procedures
0	0
dialysis_services_and_procedures	gastroenterology_procedures
0	0
ophthalmology_services_and_procedures	special_otorhinolaryngologic_services_and_procedures
0	0
cardiovascular_procedures	non_invasive_vascular_diagnostic_studies
0	0
pulmonary_procedures	allergy_and_clinical_immunology_procedures
0	0
endocrinology_services	neurology_and_neuromuscular_procedures
0	0
highly_complex_drug_or_biologic_agent_administration	therapeutic_procedures
0	0
physical_medicine_and_rehabilitation_evaluations	moderate_conscious_sedation
0	0
other_medicine_services_and_procedures	office_or_other_outpatient_services
0	0
hospital_observation_services	hospital_inpatient_services
0	0
emergency_department_services	critical_care_services
0	0
nursing_facility_services	home_services
0	0
prolonged_services	other_care_evaluation_and_management_services
0	0
organ_or_disease_oriented_panels	drug_assay
0	0
urinalysis_procedures	chemistry_procedures
0	0
hematology_and_coagulation_procedures	immunology_procedures
0	0
microbiology_procedures	cytopathology_procedures
0	0
surgical_pathology_procedures	diagnostic_radiology_procedures
0	0
diagnostic_ultrasound_procedures	radiologic_guidance
0	0
breast_and_mammography	bone_or_joint_studies
0	0
radiation_oncology_treatment	nuclear_medicine_procedures
0	0
fine_needle_aspiration_biopsy_procedures	surgical_procedures_on_the_auditory_system
0	0
surgical_procedures_on_the_cardiovascular_system	surgical_procedures_on_the_digestive_system
0	0
surgical_procedures_on_the_eye_and_ocular_adnexa	surgical_procedures_on_genital_system
0	0
surgical_procedures_on_the_hemic_and_lymphatic_systems	surgical_procedures_on_the_integumentary_system
0	0
surgical_procedures_on_the_musculoskeletal_system	surgical_procedures_on_the_nervous_system
0	0
surgical_procedures_on_the_respiratory_system	surgical_procedures_on_the_urinary_system
0	0
other_surgical_procedures	screening_examinations_and_disease_management_training
0	0
screening_examinations_and_disease_management_training.1	miscellaneous_diagnostic_and_therapeutic_services
0	0
initial_services_for_medicare_enrollment	gross_and_microscopic_examinations_prostate_biopsy
0	0
counseling_screening_and_prevention_services	miscellaneous_services
0	0
npi_year	npi
103549	103549
year	overcharge
103549	103549
nppes_provider_state	provider_type
103549	103549
number_of_hcpcs	total_unique_benes
103549	103549
number_of_med_hcpcs	total_med_services
103549	103549
total_med_unique_benes	total_med_submitted_chrg_amt
103549	103549
total_med_medicare_allowed_amt	total_med_medicare_payment_amt
103549	103549
beneficiary_average_age	beneficiary_age_less_65_count
103549	103533
beneficiary_age_65_74_count	beneficiary_age_75_84_count
103549	103549
beneficiary_age_greater_84_count	beneficiary_female_count
103540	103549
beneficiary_male_count	beneficiary_race_white_count
103498	103511
beneficiary_race_black_count	beneficiary_race_api_count
102764	99023

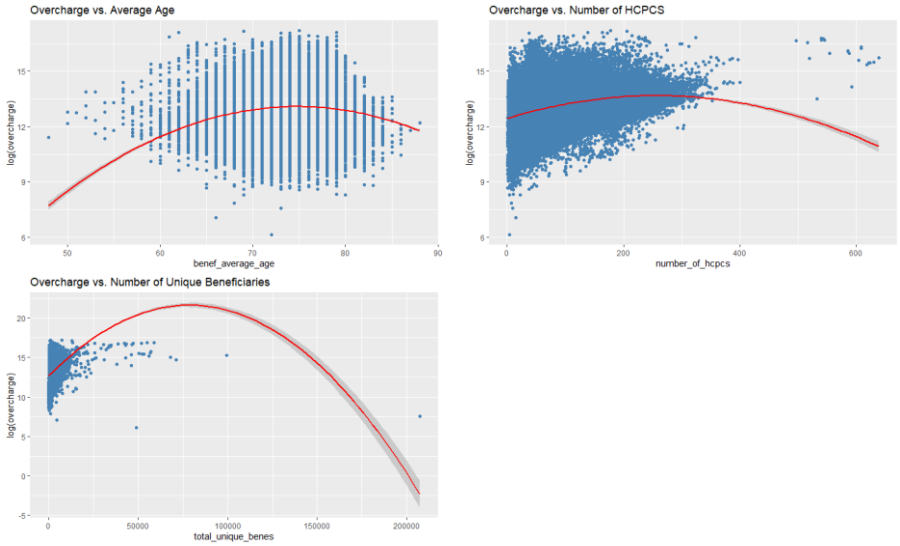
beneficiary_race_hispanic_count	101133
beneficiary_race_natind_count	23918
beneficiary_nondual_count	103549
beneficiary_dual_count	103547
beneficiary_cc_afib_percent	103549
beneficiary_cc_alzrdsd_percent	103549
beneficiary_cc_asthma_percent	103549
beneficiary_cc_cancer_percent	103549
beneficiary_cc_chf_percent	103549
beneficiary_cc_ckd_percent	103549
beneficiary_cc_copd_percent	103549
beneficiary_cc_depr_percent	103549
beneficiary_cc_diab_percent	103549
beneficiary_cc_hyperl_percent	103549
beneficiary_cc_hypert_percent	103549
beneficiary_cc_ihd_percent	103549
beneficiary_cc_ost_percent	103549
beneficiary_cc_raoa_percent	103549
beneficiary_cc_schiot_percent	103484
beneficiary_cc_strk_percent	103549
beneficiary_average_risk_score	103549
ambulatory_services	16388
anesthesia	2203
category_iii	1851
drugs_administered_other_than_oral_method	21896
chemotherapy_drugs	4337
vaccines_or_toxoids	10136
psychiatry_services_and_procedures	453
dialysis_services_and_procedures	2313
gastroenterology_procedures	580
ophthalmology_services_and_procedures	8641
special_otorhinolaryngologic_services_and_procedures	1383
cardiovascular_procedures	26763
non_invasive_vascular_diagnostic_studies	32380
pulmonary_procedures	4614
allergy_and_clinical_immunology_procedures	693
endocrinology_services	280
neurology_and_neuromuscular_procedures	3964
highly_complex_drug_or_biologic_agent_administration	10011
therapeutic_procedures	1137
physical_medicine_and_rehabilitation_evaluations	1214
moderate_conscious_sedation	3235
other_medicine_services_and_procedures	806
office_or_other_outpatient_services	51161
hospital_observation_services	37565
hospital_inpatient_services	36351
emergency_department_services	7067
critical_care_services	11865
nursing_facility_services	7077
home_services	1902
prolonged_services	2325
other_care_evaluation_and_management_services	2246
organ_or_disease_oriented_panels	4097
drug_assay	267
urinalysis_procedures	7902
chemistry_procedures	11191
hematology_and_coagulation_procedures	11780
immunology_procedures	2661
microbiology_procedures	2457
cytopathology_procedures	4299
surgical_pathology_procedures	6018
diagnostic_radiology_procedures	42171
diagnostic_ultrasound_procedures	41563
radiologic_guidance	10850
breast_and_mammography	14315
bone_or_joint_studies	16983
radiation_oncology_treatment	176
nuclear_medicine_procedures	23924
fine_needle_aspiration_biopsy_procedures	2971
surgical_procedures_on_the_auditory_system	2995
surgical_procedures_on_the_cardiovascular_system	21130
surgical_procedures_on_the_digestive_system	7190
surgical_procedures_on_the_eye_and_ocular_adnexa	7853
surgical_procedures_on_genital_system	2212
surgical_procedures_on_the_hemic_and_lymphatic_systems	2095
surgical_procedures_on_the_integumentary_system	9019
surgical_procedures_on_the_musculoskeletal_system	7811
surgical_procedures_on_the_nervous_system	4679
surgical_procedures_on_the_respiratory_system	7851
surgical_procedures_on_the_urinary_system	3934
other_surgical_procedures	219
screening_examinations_and_disease_management_training	2810
screening_examinations_and_disease_management_training.1	2810
miscellaneous_diagnostic_and_therapeutic_services	24248
initial_services_for_medicare_enrollment	1168
gross_and_microscopic_examinations_prostate_biopsy	740
counseling_screening_and_prevention_services	4291
miscellaneous_services	20962





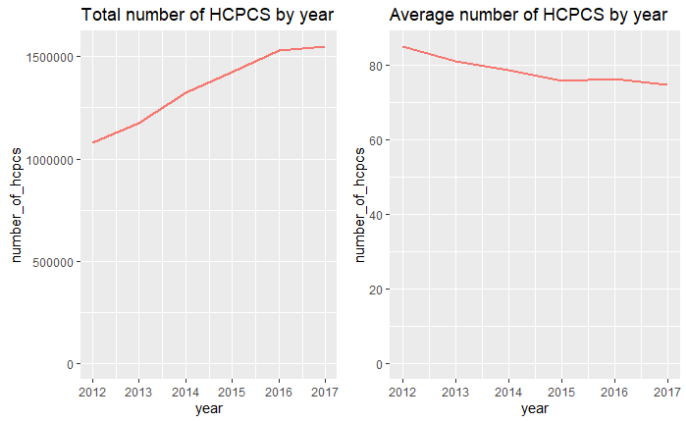
Plot overcharge vs. continuous variables

```
ggplot(d_model, aes(x=benef_average_age, y=log(overcharge))) +
  geom_point(color= "steelblue") +
  ggtitle("Overcharge vs. Average Age") +
  geom_smooth(method="lm", formula = y ~ poly(x, 2), color="red")

ggplot(d_model, aes(x=number_of_hcpcs, y=log(overcharge))) +
  geom_point(color= "steelblue") +
  ggtitle("Overcharge vs. Number of HCPCS") +
  geom_smooth(method="lm", formula = y ~ poly(x, 2), color="red")

ggplot(d_model, aes(x=total_unique_benes, y=log(overcharge))) +
  geom_point(color= "steelblue") +
  ggtitle("Overcharge vs. Number of Unique Beneficiaries") +
  geom_smooth(method="lm", formula = y ~ poly(x, 2), color="red")
```



Plot the number of procedures per year

```
g1 <- d[c("year", "number_of_hcpcs")] %>%
  group_by(year) %>%
  summarise(number_of_hcpcs = sum(number_of_hcpcs)) %>%
  ggplot(aes(year, number_of_hcpcs)) + ggtitle("Total number of HCPCS by year") +
  geom_line(aes(year, number_of_hcpcs), col="#F8766D", size=1) + ylim(0, NA)
g2 <- d[c("year", "number_of_hcpcs")] %>%
  group_by(year) %>%
  summarise(number_of_hcpcs = mean(number_of_hcpcs)) %>%
  ggplot(aes(year, number_of_hcpcs)) + ggtitle("Average number of HCPCS by year") +
  geom_line(aes(year, number_of_hcpcs), col="#F8766D", size=1) + ylim(0, NA)
grid.arrange(g1, g2, nrow=1)
```



Select variables for the models and do final transformations. Convert everything to percentages instead of counts

```
d_model <- d[ , !(names(d) %in% c("npi_year", "npi", "medicare_participation_indicator",
"total_med_submitted_chrg_amt", "total_med_medicare_allowed_amt", "total_med_medicare_payment_amt"))]
d_model$year <- as.factor(d_model$year)
# Divide percentages by 100
d_model$beneficiary_cc_afib_percent <- d_model$beneficiary_cc_afib_percent/100
d_model$beneficiary_cc_alzrdsd_percent <- d_model$beneficiary_cc_alzrdsd_percent/100
d_model$beneficiary_cc_asthma_percent <- d_model$beneficiary_cc_asthma_percent/100
d_model$beneficiary_cc_cancer_percent <- d_model$beneficiary_cc_cancer_percent/100
d_model$beneficiary_cc_chf_percent <- d_model$beneficiary_cc_chf_percent/100
d_model$beneficiary_cc_ckd_percent <- d_model$beneficiary_cc_ckd_percent/100
d_model$beneficiary_cc_copd_percent <- d_model$beneficiary_cc_copd_percent/100
d_model$beneficiary_cc_depr_percent <- d_model$beneficiary_cc_depr_percent/100
d_model$beneficiary_cc_diab_percent <- d_model$beneficiary_cc_diab_percent/100
d_model$beneficiary_cc_hyperl_percent <- d_model$beneficiary_cc_hyperl_percent/100
d_model$beneficiary_cc_hypert_percent <- d_model$beneficiary_cc_hypert_percent/100
d_model$beneficiary_cc_ihd_percent <- d_model$beneficiary_cc_ihd_percent/100
d_model$beneficiary_cc_ost_percent <- d_model$beneficiary_cc_ost_percent/100
d_model$beneficiary_cc_raoa_percent <- d_model$beneficiary_cc_raoa_percent/100
d_model$beneficiary_cc_schiot_percent <- d_model$beneficiary_cc_schiot_percent/100
d_model$beneficiary_cc_strk_percent <- d_model$beneficiary_cc_strk_percent/100
# Divide beneficiary counts by the number of beneficiaries
d_model$beneficiary_age_less_65_count <- d_model$beneficiary_age_less_65_count/d_model$total_unique_benes
d_model$beneficiary_age_65_74_count <- d_model$beneficiary_age_65_74_count/d_model$total_unique_benes
d_model$beneficiary_age_75_84_count <- d_model$beneficiary_age_75_84_count/d_model$total_unique_benes
d_model$beneficiary_age_greater_84_count <- d_model$beneficiary_age_greater_84_count/d_model$total_unique_benes
d_model$beneficiary_female_count <- d_model$beneficiary_female_count/d_model$total_unique_benes
d_model$beneficiary_male_count <- d_model$beneficiary_male_count/d_model$total_unique_benes
d_model$beneficiary_race_white_count <- d_model$beneficiary_race_white_count/d_model$total_unique_benes
d_model$beneficiary_race_black_count <- d_model$beneficiary_race_black_count/d_model$total_unique_benes
d_model$beneficiary_race_api_count <- d_model$beneficiary_race_api_count/d_model$total_unique_benes
d_model$beneficiary_race_hispanic_count <- d_model$beneficiary_race_hispanic_count/d_model$total_unique_benes
d_model$beneficiary_race_natind_count <- d_model$beneficiary_race_natind_count/d_model$total_unique_benes
d_model$beneficiary_dual_count <- d_model$beneficiary_dual_count/d_model$total_unique_benes
d_model$beneficiary_nondual_count <- d_model$beneficiary_nondual_count/d_model$total_unique_benes
# Divide procedure counts by the number of procedures
d_model$ambulatory_services <- d_model$ambulatory_services/d_model$number_of_hcpcs
d_model$anesthesia <- d_model$anesthesia/d_model$number_of_hcpcs
d_model$category_iii <- d_model$category_iii/d_model$number_of_hcpcs
d_model$drugs_administered_other_than_oral_method <-
d_model$drugs_administered_other_than_oral_method/d_model$number_of_hcpcs
d_model$chemotherapy_drugs <- d_model$chemotherapy_drugs/d_model$number_of_hcpcs
d_model$vaccines_or_toxoids <- d_model$vaccines_or_toxoids/d_model$number_of_hcpcs
d_model$psychiatry_services_and_procedures <-
d_model$psychiatry_services_and_procedures/d_model$number_of_hcpcs
d_model$dialysis_services_and_procedures <- d_model$dialysis_services_and_procedures/d_model$number_of_hcpcs
d_model$gastroenterology_procedures <- d_model$gastroenterology_procedures/d_model$number_of_hcpcs
d_model$ophthalmology_services_and_procedures <-
d_model$ophthalmology_services_and_procedures/d_model$number_of_hcpcs
d_model$special_otorhinolaryngologic_services_and_procedures <-
d_model$special_otorhinolaryngologic_services_and_procedures/d_model$number_of_hcpcs
d_model$cardiovascular_procedures <- d_model$cardiovascular_procedures/d_model$number_of_hcpcs
d_model$non_invasive_vascular_diagnostic_studies <-
d_model$non_invasive_vascular_diagnostic_studies/d_model$number_of_hcpcs
d_model$pulmonary_procedures <- d_model$pulmonary_procedures/d_model$number_of_hcpcs
d_model$allergy_and_clinical_immunology_procedures <-
d_model$allergy_and_clinical_immunology_procedures/d_model$number_of_hcpcs
d_model$endocrinology_services <- d_model$endocrinology_services/d_model$number_of_hcpcs
d_model$neurology_and_neuromuscular_procedures <-
d_model$neurology_and_neuromuscular_procedures/d_model$number_of_hcpcs
d_model$highly_complex_drug_or_biologic_agent_administration <-
d_model$highly_complex_drug_or_biologic_agent_administration/d_model$number_of_hcpcs
d_model$therapeutic_procedures <- d_model$therapeutic_procedures/d_model$number_of_hcpcs
d_model$physical_medicine_and_rehabilitation_evaluations <-
d_model$physical_medicine_and_rehabilitation_evaluations/d_model$number_of_hcpcs
d_model$moderate_conscious_sedation <- d_model$moderate_conscious_sedation/d_model$number_of_hcpcs
d_model$other_medicine_services_and_procedures <-
d_model$other_medicine_services_and_procedures/d_model$number_of_hcpcs
d_model$office_or_other_outpatient_services <-
d_model$office_or_other_outpatient_services/d_model$number_of_hcpcs
d_model$hospital_observation_services <- d_model$hospital_observation_services/d_model$number_of_hcpcs
d_model$hospital_inpatient_services <- d_model$hospital_inpatient_services/d_model$number_of_hcpcs
d_model$emergency_department_services <- d_model$emergency_department_services/d_model$number_of_hcpcs
d_model$critical_care_services <- d_model$critical_care_services/d_model$number_of_hcpcs
d_model$nursing_facility_services <- d_model$nursing_facility_services/d_model$number_of_hcpcs
d_model$home_services <- d_model$home_services/d_model$number_of_hcpcs
d_model$prolonged_services <- d_model$prolonged_services/d_model$number_of_hcpcs
d_model$other_care_evaluation_and_management_services <-
d_model$other_care_evaluation_and_management_services/d_model$number_of_hcpcs
d_model$organ_or_disease_oriented_panels <- d_model$organ_or_disease_oriented_panels/d_model$number_of_hcpcs
d_model$drug_assay <- d_model$drug_assay/d_model$number_of_hcpcs
d_model$urinalysis_procedures <- d_model$urinalysis_procedures/d_model$number_of_hcpcs
```

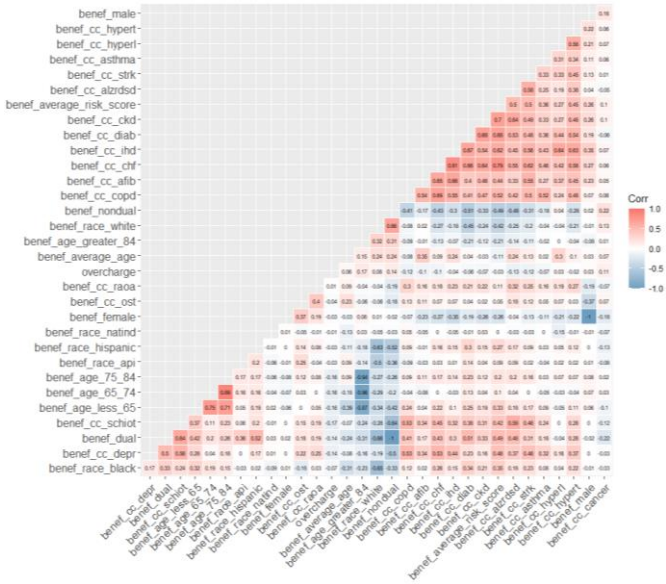
```
d_model$chemistry_procedures <- d_model$chemistry_procedures/d_model$number_of_hcpcs
d_model$hematology_and_coagulation_procedures <-
d_model$hematology_and_coagulation_procedures/d_model$number_of_hcpcs
d_model$immunology_procedures <- d_model$immunology_procedures/d_model$number_of_hcpcs
d_model$microbiology_procedures <- d_model$microbiology_procedures/d_model$number_of_hcpcs
d_model$cytopathology_procedures <- d_model$cytopathology_procedures/d_model$number_of_hcpcs
d_model$surgical_pathology_procedures <- d_model$surgical_pathology_procedures/d_model$number_of_hcpcs
d_model$diagnostic_radiology_procedures <- d_model$diagnostic_radiology_procedures/d_model$number_of_hcpcs
d_model$diagnostic_ultrasound_procedures <- d_model$diagnostic_ultrasound_procedures/d_model$number_of_hcpcs
d_model$radiologic_guidance <- d_model$radiologic_guidance/d_model$number_of_hcpcs
d_model$breast_and_mammography <- d_model$breast_and_mammography/d_model$number_of_hcpcs
d_model$bone_or_joint_studies <- d_model$bone_or_joint_studies/d_model$number_of_hcpcs
d_model$radiation_oncology_treatment <- d_model$radiation_oncology_treatment/d_model$number_of_hcpcs
d_model$nuclear_medicine_procedures <- d_model$nuclear_medicine_procedures/d_model$number_of_hcpcs
d_model$fine_needle_aspiration_biopsy_procedures <-
d_model$fine_needle_aspiration_biopsy_procedures/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_auditory_system <-
d_model$surgical_procedures_on_the_auditory_system/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_cardiovascular_system <-
d_model$surgical_procedures_on_the_cardiovascular_system/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_digestive_system <-
d_model$surgical_procedures_on_the_digestive_system/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_eye_and_ocular_adnexa <-
d_model$surgical_procedures_on_the_eye_and_ocular_adnexa/d_model$number_of_hcpcs
d_model$surgical_procedures_on_genital_system <-
d_model$surgical_procedures_on_genital_system/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_hemic_and_lymphatic_systems <-
d_model$surgical_procedures_on_the_hemic_and_lymphatic_systems/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_integumentary_system <-
d_model$surgical_procedures_on_the_integumentary_system/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_musculoskeletal_system <-
d_model$surgical_procedures_on_the_musculoskeletal_system/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_nervous_system <-
d_model$surgical_procedures_on_the_nervous_system/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_respiratory_system <-
d_model$surgical_procedures_on_the_respiratory_system/d_model$number_of_hcpcs
d_model$surgical_procedures_on_the_urinary_system <-
d_model$surgical_procedures_on_the_urinary_system/d_model$number_of_hcpcs
d_model$other_surgical_procedures <- d_model$other_surgical_procedures/d_model$number_of_hcpcs
d_model$screening_examinations_and_disease_management_training <-
d_model$screening_examinations_and_disease_management_training/d_model$number_of_hcpcs
d_model$miscellaneous_diagnostic_and_therapeutic_services <-
d_model$miscellaneous_diagnostic_and_therapeutic_services/d_model$number_of_hcpcs
d_model$initial_services_for_medicare_enrollment <-
d_model$initial_services_for_medicare_enrollment/d_model$number_of_hcpcs
d_model$gross_and_microscopic_examinations_prostate_biopsy <-
d_model$gross_and_microscopic_examinations_prostate_biopsy/d_model$number_of_hcpcs
d_model$counseling_screening_and_prevention_services <-
d_model$counseling_screening_and_prevention_services/d_model$number_of_hcpcs
d_model$miscellaneous_services <- d_model$miscellaneous_services/d_model$number_of_hcpcs
# Rename columns to make them shorter
colnames(d_model) <- gsub("_count", "", colnames(d_model), fixed = TRUE)
colnames(d_model) <- gsub("_percent", "", colnames(d_model), fixed = TRUE)
colnames(d_model) <- gsub("beneficiary", "benef", colnames(d_model), fixed = TRUE)
```

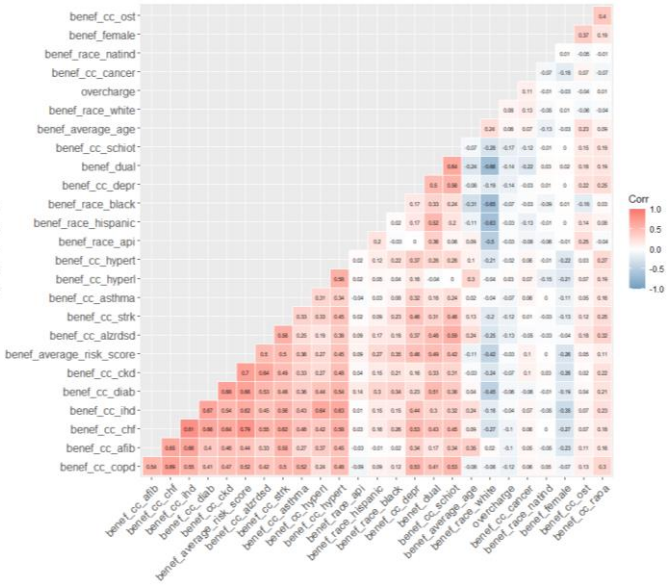
Find correlated / collinear variables

```
nums <- unlist(lapply(d_model, is.numeric))
summary(d_model[, nums])

plot = cor(d_model[, c("overcharge", "benef_average_age", "benef_age_less_65", "benef_age_65_74",
"benef_age_75_84", "benef_age_greater_84", "benef_male",  "benef_female", "benef_race_white",
"benef_race_black", "benef_race_api", "benef_race_hispanic", "benef_race_natind", "benef_dual",
"benef_nondual", "benef_cc_afib", "benef_cc_alzrdsd", "benef_cc_asthma", "benef_cc_cancer", "benef_cc_chf",
"benef_cc_ckd", "benef_cc_copd", "benef_cc_depr", "benef_cc_diab", "benef_cc_hyperl", "benef_cc_hypert",
"benef_cc_ihd", "benef_cc_ost", "benef_cc_raoa", "benef_cc_schiot", "benef_cc_strk",
"benef_average_risk_score")])
ggcorrplot(plot, hc.order = TRUE, type = "lower", outline.col = "white", ggtheme = ggplot2::theme_gray, colors
= c("#6D9EC1", "white", "#F8766D"), lab = TRUE,  lab_size = 2)

plot = cor(d_model[, c("overcharge", "benef_average_age",  "benef_female", "benef_race_white",
"benef_race_black", "benef_race_api", "benef_race_hispanic", "benef_race_natind", "benef_dual",
"benef_cc_afib", "benef_cc_alzrdsd", "benef_cc_asthma", "benef_cc_cancer", "benef_cc_chf", "benef_cc_ckd",
"benef_cc_copd", "benef_cc_depr", "benef_cc_diab", "benef_cc_hyperl", "benef_cc_hypert", "benef_cc_ihd",
"benef_cc_ost", "benef_cc_raoa", "benef_cc_schiot", "benef_cc_strk", "benef_average_risk_score")])
ggcorrplot(plot, hc.order = TRUE, type = "lower", outline.col = "white", ggtheme = ggplot2::theme_gray, colors
= c("#6D9EC1", "white", "#F8766D"), lab = TRUE,  lab_size = 2)

lincomb <- caret::findLinearCombos(d_model[, nums])
lapply(lincomb$linearCombos, function(x) colnames(d_model)[x])
```

```
  overcharge       number_of_hcpcs  total_unique_benes number_of_med_hcpcs total_med_services total_med_unique_benes benef_average_age
Min.   :     461   Min.   :  1.00   Min.   :    66     Min.   :  1.00      Min.   :     75    Min.   :    66         Min.   :48.00
1st Qu.:  247103   1st Qu.: 31.00   1st Qu.:   668     1st Qu.: 30.00      1st Qu.:   1793    1st Qu.:   667         1st Qu.:72.00
Median :  455920   Median : 55.00   Median :  1115     Median : 51.00      Median :   3180    Median :  1115         Median :74.00
Mean   :  719831   Mean   : 78.03   Mean   :  1562     Mean   : 75.54      Mean   :   4246    Mean   :  1562         Mean   :73.53
3rd Qu.:  807023   3rd Qu.:113.00   3rd Qu.:  2028     3rd Qu.:108.00      3rd Qu.:   5106    3rd Qu.:  2028         3rd Qu.:75.00
Max.   :29448020   Max.   :639.00   Max.   :207271     Max.   :639.00      Max.   :2931825    Max.   :207271         Max.   :88.00
benef_age_less_65 benef_age_65_74  benef_age_75_84   benef_age_greater_84  benef_female       benef_male     benef_race_white
Min.   :0.00000   Min.   :0.04384  Min.   :0.03275   Min.   :0.0000       Min.   :0.02163   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.08537   1st Qu.:0.32173  1st Qu.:0.27494   1st Qu.:0.1234       1st Qu.:0.52632   1st Qu.:0.3882   1st Qu.:0.6730
Median :0.13395   Median :0.37396  Median :0.30375   Median :0.1643       Median :0.57143   Median :0.4286   Median :0.7940
Mean   :0.14524   Mean   :0.37581  Mean   :0.30354   Mean   :0.1754       Mean   :0.56848   Mean   :0.4315   Mean   :0.7466
3rd Qu.:0.18933   3rd Qu.:0.42809  3rd Qu.:0.33418   3rd Qu.:0.2153       3rd Qu.:0.61181   3rd Qu.:0.4737   3rd Qu.:0.8733
Max.   :0.89863   Max.   :0.73313  Max.   :0.69431   Max.   :0.7616       Max.   :1.00000   Max.   :0.9784   Max.   :1.0000
benef_race_black  benef_race_api    benef_race_hispanic benef_race_natind  benef_nondual        benef_dual     benef_cc_afib
Min.   :0.00000   Min.   :0.00000   Min.   :0.00000     Min.   :0.000000   Min.   :0.006869   Min.   :0.0000   Min.   :0.0200
1st Qu.:0.03313   1st Qu.:0.01120   1st Qu.:0.02332     1st Qu.:0.000000   1st Qu.:0.654344   1st Qu.:0.1430   1st Qu.:0.1200
Median :0.07164   Median :0.02176   Median :0.04581     Median :0.000000   Median :0.773398   Median :0.2266   Median :0.1700
Mean   :0.10954   Mean   :0.04126   Mean   :0.07609     Mean   :0.004998   Mean   :0.730972   Mean   :0.2690   Mean   :0.1827
3rd Qu.:0.14061   3rd Qu.:0.04570   3rd Qu.:0.09062     3rd Qu.:0.000000   3rd Qu.:0.856974   3rd Qu.:0.3457   3rd Qu.:0.2300
Max.   :0.96736   Max.   :0.96544   Max.   :1.00000     Max.   :1.000000   Max.   :1.000000   Max.   :0.9931   Max.   :0.7500
benef_cc_alzrdsd benef_cc_asthma  benef_cc_cancer   benef_cc_chf     benef_cc_ckd    benef_cc_copd    benef_cc_depr    benef_cc_diab
Min.   :0.0100   Min.   :0.0200   Min.   :0.0200   Min.   :0.0200   Min.   :0.0200   Min.   :0.0200   Min.   :0.0300   Min.   :0.0200
1st Qu.:0.1000   1st Qu.:0.1000   1st Qu.:0.1200   1st Qu.:0.2100   1st Qu.:0.2400   1st Qu.:0.1600   1st Qu.:0.2300   1st Qu.:0.3000
Median :0.1400   Median :0.1400   Median :0.1400   Median :0.3100   Median :0.3400   Median :0.2300   Median :0.3000   Median :0.3800
Mean   :0.1676   Mean   :0.1671   Mean   :0.1526   Mean   :0.3195   Mean   :0.3534   Mean   :0.2344   Mean   :0.3125   Mean   :0.3837
3rd Qu.:0.2100   3rd Qu.:0.2100   3rd Qu.:0.1700   3rd Qu.:0.4100   3rd Qu.:0.4500   3rd Qu.:0.2900   3rd Qu.:0.3800   3rd Qu.:0.4500
Max.   :0.7500   Max.   :0.7500   Max.   :0.7500   Max.   :0.7500   Max.   :0.7500   Max.   :0.7500   Max.   :0.7500   Max.   :0.7500
benef_cc_hyperl  benef_cc_hypert   benef_cc_ihd     benef_cc_ost    benef_cc_raoa    benef_cc_schiot   benef_cc_strk
Min.   :0.1300   Min.   :0.1900   Min.   :0.0700   Min.   :0.0100   Min.   :0.1000   Min.   :0.00000   Min.   :0.0100
1st Qu.:0.5600   1st Qu.:0.7100   1st Qu.:0.3700   1st Qu.:0.0900   1st Qu.:0.4000   1st Qu.:0.03000   1st Qu.:0.0600
Median :0.6300   Median :0.7500   Median :0.4700   Median :0.1000   Median :0.4500   Median :0.05000   Median :0.0900
Mean   :0.6226   Mean   :0.7213   Mean   :0.4821   Mean   :0.1099   Mean   :0.4566   Mean   :0.06085   Mean   :0.1013
3rd Qu.:0.6900   3rd Qu.:0.7500   3rd Qu.:0.5900   3rd Qu.:0.1300   3rd Qu.:0.5000   3rd Qu.:0.08000   3rd Qu.:0.1300
Max.   :0.7500   Max.   :0.7500   Max.   :0.7500   Max.   :0.7500   Max.   :0.7500   Max.   :0.75000   Max.   :0.7500
benef_average_risk_score ambulatory_services   anesthesia        category_iii      drugs_administered_other_than_oral_method
Min.   : 0.6415          Min.   :   0.000     Min.   :   0.000   Min.   : 0.00000   Min.   :    0.00
1st Qu.: 1.4093          1st Qu.:   0.000     1st Qu.:   0.000   1st Qu.: 0.00000   1st Qu.:    0.00
Median : 1.7462          Median :   0.000     Median :   0.000   Median : 0.00000   Median :    0.00
Mean   : 1.8578          Mean   :   2.644     Mean   :   1.381   Mean   : 0.01564   Mean   :   17.76
3rd Qu.: 2.1539          3rd Qu.:   0.000     3rd Qu.:   0.000   3rd Qu.: 0.00000   3rd Qu.:    0.00
Max.   :10.8422          Max.   :3601.250     Max.   :1322.500   Max.   :26.17647   Max.   :40693.65
chemotherapy_drugs vaccines_or_toxoids psychiatry_services_and_procedures dialysis_services_and_procedures gastroenterology_procedures
Min.   :   0.000   Min.   :   0.0000   Min.   :   0.0000                  Min.   :  0.0000                 Min.   : 0.000000
1st Qu.:   0.000   1st Qu.:   0.0000   1st Qu.:   0.0000                  1st Qu.:  0.0000                 1st Qu.: 0.000000
Median :   0.000   Median :   0.0000   Median :   0.0000                  Median :  0.0000                 Median : 0.000000
Mean   :   1.114   Mean   :   0.3667   Mean   :   0.5074                  Mean   :  0.6279                 Mean   : 0.005856
3rd Qu.:   0.000   3rd Qu.:   0.0000   3rd Qu.:   0.0000                  3rd Qu.:  0.0000                 3rd Qu.: 0.000000
Max.   :1212.793   Max.   :2493.0000   Max.   :2477.0000                  Max.   :504.2917                 Max.   :16.703704
ophthalmology_services_and_procedures special_otorhinolaryngologic_services_and_procedures cardiovascular_procedures
Min.   :   0.000                      Min.   :   0.000                                      Min.   :   0.000
1st Qu.:   0.000                      1st Qu.:   0.000                                      1st Qu.:   0.000
Median :   0.000                      Median :   0.000                                      Median :   0.000
Mean   :   8.212                      Mean   :   0.263                                      Mean   :   8.602
3rd Qu.:   0.000                      3rd Qu.:   0.000                                      3rd Qu.:   0.294
Max.   :2395.857                      Max.   :2204.000                                      Max.   :5057.000
non_invasive_vascular_diagnostic_studies pulmonary_procedures allergy_and_clinical_immunology_procedures endocrinology_services
Min.   :   0.0000                        Min.   :  0.0000     Min.   :   0.000                          Min.   : 0.00000
1st Qu.:   0.0000                        1st Qu.:  0.0000     1st Qu.:   0.000                          1st Qu.: 0.00000
Median :   0.0000                        Median :  0.0000     Median :   0.000                          Median : 0.00000
Mean   :   0.8009                        Mean   :  0.4567     Mean   :   0.318                          Mean   : 0.01177
3rd Qu.:   0.3864                        3rd Qu.:  0.0000     3rd Qu.:   0.000                          3rd Qu.: 0.00000
Max.   :1049.2727                        Max.   :622.5333     Max.   :1451.133                          Max.   :82.83333
neurology_and_neuromuscular_procedures highly_complex_drug_or_biologic_agent_administration therapeutic_procedures
Min.   :   0.000                       Min.   :  0.0000                                    Min.   :   0.000
1st Qu.:   0.000                       1st Qu.:  0.0000                                    1st Qu.:   0.000
Median :   0.000                       Median :  0.0000                                    Median :   0.000
Mean   :   0.487                       Mean   :  0.4708                                    Mean   :   0.515
3rd Qu.:   0.000                       3rd Qu.:  0.0000                                    3rd Qu.:   0.000
Max.   :3576.636                       Max.   :190.0267                                    Max.   :3825.286
physical_medicine_and_rehabilitation_evaluations moderate_conscious_sedation other_medicine_services_and_procedures
Min.   :   0.0000                                Min.   : 0.00000            Min.   : 0.00000
1st Qu.:   0.0000                                1st Qu.: 0.00000            1st Qu.: 0.00000
Median :   0.0000                                Median : 0.00000            Median : 0.00000
Mean   :   0.0725                                Mean   : 0.02211            Mean   : 0.03196
3rd Qu.:   0.0000                                3rd Qu.: 0.00000            3rd Qu.: 0.00000
Max.   :1885.8182                                Max.   :24.52381            Max.   :57.54930
office_or_other_outpatient_services hospital_observation_services hospital_inpatient_services emergency_department_services
Min.   : 0.00                       Min.   :  0.000               Min.   :   0.000           Min.   :  0.000
1st Qu.:  0.00                      1st Qu.:  0.000               1st Qu.:   0.000           1st Qu.:  0.000
Median :  0.00                      Median :  0.000               Median :   0.000           Median :  0.000
Mean   : 10.31                      Mean   :  1.879               Mean   :   8.002           Mean   :  1.284
3rd Qu.: 15.09                      3rd Qu.:  1.241               3rd Qu.:   3.310           3rd Qu.:  0.000
Max.   :989.60                      Max.   :315.556               Max.   :2376.000           Max.   :434.625
critical_care_services nursing_facility_services home_services     prolonged_services  other_care_evaluation_and_management_services
Min.   :  0.0000       Min.   :   0.000          Min.   :  0.0000   Min.   :  0.00000   Min.   :  0.00000
1st Qu.:  0.0000       1st Qu.:   0.000          1st Qu.:  0.0000   1st Qu.:  0.00000   1st Qu.:  0.00000
Median :  0.0000       Median :   0.000          Median :  0.0000   Median :  0.00000   Median :  0.00000
Mean   :  0.4598       Mean   :   3.416          Mean   :  0.3053   Mean   :  0.03455   Mean   :  0.07394
3rd Qu.:  0.0000       3rd Qu.:   0.000          3rd Qu.:  0.0000   3rd Qu.:  0.00000   3rd Qu.:  0.00000
Max.   :205.5217       Max.   :3692.000          Max.   :779.7143   Max.   :164.45455   Max.   :171.46667
organ_or_disease_oriented_panels   drug_assay        urinalysis_procedures chemistry_procedures hematology_and_coagulation_procedures
Min.   :  0.0000                   Min.   :  0.00000   Min.   :  0.0000      Min.   :   0.000     Min.   :  0.0000
1st Qu.:  0.0000                   1st Qu.:  0.00000   1st Qu.:  0.0000      1st Qu.:   0.000     1st Qu.:  0.0000
Median :  0.0000                   Median :  0.00000   Median :  0.0000      Median :   0.000     Median :  0.0000
Mean   :  0.4472                   Mean   :  0.01448   Mean   :  0.4352      Mean   :   1.795     Mean   :  0.9553
3rd Qu.:  0.0000                   3rd Qu.:  0.00000   3rd Qu.:  0.0000      3rd Qu.:   0.000     3rd Qu.:  0.0000
Max.   :655.2089                   Max.   :223.24540   Max.   :394.3333      Max.   :2068.918     Max.   :646.7200
immunology_procedures microbiology_procedures cytopathology_procedures surgical_pathology_procedures diagnostic_radiology_procedures
Min.   : 0.000        Min.   :   0.0000       Min.   :   0.000         Min.   :    0.00             Min.   :   0.000
```

```
 1st Qu.:  0.000       1st Qu.:   0.0000       1st Qu.:   0.000         1st Qu.:    0.00             1st Qu.:   0.000
 Median :  0.000       Median :   0.0000       Median :   0.000         Median :    0.00             Median :   0.000
 Mean   :  0.367       Mean   :   0.1519       Mean   :   0.954         Mean   :   10.67             Mean   :   7.346
 3rd Qu.:  0.000       3rd Qu.:   0.0000       3rd Qu.:   0.000         3rd Qu.:    0.00             3rd Qu.:  10.400
 Max.   :886.400       Max.   :1213.8462       Max.   :4708.333         Max.   :11267.60             Max.   :1293.000
 diagnostic_ultrasound_procedures radiologic_guidance breast_and_mammography bone_or_joint_studies radiation_oncology_treatment
 Min.   :  0.000                  Min.   :  0.00000   Min.   :  0.0000       Min.   :  0.0000      Min.   :  0.00000
 1st Qu.:  0.000                  1st Qu.:  0.00000   1st Qu.:  0.0000       1st Qu.:  0.0000      1st Qu.:  0.00000
 Median :  0.000                  Median :  0.00000   Median :  0.0000       Median :  0.0000      Median :  0.00000
 Mean   :  0.947                  Mean   :  0.07373   Mean   :  0.8461       Mean   :  0.3031      Mean   :  0.06207
 3rd Qu.:  0.978                  3rd Qu.:  0.00000   3rd Qu.:  0.0000       3rd Qu.:  0.0000      3rd Qu.:  0.00000
 Max.   :879.286                  Max.   :275.57143   Max.   :801.0000       Max.   :566.3333      Max.   :289.96774
 nuclear_medicine_procedures fine_needle_aspiration_biopsy_procedures surgical_procedures_on_the_auditory_system
 Min.   :  0.0000            Min.   :  0.00000                       Min.   :   0.0000
 1st Qu.:  0.0000            1st Qu.:  0.00000                       1st Qu.:   0.0000
 Median :  0.0000            Median :  0.00000                       Median :   0.0000
 Mean   :  0.5103            Mean   :  0.01935                       Mean   :   0.1892
 3rd Qu.:  0.0000            3rd Qu.:  0.00000                       3rd Qu.:   0.0000
 Max.   :483.0000            Max.   :138.15152                       Max.   :2456.5000
 surgical_procedures_on_the_cardiovascular_system surgical_procedures_on_the_digestive_system
surgical_procedures_on_the_eye_and_ocular_adnexa
 Min.   :   0.000                                 Min.   : 0.0000                            Min.   :  0.00
 1st Qu.:   0.000                                 1st Qu.: 0.0000                            1st Qu.:  0.00
 Median :   0.000                                 Median : 0.0000                            Median :  0.00
 Mean   :   1.071                                 Mean   : 0.2365                            Mean   :  1.16
 3rd Qu.:   0.000                                 3rd Qu.: 0.0000                            3rd Qu.:  0.00
 Max.   :1542.737                                 Max.   :72.4737                            Max.   :497.89
 surgical_procedures_on_genital_system surgical_procedures_on_the_hemic_and_lymphatic_systems
surgical_procedures_on_the_integumentary_system
 Min.   : 0.000000                     Min.   :0.000000                                      Min.   :   0.000
 1st Qu.: 0.000000                     1st Qu.:0.000000                                      1st Qu.:   0.000
 Median : 0.000000                     Median :0.000000                                      Median :   0.000
 Mean   : 0.008678                     Mean   :0.005834                                      Mean   :   2.095
 3rd Qu.: 0.000000                     3rd Qu.:0.000000                                      3rd Qu.:   0.000
 Max.   :15.439024                     Max.   :6.200000                                      Max.   :1795.486
 surgical_procedures_on_the_musculoskeletal_system surgical_procedures_on_the_nervous_system surgical_procedures_on_the_respiratory_system
 Min.   :  0.0000                                  Min.   :  0.0000                         Min.   :  0.0000
 1st Qu.:  0.0000                                  1st Qu.:  0.0000                         1st Qu.:  0.0000
 Median :  0.0000                                  Median :  0.0000                         Median :  0.0000
 Mean   :  0.2462                                  Mean   :  0.1018                         Mean   :  0.1201
 3rd Qu.:  0.0000                                  3rd Qu.:  0.0000                         3rd Qu.:  0.0000
 Max.   :275.0000                                  Max.   :415.2941                         Max.   :464.5556
 surgical_procedures_on_the_urinary_system other_surgical_procedures screening_examinations_and_disease_management_training
 Min.   :  0.0000                          Min.   :0.0000000        Min.   :0.00e+00
 1st Qu.:  0.0000                          1st Qu.:0.0000000        1st Qu.:0.00e+00
 Median :  0.0000                          Median :0.0000000        Median :0.00e+00
 Mean   :  0.2849                          Mean   :0.0005568        Mean   :7.53e-02
 3rd Qu.:  0.0000                          3rd Qu.:0.0000000        3rd Qu.:0.00e+00
 Max.   :218.2292                          Max.   :3.1034483        Max.   :1.31e+03
 miscellaneous_diagnostic_and_therapeutic_services initial_services_for_medicare_enrollment
gross_and_microscopic_examinations_prostate_biopsy
 Min.   :  0.000                                   Min.   : 0.0000                          Min.   : 0.00000
 1st Qu.:  0.000                                   1st Qu.: 0.0000                          1st Qu.: 0.00000
 Median :  0.000                                   Median : 0.0000                          Median : 0.00000
 Mean   :  1.352                                   Mean   : 0.0066                          Mean   : 0.02319
 3rd Qu.:  0.000                                   3rd Qu.: 0.0000                          3rd Qu.: 0.00000
 Max.   :858.333                                   Max.   :47.3333                          Max.   :70.25000
 counseling_screening_and_prevention_services miscellaneous_services
 Min.   :  0.0000                             Min.   :     0.0
 1st Qu.:  0.0000                             1st Qu.:     0.0
 Median :  0.0000                             Median :     0.0
 Mean   :  0.1297                             Mean   :    14.7
 3rd Qu.:  0.0000                             3rd Qu.:     0.0
 Max.   :586.0000                             Max.   :325751.7
```






```
[[1]]
[1] "benef_age_75_84"        "total_med_services"     "total_med_unique_benes" "benef_average_age"      "benef_age_less_65"
[6] "benef_age_65_74"
```

```
[[2]]
[1] "benef_race_hispanic"    "total_med_services"     "total_med_unique_benes" "benef_average_age"      "benef_age_less_65"
[6] "benef_race_api"
```

Histograms of numeric variables

```
hist.data.frame(d_model[, nums])
```

Train-test split stratified by provider type

```
set.seed(1234)
df_index <- createDataPartition(d_model$provider_type, p = .75, list = FALSE)
df_train <- d_model[df_index,]
df_test <- d_model[-df_index,]
```

Model 1

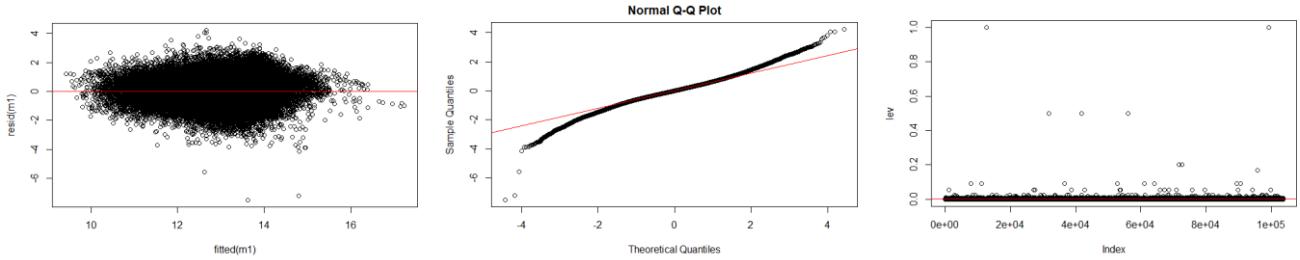
```
m1 <- lmer(log(overcharge) ~ year + nppes_provider_state + (1 | provider_type) + log(total_unique_benes) +
log(number_of_hcpcs) + benef_average_age + log(benef_female) + log(benef_race_white+0.01) +
log(benef_race_black+0.01) + log(benef_race_api+0.01) + log(benef_race_hispanic+0.01) +
log(benef_race_natind+0.01) + log(benef_dual+0.01) + log(benef_cc_afib+0.01) + log(benef_cc_alzrdsd) +
log(benef_cc_asthma) + log(benef_cc_cancer) + log(benef_cc_chf) + log(benef_cc_ckd) + log(benef_cc_copd) +
log(benef_cc_depr) + log(benef_cc_diab) + log(benef_cc_hyperl) + log(benef_cc_hypert) + log(benef_cc_ihd) +
```

```
log(benef_cc_ost) + log(benef_cc_raoa) + log(benef_cc_schiot+0.01) + log(benef_cc_strk) +
benef_average_risk_score, data=d_model, REML=FALSE)
summary(m1)
AIC(m1)
ranef(m1)
plot(resid(m1) ~ fitted(m1))
abline(0,0, col='red')
qqnorm(resid(m1))
qqline(resid(m1), col="red")
vif(m1)
lev = hat(model.matrix(m1))
plot(lev)
abline(5*mean(lev),0,col="red")
d_model[lev>0.02,]
```

```
      AIC       BIC    logLik  deviance  df.resid
 217693.2  218581.2 -108753.6  217507.2    103455

Scaled residuals:
     Min       1Q   Median       3Q      Max
-10.8422  -0.5854  -0.0049   0.5966   6.1393

Random effects:
 Groups        Name        Variance Std.Dev.
 provider_type (Intercept) 0.2391   0.4890
 Residual                  0.4769   0.6906
Number of obs: 103548, groups:  provider_type, 72

Fixed effects:
                                Estimate Std. Error t value
(Intercept)                    3.5250453  0.3161264  11.151
year2013                       0.0763932  0.0084721   9.017
year2014                       0.0799563  0.0124887   6.402
year2015                       0.1975352  0.0125780  15.705
year2016                       0.3236192  0.0149824  21.600
year2017                       0.1342575  0.0144391   9.298
nppes_provider_stateAE         0.5670144  0.7461291   0.760
nppes_provider_stateAK         1.0587298  0.2843567   3.723
nppes_provider_stateAL         0.1872822  0.2843232   0.659
nppes_provider_stateAP         0.5821487  0.4185875   1.391
nppes_provider_stateAR         0.2814818  0.2838493   0.992
nppes_provider_stateAZ         0.2780641  0.2824432   0.984
nppes_provider_stateCA         0.4765929  0.2824270   1.687
nppes_provider_stateCO         0.2267982  0.2834106   0.800
nppes_provider_stateCT         0.5017521  0.2828446   1.774
nppes_provider_stateDC         0.4065915  0.2839518   1.432
nppes_provider_stateDE         0.4153041  0.2835515   1.465
nppes_provider_stateFL         0.4235842  0.2825210   1.499
nppes_provider_stateGA         0.4314911  0.2828182   1.526
nppes_provider_stateGU         0.6976708  0.5640582   1.237
nppes_provider_stateHI        -0.0142243  0.2874652  -0.049
nppes_provider_stateIA         0.4482216  0.2854859   1.570
nppes_provider_stateID        -0.0160150  0.2878976  -0.056
nppes_provider_stateIL         0.4935884  0.2824972   1.747
nppes_provider_stateIN         0.2990689  0.2835885   1.055
nppes_provider_stateKS         0.3099101  0.2834199   1.093
nppes_provider_stateKY         0.2547538  0.2843348   0.896
nppes_provider_stateLA         0.2481476  0.2840038   0.874
nppes_provider_stateMA         0.4276823  0.2826845   1.513
nppes_provider_stateMD         0.4270835  0.2826050   1.511
nppes_provider_stateME         0.3905502  0.2906051   1.344
nppes_provider_stateMI         0.1431955  0.2827239   0.506
nppes_provider_stateMN         0.1192559  0.2831288   0.421
nppes_provider_stateMO         0.2759430  0.2834991   0.973
nppes_provider_stateMP         1.2320393  0.7465173   1.650
nppes_provider_stateMS         0.4171861  0.2852230   1.463
nppes_provider_stateMT         0.1540691  0.2956662   0.521
nppes_provider_stateNC         0.5237973  0.2828446   1.852
nppes_provider_stateND         0.1850962  0.2911493   0.636
nppes_provider_stateNE         0.3210800  0.2851011   1.126
nppes_provider_stateNH         0.5067595  0.2904889   1.745
nppes_provider_stateNJ         0.4146726  0.2824411   1.468
nppes_provider_stateNM         0.3071614  0.2837874   1.082
nppes_provider_stateNV         0.4859629  0.2828305   1.718
nppes_provider_stateNY         0.4622613  0.2824580   1.637
nppes_provider_stateOH         0.2977171  0.2828652   1.053
nppes_provider_stateOK         0.1473016  0.2825737   0.521
nppes_provider_stateOR         0.2463729  0.2834882   0.869
nppes_provider_statePA         0.2902197  0.2825658   1.027
nppes_provider_statePR        -0.1864057  0.3248546  -0.574
nppes_provider_stateRI         0.3233363  0.2851056   1.134
nppes_provider_stateSC         0.4754323  0.2837073   1.676
nppes_provider_stateSD        -0.1063534  0.2913089  -0.365
nppes_provider_stateTN         0.5143039  0.2837462   1.813
nppes_provider_stateTX         0.4668135  0.2825198   1.652
nppes_provider_stateUT         0.0825019  0.2851660   0.289
nppes_provider_stateVA         0.4444903  0.2826512   1.573
nppes_provider_stateVI         0.6072585  0.3517452   1.726
nppes_provider_stateVT         0.3033493  0.3015484   1.006
nppes_provider_stateWA         0.4272231  0.2824999   1.512
nppes_provider_stateWI         0.9423948  0.2835218   3.324
nppes_provider_stateWV         0.2425210  0.2885122   0.841
nppes_provider_stateWY         0.0265290  0.2974464   0.089
nppes_provider_stateXX         0.8604687  0.5642244   1.525
log(total_unique_benes)        0.6276684  0.0045186 138.906
log(number_of_hcpcs)           0.3938105  0.0041749  94.327
benef_average_age              0.0206859  0.0013743  15.052
log(benef_female)             -0.2314898  0.0207765 -11.142
```

```
log(benef_race_white + 0.01)     0.1839893  0.0090985  20.222
log(benef_race_black + 0.01)    -0.0006835  0.0040177  -0.170
log(benef_race_api + 0.01)       0.0565793  0.0045653  12.393
log(benef_race_hispanic + 0.01)  0.0387389  0.0043314   8.944
log(benef_race_natind + 0.01)   -0.0435039  0.0081051  -5.367
log(benef_dual + 0.01)           0.1060253  0.0082619  12.833
log(benef_cc_afib + 0.01)       -0.3443350  0.0106428 -32.354
log(benef_cc_alzrdsd)           -0.1902866  0.0100076 -19.014
log(benef_cc_asthma)            -0.2458124  0.0101341 -24.256
log(benef_cc_cancer)             0.0951188  0.0103215   9.216
log(benef_cc_chf)               -0.3202421  0.0144929 -22.096
log(benef_cc_ckd)               -0.0931940  0.0133785  -6.966
log(benef_cc_copd)               0.0569017  0.0122323   4.652
log(benef_cc_depr)              -0.0599352  0.0125701  -4.768
log(benef_cc_diab)               0.0090811  0.0144125   0.630
log(benef_cc_hyperl)             0.2802715  0.0258067  10.860
log(benef_cc_hypert)             0.3230801  0.0484626   6.667
log(benef_cc_ihd)                0.2966457  0.0211069  14.054
log(benef_cc_ost)               -0.1033183  0.0100235 -10.308
log(benef_cc_raoa)               0.3623999  0.0185017  19.587
log(benef_cc_schiot + 0.01)     -0.1934341  0.0082790 -23.364
log(benef_cc_strk)               0.2200209  0.0093918  23.427
benef_average_risk_score         0.3321479  0.0073577  45.143

Correlation matrix not shown by default, as p = 91 > 12.
Use print(x, correlation=TRUE)  or
    vcov(x)        if you need it

[1] 217693.2
$provider_type
                                                (Intercept)
allergy/immunology                               0.128939317
anesthesiologist assistants                     -0.025200215
anesthesiology                                   0.467365045
audiologist                                     -0.979256508
cardiac electrophysiology                       -0.114992857
cardiac surgery                                  0.279004850
cardiology                                      -0.103018906
certified clinical nurse specialist             -0.294819380
certified registered nurse anesthetist (crna)    0.302542223
chiropractic                                    -0.304049648
clinical psychologist                            0.498571740
colorectal surgery (proctology)                  0.162684217
critical care (intensivists)                     0.372628945
crna                                             0.052041676
dermatology                                     -0.463513683
diagnostic radiology                            -0.784809552
emergency medicine                               0.352334110
endocrinology                                   -0.317114986
family practice                                 -0.480022603
gastroenterology                                 0.261768759
general practice                                -0.337311522
general surgery                                 -0.036551445
geriatric medicine                               0.004297997
geriatric psychiatry                            -0.076375497
gynecological/oncology                           0.227465263
hand surgery                                     0.005579290
hematology                                       0.842627479
hematology/oncology                              1.126252129
hospice and palliative care                     -0.115815019
hospitalist                                      0.091699066
independent diagnostic testing facility         -0.282732372
infectious disease                               0.002131138
internal medicine                               -0.171802602
interventional radiology                        -0.547963311
licensed clinical social worker                  0.616519267
maxillofacial surgery                           -1.101690745
medical oncology                                 1.054362264
nephrology                                      -0.166697529
neurology                                       -0.100339238
neuropsychiatry                                 -0.105800102
neurosurgery                                     0.901089798
nuclear medicine                                -0.346984449
nurse practitioner                              -0.181073094
obstetrics/gynecology                           -0.563153504
occupational therapist                           0.667826796
ophthalmology                                    0.191212146
optometry                                       -0.882847947
oral surgery (dentist only)                     -0.277271128
orthopedic surgery                               0.443256442
osteopathic manipulative medicine                0.218337291
otolaryngology                                  -0.442751514
pain management                                  0.826576248
pathology                                       -0.202181353
pediatric medicine                              -0.508230697
peripheral vascular disease                     -0.019423254
physical medicine and rehabilitation             0.072886056
physical therapist                               0.263164208
physician assistant                             -0.140464094
plastic and reconstructive surgery               0.187772499
podiatry                                        -0.864734448
preventive medicine                             -0.612163454
psychiatry                                       0.322836935
pulmonary disease                                0.104568218
radiation oncology                               0.098405253
registered dietician/nutrition professional     -0.765939653
rheumatology                                     0.403595261
sleep medicine                                   0.022071904
sports medicine                                  0.173974702
surgical oncology                                1.077577268
thoracic surgery                                 0.024570185
```

```
urology                                    -0.100775835
vascular surgery                           -0.030663823

with conditional variances for "provider_type"
                                    GVIF Df GVIF^(1/(2*Df))
year                           27.593440  5        1.393415
nppes_provider_state           58.867437 58        1.035756
log(total_unique_benes)         1.406931  1        1.186141
log(number_of_hcpcs)            1.219996  1        1.104534
benef_average_age               2.952247  1        1.718210
log(benef_female)               1.560677  1        1.249270
log(benef_race_white + 0.01)    2.243690  1        1.497895
log(benef_race_black + 0.01)    2.536843  1        1.592747
log(benef_race_api + 0.01)      2.327185  1        1.525511
log(benef_race_hispanic + 0.01) 2.633769  1        1.622889
log(benef_race_natind + 0.01)   2.790956  1        1.670615
log(benef_dual + 0.01)          5.425685  1        2.329310
log(benef_cc_afib + 0.01)       3.248277  1        1.802298
log(benef_cc_alzrdsd)           4.968346  1        2.228979
log(benef_cc_asthma)            4.434714  1        2.105876
log(benef_cc_cancer)            1.725239  1        1.313483
log(benef_cc_chf)               6.230131  1        2.496023
log(benef_cc_ckd)               6.926388  1        2.631803
log(benef_cc_copd)              4.146799  1        2.036369
log(benef_cc_depr)              3.703466  1        1.924439
log(benef_cc_diab)              4.617867  1        2.148922
log(benef_cc_hyperl)            2.874412  1        1.695409
log(benef_cc_hypert)            2.680472  1        1.637215
log(benef_cc_ihd)               5.968396  1        2.443030
log(benef_cc_ost)               1.933574  1        1.390530
log(benef_cc_raoa)              2.039872  1        1.428241
log(benef_cc_schiot + 0.01)     3.619516  1        1.902503
log(benef_cc_strk)              3.600526  1        1.897505
benef_average_risk_score        3.235185  1        1.798662
```



Model 1 validation

```
m1_train <- lmer(log(overcharge) ~ year + nppes_provider_state + (1 | provider_type) + log(total_unique_benes)
+ log(number_of_hcpcs) + benef_average_age + log(benef_female) + log(benef_race_white+0.01) +
log(benef_race_black+0.01) + log(benef_race_api+0.01) + log(benef_race_hispanic+0.01) +
log(benef_race_natind+0.01) + log(benef_dual+0.01) + log(benef_cc_afib+0.01) + log(benef_cc_alzrdsd) +
log(benef_cc_asthma) + log(benef_cc_cancer) + log(benef_cc_chf) + log(benef_cc_ckd) + log(benef_cc_copd) +
log(benef_cc_depr) + log(benef_cc_diab) + log(benef_cc_hyperl) + log(benef_cc_hypert) + log(benef_cc_ihd) +
log(benef_cc_ost) + log(benef_cc_raoa) + log(benef_cc_schiot+0.01) + log(benef_cc_strk) +
benef_average_risk_score, REML=FALSE, data=df_train)
m1_pred <- predict(m1_train, df_test)
sqrt(mean((df_test$overcharge - exp(m1_pred)) ^ 2))
sqrt(mean((df_test[sample(nrow(df_test)), "overcharge"] - exp(m1_pred)) ^ 2)) # baseline 1
sqrt(mean((df_test$overcharge - mean(df_test$overcharge)) ^ 2)) # baseline 2
```

```
[1] 877635.1
[1] 1152428
[1] 1057204
```

Model 2

```
m2 <- lm(log(overcharge) ~ year + nppes_provider_state + provider_type + log(total_unique_benes) +
log(number_of_hcpcs) + benef_average_age + log(benef_female) + log(benef_race_white+0.01) +
log(benef_race_black+0.01) + log(benef_race_api+0.01) + log(benef_race_hispanic+0.01) +
log(benef_race_natind+0.01) + log(benef_dual+0.01) + log(benef_cc_afib+0.01) + log(benef_cc_alzrdsd) +
log(benef_cc_asthma) + log(benef_cc_cancer) + log(benef_cc_chf) + log(benef_cc_ckd) + log(benef_cc_copd) +
log(benef_cc_depr) + log(benef_cc_diab) + log(benef_cc_hyperl) + log(benef_cc_hypert) + log(benef_cc_ihd) +
log(benef_cc_ost) + log(benef_cc_raoa) + log(benef_cc_schiot+0.01) + log(benef_cc_strk) +
benef_average_risk_score + log(ambulatory_services+0.01) + log(anesthesia+0.01) + log(category_iii+0.01) +
log(drugs_administered_other_than_oral_method+0.01) + log(chemotherapy_drugs+0.01) +
log(vaccines_or_toxoids+0.01) + log(psychiatry_services_and_procedures+0.01) +
log(dialysis_services_and_procedures+0.01) + log(gastroenterology_procedures+0.01) +
log(ophthalmology_services_and_procedures+0.01) +
log(special_otorhinolaryngologic_services_and_procedures+0.01) + log(cardiovascular_procedures+0.01) +
log(non_invasive_vascular_diagnostic_studies+0.01) + log(pulmonary_procedures+0.01) +
log(allergy_and_clinical_immunology_procedures+0.01) + log(endocrinology_services+0.01) +
log(neurology_and_neuromuscular_procedures+0.01) +
log(highly_complex_drug_or_biologic_agent_administration+0.01) + log(therapeutic_procedures+0.01) +
```

```
log(physical_medicine_and_rehabilitation_evaluations+0.01) + log(moderate_conscious_sedation+0.01) +
log(other_medicine_services_and_procedures+0.01) + log(office_or_other_outpatient_services+0.01) +
log(hospital_observation_services+0.01) + log(hospital_inpatient_services+0.01) +
log(emergency_department_services+0.01) + log(critical_care_services+0.01) +
log(nursing_facility_services+0.01) + log(home_services+0.01) + log(prolonged_services+0.01) +
log(other_care_evaluation_and_management_services+0.01) + log(organ_or_disease_oriented_panels+0.01) +
log(drug_assay+0.01) + log(urinalysis_procedures+0.01) + log(chemistry_procedures+0.01) +
log(hematology_and_coagulation_procedures+0.01) + log(immunology_procedures+0.01) +
log(microbiology_procedures+0.01) + log(cytopathology_procedures+0.01) +
log(surgical_pathology_procedures+0.01) + log(diagnostic_radiology_procedures+0.01) +
log(diagnostic_ultrasound_procedures+0.01) + log(radiologic_guidance+0.01) + log(breast_and_mammography+0.01) +
log(bone_or_joint_studies+0.01) + log(radiation_oncology_treatment+0.01) +
log(nuclear_medicine_procedures+0.01) + log(fine_needle_aspiration_biopsy_procedures+0.01) +
log(surgical_procedures_on_the_auditory_system+0.01) +
log(surgical_procedures_on_the_cardiovascular_system+0.01) +
log(surgical_procedures_on_the_digestive_system+0.01) +
log(surgical_procedures_on_the_eye_and_ocular_adnexa+0.01) + log(surgical_procedures_on_genital_system+0.01) +
log(surgical_procedures_on_the_hemic_and_lymphatic_systems+0.01) +
log(surgical_procedures_on_the_integumentary_system+0.01) +
log(surgical_procedures_on_the_musculoskeletal_system+0.01) +
log(surgical_procedures_on_the_nervous_system+0.01) + log(surgical_procedures_on_the_respiratory_system+0.01) +
log(surgical_procedures_on_the_urinary_system+0.01) + log(other_surgical_procedures+0.01) +
log(screening_examinations_and_disease_management_training+0.01) +
log(miscellaneous_diagnostic_and_therapeutic_services+0.01) +
log(initial_services_for_medicare_enrollment+0.01) +
log(gross_and_microscopic_examinations_prostate_biopsy+0.01) +
log(counseling_screening_and_prevention_services+0.01) + log(miscellaneous_services+0.01), data=d_model)
summary(m2)
AIC(m2)
plot(resid(m2) ~ fitted(m2))
abline(0,0, col='red')
qqnorm(resid(m2))
qqline(resid(m2), col="red")
vif(m2)
lev = hat(model.matrix(m2))
plot(lev)
abline(0.4,0,col="red")
d_model[lev>0.4,]
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-8.1552 -0.3421  0.0142  0.3661  4.1897

Coefficients:
                                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                                       12.0269864  0.3357937  35.817  < 2e-16 ***
year2013                                           0.0766963  0.0075282  10.188  < 2e-16 ***
year2014                                           0.1031479  0.0111226   9.274  < 2e-16 ***
year2015                                           0.2156645  0.0112275  19.209  < 2e-16 ***
year2016                                           0.3208013  0.0134399  23.869  < 2e-16 ***
year2017                                           0.2124019  0.0129562  16.394  < 2e-16 ***
nppes_provider_stateAE                             0.1142558  0.6625909   0.172 0.863094
nppes_provider_stateAK                             0.5912819  0.2525754   2.341 0.019233 *
nppes_provider_stateAL                            -0.1686292  0.2525398  -0.668 0.504305
nppes_provider_stateAP                             0.1913265  0.3718237   0.515 0.606860
nppes_provider_stateAR                            -0.0586293  0.2521274  -0.233 0.816121
nppes_provider_stateAZ                            -0.2215489  0.2509011  -0.883 0.377231
nppes_provider_stateCA                             0.0591957  0.2508690   0.236 0.813462
nppes_provider_stateCO                            -0.1234034  0.2517494  -0.490 0.624005
nppes_provider_stateCT                             0.1122515  0.2512612   0.447 0.655055
nppes_provider_stateDC                             0.0053630  0.2522245   0.021 0.983036
nppes_provider_stateDE                             0.0445133  0.2518661   0.177 0.859718
nppes_provider_stateFL                            -0.0106066  0.2509514  -0.042 0.966287
nppes_provider_stateGA                             0.0825272  0.2512120   0.329 0.742522
nppes_provider_stateGU                             0.2161410  0.5008930   0.432 0.666097
nppes_provider_stateHI                            -0.3477835  0.2553468  -1.362 0.173199
nppes_provider_stateIA                             0.0298641  0.2535917   0.118 0.906254
nppes_provider_stateID                            -0.4088803  0.2557289  -1.599 0.109850
nppes_provider_stateIL                             0.1596945  0.2509217   0.636 0.524496
nppes_provider_stateIN                             0.0254386  0.2518952   0.101 0.919560
nppes_provider_stateKS                             0.0203723  0.2517393   0.081 0.935501
nppes_provider_stateKY                            -0.0237101  0.2525496  -0.094 0.925202
nppes_provider_stateLA                            -0.0341628  0.2522421  -0.135 0.892267
nppes_provider_stateMA                             0.0787271  0.2511051   0.314 0.753884
nppes_provider_stateMD                             0.0018616  0.2510342   0.007 0.994083
nppes_provider_stateME                             0.1489166  0.2581275   0.577 0.564001
nppes_provider_stateMI                            -0.1536646  0.2511272  -0.612 0.540606
nppes_provider_stateMN                            -0.3745436  0.2515309  -1.489 0.136476
nppes_provider_stateMO                             0.0473619  0.2518131   0.188 0.850811
nppes_provider_stateMP                             0.7906500  0.6628887   1.193 0.232976
nppes_provider_stateMS                             0.0102538  0.2533652   0.040 0.967718
nppes_provider_stateMT                            -0.0967516  0.2625787  -0.368 0.712526
nppes_provider_stateNC                             0.1260534  0.2512422   0.502 0.615865
nppes_provider_stateND                             0.0181265  0.2585922   0.070 0.944117
nppes_provider_stateNE                             0.0283014  0.2532065   0.112 0.911004
nppes_provider_stateNH                             0.2026490  0.2580155   0.785 0.432212
nppes_provider_stateNJ                             0.0247402  0.2508792   0.099 0.921445
nppes_provider_stateNM                            -0.1322162  0.2520639  -0.525 0.599908
nppes_provider_stateNV                             0.0679822  0.2512163   0.271 0.786690
nppes_provider_stateNY                             0.0641912  0.2509068   0.256 0.798077
nppes_provider_stateOH                             0.0695097  0.2512438   0.277 0.782040
nppes_provider_stateOK                            -0.1221106  0.2509726  -0.487 0.626579
```

nppes_provider_stateOR	-0.1384059	0.2518101	-0.550	0.582565	
nppes_provider_statePA	-0.0147334	0.2509872	-0.059	0.953190	
nppes_provider_statePR	-0.4980393	0.2885439	-1.726	0.084343	.
nppes_provider_stateRI	-0.0879586	0.2532530	-0.347	0.728355	
nppes_provider_stateSC	0.0932014	0.2519991	0.370	0.711496	
nppes_provider_stateSD	-0.4086357	0.2587671	-1.579	0.114301	
nppes_provider_stateTN	0.1660525	0.2520470	0.659	0.510016	
nppes_provider_stateTX	0.1098870	0.2509421	0.438	0.661461	
nppes_provider_stateUT	-0.1769395	0.2533035	-0.699	0.484849	
nppes_provider_stateVA	0.0528949	0.2510698	0.211	0.833139	
nppes_provider_stateVI	0.0249831	0.3125220	0.080	0.936285	
nppes_provider_stateVT	0.0832585	0.2678821	0.311	0.755951	
nppes_provider_stateWA	-0.0474937	0.2509517	-0.189	0.849894	
nppes_provider_stateWI	0.6268786	0.2518439	2.489	0.012806	*
nppes_provider_stateWV	0.0463289	0.2562383	0.181	0.856522	
nppes_provider_stateWY	-0.6932231	0.2642686	-2.623	0.008713	**
nppes_provider_stateXX	0.6771907	0.5016840	1.350	0.177072	
provider_typeanesthesiologist assistants	0.0412220	0.3012726	0.137	0.891168	
provider_typeanesthesiology	0.2808274	0.1290311	2.176	0.029525	*
provider_typeaudiologist	-0.8567072	0.1471669	-5.821	5.86e-09	***
provider_typecardiac electrophysiology	0.1232318	0.1230442	1.002	0.316575	
provider_typecardiac surgery	0.3796339	0.1468344	2.585	0.009726	**
provider_typecardiology	-0.0884797	0.1212340	-0.730	0.465498	
provider_typecertified clinical nurse specialist	-0.3880912	0.1782510	-2.177	0.029467	*
provider_typecertified registered nurse anesthetist (crna)	0.0568447	0.1422697	0.400	0.689485	
provider_typechiropractic	-0.4291527	0.4523424	-0.949	0.342758	
provider_typeclinical psychologist	0.0878887	0.1539462	0.571	0.568065	
provider_typecolorectal surgery (proctology)	0.4590469	0.1687461	2.720	0.006523	**
provider_typecritical care (intensivists)	0.1993551	0.1258951	1.584	0.113310	
provider_typecrna	-0.0573492	0.1310992	-0.437	0.661787	
provider_typedermatology	-0.4527770	0.1235169	-3.666	0.000247	***
provider_typediagnostic radiology	-0.5999289	0.1212636	-4.947	7.54e-07	***
provider_typeemergency medicine	-0.0449311	0.1219867	-0.368	0.712630	
provider_typeendocrinology	-0.2158049	0.1229512	-1.755	0.079227	.
provider_typefamily practice	-0.2533358	0.1209871	-2.094	0.036271	*
provider_typegastroenterology	0.2944147	0.1235280	2.383	0.017156	*
provider_typegeneral practice	-0.1978604	0.1272428	-1.555	0.119953	
provider_typegeneral surgery	0.1582115	0.1235403	1.281	0.200321	
provider_typegeriatric medicine	0.0837522	0.1329516	0.630	0.528732	
provider_typegeriatric psychiatry	-0.2689443	0.2399028	-1.121	0.262267	
provider_typegynecological/oncology	0.8336488	0.2639251	3.159	0.001585	**
provider_typehand surgery	0.1179495	0.1442455	0.818	0.413530	
provider_typehematology	0.4484557	0.1454645	3.083	0.002050	**
provider_typehematology/oncology	0.5712812	0.1235230	4.625	3.75e-06	***
provider_typehospice and palliative care	-0.1756162	0.1571546	-1.117	0.263794	
provider_typehospitalist	0.1639188	0.1526631	1.074	0.282947	
provider_typeindependent diagnostic testing facility	-0.2076324	0.3745613	-0.554	0.579351	
provider_typeinfectious disease	0.0402936	0.1228887	0.328	0.742998	
provider_typeinternal medicine	-0.0635402	0.1204898	-0.527	0.597952	
provider_typeinterventional radiology	-0.4548286	0.1232521	-3.690	0.000224	***
provider_typelicensed clinical social worker	0.4101916	0.2869497	1.429	0.152867	
provider_typemaxillofacial surgery	-1.5344535	0.2622185	-5.852	4.88e-09	***
provider_typemedical oncology	0.5346323	0.1293112	4.134	3.56e-05	***
provider_typenephrology	-0.0189698	0.1249605	-0.152	0.879340	
provider_typeneurology	-0.2761892	0.1226836	-2.251	0.024373	*
provider_typeneuropsychiatry	-0.4483856	0.2646506	-1.694	0.090220	.
provider_typeneurosurgery	1.0904045	0.1335987	8.162	3.34e-16	***
provider_typenuclear medicine	-0.1086982	0.1249660	-0.870	0.384400	
provider_typenurse practitioner	-0.1595925	0.1221333	-1.307	0.191315	
provider_typeobstetrics/gynecology	-0.3139102	0.1503183	-2.088	0.036773	*
provider_typeoccupational therapist	1.1047143	0.4522266	2.443	0.014574	*
provider_typeophthalmology	-1.2042475	0.1304954	-9.228	< 2e-16	***
provider_typeoptometry	-1.5421370	0.1318118	-11.700	< 2e-16	***
provider_typeoral surgery (dentist only)	-0.5866661	0.2487791	-2.358	0.018367	*
provider_typeorthopedic surgery	0.6605399	0.1228931	5.375	7.68e-08	***
provider_typeosteopathic manipulative medicine	-0.0402887	0.2035235	-0.198	0.843080	
provider_typeotolaryngology	-0.1741344	0.1235978	-1.409	0.158874	
provider_typepain management	0.4544044	0.1325655	3.428	0.000609	***
provider_typepathology	-0.4571625	0.1238240	-3.692	0.000223	***
provider_typepediatric medicine	-0.4240618	0.1498656	-2.830	0.004661	**
provider_typeperipheral vascular disease	0.1288365	0.1742811	0.739	0.459760	
provider_typephysical medicine and rehabilitation	-0.0204309	0.1235365	-0.165	0.868643	
provider_typephysical therapist	0.5128612	0.1764984	2.906	0.003664	**
provider_typephysician assistant	-0.1494984	0.1235302	-1.210	0.226198	
provider_typeplastic and reconstructive surgery	0.4302277	0.1781623	2.415	0.015745	*
provider_typepodiatry	-0.7651742	0.1235825	-6.192	5.98e-10	***
provider_typepreventive medicine	-0.4914346	0.2284692	-2.151	0.031479	*
provider_typepsychiatry	-0.0710871	0.1334405	-0.533	0.594225	
provider_typepulmonary disease	-0.0003801	0.1211156	-0.003	0.997496	
provider_typeradiation oncology	-0.2915978	0.1336215	-2.182	0.029092	*
provider_typeregistered dietician/nutrition professional	-0.9827464	0.2646795	-3.713	0.000205	***
provider_typerheumatology	0.4054310	0.1269044	3.195	0.001400	**
provider_typesleep medicine	0.1067834	0.1955724	0.546	0.585064	
provider_typesports medicine	0.3649507	0.1713757	2.130	0.033212	*
provider_typesurgical oncology	1.4817268	0.2484376	5.964	2.47e-09	***
provider_typethoracic surgery	0.3004527	0.1518055	1.979	0.047797	*
provider_typeurology	0.0712325	0.1282529	0.555	0.578618	
provider_typevascular surgery	0.2675581	0.1227720	2.179	0.029311	*
log(total_unique_benes)	0.5828783	0.0043913	132.734	< 2e-16	***
log(number_of_hcpcs)	0.2634526	0.0048987	53.780	< 2e-16	***
benef_average_age	0.0105587	0.0012824	8.233	< 2e-16	***
log(benef_female)	-0.2017933	0.0204020	-9.891	< 2e-16	***
log(benef_race_white + 0.01)	0.1437499	0.0081334	17.674	< 2e-16	***
log(benef_race_black + 0.01)	0.0242350	0.0036026	6.727	1.74e-11	***
log(benef_race_api + 0.01)	0.0479643	0.0040939	11.716	< 2e-16	***
log(benef_race_hispanic + 0.01)	0.0420662	0.0038620	10.892	< 2e-16	***
log(benef_race_natind + 0.01)	-0.0036366	0.0072323	-0.503	0.615087	
log(benef_dual + 0.01)	0.0975967	0.0074335	13.129	< 2e-16	***
log(benef_cc_afib + 0.01)	-0.2289775	0.0095901	-23.876	< 2e-16	***
log(benef_cc_alzrdsd)	-0.1390085	0.0089609	-15.513	< 2e-16	***
log(benef_cc_asthma)	-0.1608223	0.0091709	-17.536	< 2e-16	***

```
log(benef_cc_cancer)                                                0.0865106  0.0095499   9.059  < 2e-16 ***
log(benef_cc_chf)                                                  -0.2132997  0.0130491 -16.346  < 2e-16 ***
log(benef_cc_ckd)                                                  -0.1138670  0.0120564  -9.445  < 2e-16 ***
log(benef_cc_copd)                                                 -0.0109453  0.0110185  -0.993 0.320539
log(benef_cc_depr)                                                 -0.0647628  0.0112490  -5.757 8.58e-09 ***
log(benef_cc_diab)                                                 -0.0479550  0.0128924  -3.720 0.000200 ***
log(benef_cc_hyperl)                                                0.2896621  0.0234233  12.366  < 2e-16 ***
log(benef_cc_hypert)                                                0.1450866  0.0437406   3.317 0.000910 ***
log(benef_cc_ihd)                                                   0.1830120  0.0190720   9.596  < 2e-16 ***
log(benef_cc_ost)                                                  -0.0326632  0.0091077  -3.586 0.000336 ***
log(benef_cc_raoa)                                                  0.1632097  0.0172954   9.437  < 2e-16 ***
log(benef_cc_schiot + 0.01)                                        -0.1195346  0.0074893 -15.961  < 2e-16 ***
log(benef_cc_strk)                                                  0.1302987  0.0086203  15.115  < 2e-16 ***
benef_average_risk_score                                            0.2457659  0.0069176  35.528  < 2e-16 ***
log(ambulatory_services + 0.01)                                     0.0754848  0.0011811  63.912  < 2e-16 ***
log(anesthesia + 0.01)                                              0.0746724  0.0062196  12.006  < 2e-16 ***
log(category_iii + 0.01)                                            0.0325295  0.0036972   8.798  < 2e-16 ***
log(drugs_administered_other_than_oral_method + 0.01)               0.0249341  0.0010912  22.850  < 2e-16 ***
log(chemotherapy_drugs + 0.01)                                      0.0876360  0.0022100  39.655  < 2e-16 ***
log(vaccines_or_toxoids + 0.01)                                    -0.0399222  0.0020182 -19.781  < 2e-16 ***
log(psychiatry_services_and_procedures + 0.01)                      0.0463513  0.0067184   6.899 5.26e-12 ***
log(dialysis_services_and_procedures + 0.01)                        0.0051427  0.0043692   1.177 0.239177
log(gastroenterology_procedures + 0.01)                             0.0362473  0.0068427   5.297 1.18e-07 ***
log(ophthalmology_services_and_procedures + 0.01)                   0.0621903  0.0060483  10.282  < 2e-16 ***
log(special_otorhinolaryngologic_services_and_procedures + 0.01)    0.0135380  0.0041683   3.248 0.001163 **
log(cardiovascular_procedures + 0.01)                              -0.0073654  0.0014804  -4.975 6.53e-07 ***
log(non_invasive_vascular_diagnostic_studies + 0.01)                0.0185540  0.0012753  14.549  < 2e-16 ***
log(pulmonary_procedures + 0.01)                                    0.0020872  0.0026735   0.781 0.434986
log(allergy_and_clinical_immunology_procedures + 0.01)              0.0311406  0.0035183   8.851  < 2e-16 ***
log(endocrinology_services + 0.01)                                  0.0097512  0.0076037   1.282 0.199698
log(neurology_and_neuromuscular_procedures + 0.01)                  0.0368057  0.0024421  15.072  < 2e-16 ***
log(highly_complex_drug_or_biologic_agent_administration + 0.01)   -0.0035660  0.0020481  -1.741 0.081664 .
log(therapeutic_procedures + 0.01)                                  0.0039500  0.0034128   1.157 0.247108
log(physical_medicine_and_rehabilitation_evaluations + 0.01)        0.0123267  0.0042239   2.918 0.003520 **
log(moderate_conscious_sedation + 0.01)                             0.0169562  0.0033897   5.002 5.67e-07 ***
log(other_medicine_services_and_procedures + 0.01)                  0.0073072  0.0045969   1.590 0.111929
log(office_or_other_outpatient_services + 0.01)                     0.0022643  0.0012661   1.788 0.073706 .
log(hospital_observation_services + 0.01)                           0.0073224  0.0019286   3.797 0.000147 ***
log(hospital_inpatient_services + 0.01)                             0.0193529  0.0017140  11.291  < 2e-16 ***
log(emergency_department_services + 0.01)                           0.0794753  0.0026625  29.849  < 2e-16 ***
log(critical_care_services + 0.01)                                  0.0433490  0.0016526  26.231  < 2e-16 ***
log(nursing_facility_services + 0.01)                               0.0227687  0.0014151  16.090  < 2e-16 ***
log(home_services + 0.01)                                           0.0093211  0.0025764   3.618 0.000297 ***
log(prolonged_services + 0.01)                                      0.0171569  0.0031163   5.506 3.69e-08 ***
log(other_care_evaluation_and_management_services + 0.01)           0.0183462  0.0030802   5.956 2.59e-09 ***
log(organ_or_disease_oriented_panels + 0.01)                        0.0028970  0.0024014   1.206 0.227672
log(drug_assay + 0.01)                                              0.0602716  0.0080948   7.446 9.72e-14 ***
log(urinalysis_procedures + 0.01)                                  -0.0373142  0.0023772 -15.697  < 2e-16 ***
log(chemistry_procedures + 0.01)                                    0.0005817  0.0015431   0.377 0.706213
log(hematology_and_coagulation_procedures + 0.01)                   0.0061794  0.0014255   4.335 1.46e-05 ***
log(immunology_procedures + 0.01)                                   0.0024393  0.0027807   0.877 0.380354
log(microbiology_procedures + 0.01)                                -0.0336169  0.0029689 -11.323  < 2e-16 ***
log(cytopathology_procedures + 0.01)                               -0.0034011  0.0025288  -1.345 0.178650
log(surgical_pathology_procedures + 0.01)                           0.0698497  0.0027819  25.108  < 2e-16 ***
log(diagnostic_radiology_procedures + 0.01)                         0.0148578  0.0015028   9.886  < 2e-16 ***
log(diagnostic_ultrasound_procedures + 0.01)                        0.0040394  0.0014949   2.702 0.006892 **
log(radiologic_guidance + 0.01)                                     0.0269125  0.0026300  10.233  < 2e-16 ***
log(breast_and_mammography + 0.01)                                 -0.0114206  0.0020441  -5.587 2.31e-08 ***
log(bone_or_joint_studies + 0.01)                                  -0.0094072  0.0017058  -5.515 3.50e-08 ***
log(radiation_oncology_treatment + 0.01)                            0.1546038  0.0080742  19.148  < 2e-16 ***
log(nuclear_medicine_procedures + 0.01)                            -0.0003536  0.0014416  -0.245 0.806243
log(fine_needle_aspiration_biopsy_procedures + 0.01)                0.0185857  0.0037923   4.901 9.56e-07 ***
log(surgical_procedures_on_the_auditory_system + 0.01)             -0.0050072  0.0038535  -1.299 0.193811
log(surgical_procedures_on_the_cardiovascular_system + 0.01)        0.0303972  0.0012804  23.741  < 2e-16 ***
log(surgical_procedures_on_the_digestive_system + 0.01)             0.0258627  0.0033472   7.727 1.11e-14 ***
log(surgical_procedures_on_the_eye_and_ocular_adnexa + 0.01)        0.1710623  0.0037292  45.871  < 2e-16 ***
log(surgical_procedures_on_genital_system + 0.01)                  -0.0275467  0.0076086  -3.620 0.000294 ***
log(surgical_procedures_on_the_hemic_and_lymphatic_systems + 0.01)  0.0178401  0.0049855   3.578 0.000346 ***
log(surgical_procedures_on_the_integumentary_system + 0.01)         0.0063206  0.0022678   2.787 0.005318 **
log(surgical_procedures_on_the_musculoskeletal_system + 0.01)       0.0047257  0.0026040   1.815 0.069561 .
log(surgical_procedures_on_the_nervous_system + 0.01)               0.0428802  0.0027172  15.781  < 2e-16 ***
log(surgical_procedures_on_the_respiratory_system + 0.01)           0.0032207  0.0029313   1.099 0.271897
log(surgical_procedures_on_the_urinary_system + 0.01)               0.0046308  0.0051234   0.904 0.366068
log(other_surgical_procedures + 0.01)                               0.0417089  0.0143277   2.911 0.003603 **
log(screening_examinations_and_disease_management_training + 0.01) -0.0118195  0.0032885  -3.594 0.000326 ***
log(miscellaneous_diagnostic_and_therapeutic_services + 0.01)       0.0052313  0.0015218   3.437 0.000587 ***
log(initial_services_for_medicare_enrollment + 0.01)                0.0034346  0.0059774   0.575 0.565560
log(gross_and_microscopic_examinations_prostate_biopsy + 0.01)      0.0333433  0.0051247   6.506 7.73e-11 ***
log(counseling_screening_and_prevention_services + 0.01)           -0.0017267  0.0028426  -0.607 0.543556
log(miscellaneous_services + 0.01)                                  0.0329315  0.0008487  38.804  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6131 on 103320 degrees of freedom
Multiple R-squared:  0.5987,  Adjusted R-squared:  0.5978
F-statistic: 679.1 on 227 and 103320 DF,  p-value: < 2.2e-16

[1] 192771.2


                                                          GVIF Df GVIF^(1/(2*Df))
year                                                3.049958e+01  5        1.407438
nppes_provider_state                                1.191106e+02 58        1.042068
provider_type                                       4.749749e+13 71        1.248287
log(total_unique_benes)                             3.327034e+00  1        1.824016
log(number_of_hcpcs)                                4.972237e+00  1        2.229851
benef_average_age                                   4.000266e+00  1        2.000067
log(benef_female)                                   4.804715e+00  1        2.191966
log(benef_race_white + 0.01)                        2.384044e+00  1        1.544035
log(benef_race_black + 0.01)                        2.703474e+00  1        1.644224
log(benef_race_api + 0.01)                          2.591417e+00  1        1.609788
```

log(benef_race_hispanic + 0.01)	2.770088e+00	1	1.664358
log(benef_race_natind + 0.01)	3.175394e+00	1	1.781964
log(benef_dual + 0.01)	6.656169e+00	1	2.579955
log(benef_cc_afib + 0.01)	5.027064e+00	1	2.242111
log(benef_cc_alzrdsd)	6.834911e+00	1	2.614366
log(benef_cc_asthma)	5.705660e+00	1	2.388652
log(benef_cc_cancer)	3.050780e+00	1	1.746648
log(benef_cc_chf)	1.041894e+01	1	3.227839
log(benef_cc_ckd)	9.282172e+00	1	3.046666
log(benef_cc_copd)	6.610189e+00	1	2.571029
log(benef_cc_depr)	4.830271e+00	1	2.197788
log(benef_cc_diab)	5.898812e+00	1	2.428747
log(benef_cc_hyperl)	3.692601e+00	1	1.921614
log(benef_cc_hypert)	3.250890e+00	1	1.803023
log(benef_cc_ihd)	9.747511e+00	1	3.122100
log(benef_cc_ost)	2.524315e+00	1	1.588809
log(benef_cc_raoa)	2.996994e+00	1	1.731183
log(benef_cc_schiot + 0.01)	5.371165e+00	1	2.317577
log(benef_cc_strk)	5.949484e+00	1	2.439156
benef_average_risk_score	6.205251e+00	1	2.491034
log(ambulatory_services + 0.01)	2.068057e+00	1	1.438074
log(anesthesia + 0.01)	1.182194e+01	1	3.438305
log(category_iii + 0.01)	1.070954e+00	1	1.034869
log(drugs_administered_other_than_oral_method + 0.01)	2.669716e+00	1	1.633927
log(chemotherapy_drugs + 0.01)	2.245554e+00	1	1.498517
log(vaccines_or_toxoids + 0.01)	2.758722e+00	1	1.660940
log(psychiatry_services_and_procedures + 0.01)	3.323109e+00	1	1.822940
log(dialysis_services_and_procedures + 0.01)	6.614825e+00	1	2.571930
log(gastroenterology_procedures + 0.01)	1.234016e+00	1	1.110863
log(ophthalmology_services_and_procedures + 0.01)	6.110986e+01	1	7.817280
log(special_otorhinolaryngologic_services_and_procedures + 0.01)	2.402202e+00	1	1.549904
log(cardiovascular_procedures + 0.01)	6.161754e+00	1	2.482288
log(non_invasive_vascular_diagnostic_studies + 0.01)	2.243873e+00	1	1.497956
log(pulmonary_procedures + 0.01)	2.862244e+00	1	1.691817
log(allergy_and_clinical_immunology_procedures + 0.01)	1.170284e+00	1	1.081796
log(endocrinology_services + 0.01)	1.213573e+00	1	1.101623
log(neurology_and_neuromuscular_procedures + 0.01)	2.128772e+00	1	1.459031
log(highly_complex_drug_or_biologic_agent_administration + 0.01)	2.887611e+00	1	1.699297
log(therapeutic_procedures + 0.01)	1.424943e+00	1	1.193710
log(physical_medicine_and_rehabilitation_evaluations + 0.01)	1.520258e+00	1	1.232987
log(moderate_conscious_sedation + 0.01)	1.271115e+00	1	1.127437
log(other_medicine_services_and_procedures + 0.01)	1.252261e+00	1	1.119045
log(office_or_other_outpatient_services + 0.01)	5.943152e+00	1	2.437858
log(hospital_observation_services + 0.01)	7.590800e+00	1	2.755141
log(hospital_inpatient_services + 0.01)	8.741131e+00	1	2.956540
log(emergency_department_services + 0.01)	6.339414e+00	1	2.517819
log(critical_care_services + 0.01)	2.260305e+00	1	1.503431
log(nursing_facility_services + 0.01)	1.764562e+00	1	1.328368
log(home_services + 0.01)	1.361944e+00	1	1.167024
log(prolonged_services + 0.01)	1.092081e+00	1	1.045027
log(other_care_evaluation_and_management_services + 0.01)	1.299457e+00	1	1.139937
log(organ_or_disease_oriented_panels + 0.01)	2.274932e+00	1	1.508288
log(drug_assay + 0.01)	1.120253e+00	1	1.058420
log(urinalysis_procedures + 0.01)	3.433172e+00	1	1.852882
log(chemistry_procedures + 0.01)	2.258097e+00	1	1.502697
log(hematology_and_coagulation_procedures + 0.01)	2.096224e+00	1	1.447834
log(immunology_procedures + 0.01)	1.675815e+00	1	1.294533
log(microbiology_procedures + 0.01)	1.381763e+00	1	1.175484
log(cytopathology_procedures + 0.01)	3.192999e+00	1	1.786896
log(surgical_pathology_procedures + 0.01)	9.475655e+00	1	3.078255
log(diagnostic_radiology_procedures + 0.01)	7.462263e+00	1	2.731714
log(diagnostic_ultrasound_procedures + 0.01)	3.761808e+00	1	1.939538
log(radiologic_guidance + 0.01)	2.114130e+00	1	1.454005
log(breast_and_mammography + 0.01)	4.314975e+00	1	2.077252
log(bone_or_joint_studies + 0.01)	2.019271e+00	1	1.421011
log(radiation_oncology_treatment + 0.01)	1.391839e+00	1	1.179762
log(nuclear_medicine_procedures + 0.01)	2.207631e+00	1	1.485810
log(fine_needle_aspiration_biopsy_procedures + 0.01)	1.209032e+00	1	1.099560
log(surgical_procedures_on_the_auditory_system + 0.01)	2.422732e+00	1	1.556513
log(surgical_procedures_on_the_cardiovascular_system + 0.01)	2.032705e+00	1	1.425730
log(surgical_procedures_on_the_digestive_system + 0.01)	4.214368e+00	1	2.052893
log(surgical_procedures_on_the_eye_and_ocular_adnexa + 0.01)	1.247418e+01	1	3.531881
log(surgical_procedures_on_genital_system + 0.01)	4.161321e+00	1	2.039932
log(surgical_procedures_on_the_hemic_and_lymphatic_systems + 0.01)	1.206353e+00	1	1.098341
log(surgical_procedures_on_the_integumentary_system + 0.01)	3.532597e+00	1	1.879520
log(surgical_procedures_on_the_musculoskeletal_system + 0.01)	2.832389e+00	1	1.682970
log(surgical_procedures_on_the_nervous_system + 0.01)	1.472657e+00	1	1.213531
log(surgical_procedures_on_the_respiratory_system + 0.01)	2.403793e+00	1	1.550417
log(surgical_procedures_on_the_urinary_system + 0.01)	9.148825e+00	1	3.024702
log(other_surgical_procedures + 0.01)	1.047109e+00	1	1.023284
log(screening_examinations_and_disease_management_training + 0.01)	1.586582e+00	1	1.259596
log(miscellaneous_diagnostic_and_therapeutic_services + 0.01)	3.322483e+00	1	1.822768
log(initial_services_for_medicare_enrollment + 0.01)	1.472483e+00	1	1.213459
log(gross_and_microscopic_examinations_prostate_biopsy + 0.01)	1.195442e+00	1	1.093363
log(counseling_screening_and_prevention_services + 0.01)	2.377121e+00	1	1.541792
log(miscellaneous_services + 0.01)	1.711179e+00	1	1.308120

Residuals:

Min	1Q	Median	3Q	Max
-8.1552	-0.3421	0.0142	0.3661	4.1897

Coefficients:

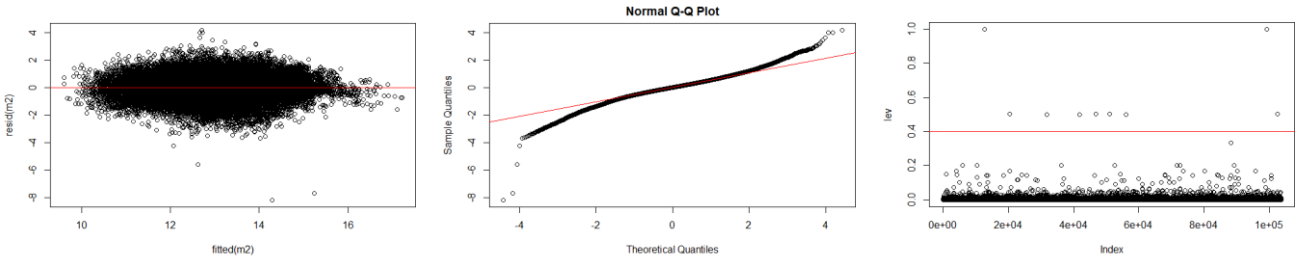
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	12.0269864	0.3357937	35.817	< 2e-16	***
year2013	0.0766963	0.0075282	10.188	< 2e-16	***
year2014	0.1031479	0.0111226	9.274	< 2e-16	***
year2015	0.2156645	0.0112275	19.209	< 2e-16	***
year2016	0.3208013	0.0134399	23.869	< 2e-16	***
year2017	0.2124019	0.0129562	16.394	< 2e-16	***
nppes_provider_stateAE	0.1142558	0.6625909	0.172	0.863094	

nppes_provider_stateAK	0.5912819	0.2525754	2.341	0.019233	*
nppes_provider_stateAL	-0.1686292	0.2525398	-0.668	0.504305	
nppes_provider_stateAP	0.1913265	0.3718237	0.515	0.606860	
nppes_provider_stateAR	-0.0586293	0.2521274	-0.233	0.816121	
nppes_provider_stateAZ	-0.2215489	0.2509011	-0.883	0.377231	
nppes_provider_stateCA	0.0591957	0.2508690	0.236	0.813462	
nppes_provider_stateCO	-0.1234034	0.2517494	-0.490	0.624005	
nppes_provider_stateCT	0.1122515	0.2512612	0.447	0.655055	
nppes_provider_stateDC	0.0053630	0.2522245	0.021	0.983036	
nppes_provider_stateDE	0.0445133	0.2518661	0.177	0.859718	
nppes_provider_stateFL	-0.0106066	0.2509514	-0.042	0.966287	
nppes_provider_stateGA	0.0825272	0.2512120	0.329	0.742522	
nppes_provider_stateGU	0.2161410	0.5008930	0.432	0.666097	
nppes_provider_stateHI	-0.3477835	0.2553468	-1.362	0.173199	
nppes_provider_stateIA	0.0298641	0.2535917	0.118	0.906254	
nppes_provider_stateID	-0.4088803	0.2557289	-1.599	0.109850	
nppes_provider_stateIL	0.1596945	0.2509217	0.636	0.524496	
nppes_provider_stateIN	0.0254386	0.2518952	0.101	0.919560	
nppes_provider_stateKS	0.0203723	0.2517393	0.081	0.935501	
nppes_provider_stateKY	-0.0237101	0.2525496	-0.094	0.925202	
nppes_provider_stateLA	-0.0341628	0.2522421	-0.135	0.892267	
nppes_provider_stateMA	0.0787271	0.2511051	0.314	0.753884	
nppes_provider_stateMD	0.0018616	0.2510342	0.007	0.994083	
nppes_provider_stateME	0.1489166	0.2581275	0.577	0.564001	
nppes_provider_stateMI	-0.1536646	0.2511272	-0.612	0.540606	
nppes_provider_stateMN	-0.3745436	0.2515309	-1.489	0.136476	
nppes_provider_stateMO	0.0473619	0.2518131	0.188	0.850811	
nppes_provider_stateMP	0.7906500	0.6628887	1.193	0.232976	
nppes_provider_stateMS	0.0102538	0.2533652	0.040	0.967718	
nppes_provider_stateMT	-0.0967516	0.2625787	-0.368	0.712526	
nppes_provider_stateNC	0.1260534	0.2512422	0.502	0.615865	
nppes_provider_stateND	0.0181265	0.2585922	0.070	0.944117	
nppes_provider_stateNE	0.0283014	0.2532065	0.112	0.911004	
nppes_provider_stateNH	0.2026490	0.2580155	0.785	0.432212	
nppes_provider_stateNJ	0.0247402	0.2508792	0.099	0.921445	
nppes_provider_stateNM	-0.1322162	0.2520639	-0.525	0.599908	
nppes_provider_stateNV	0.0679822	0.2512163	0.271	0.786690	
nppes_provider_stateNY	0.0641912	0.2509068	0.256	0.798077	
nppes_provider_stateOH	0.0695097	0.2512438	0.277	0.782040	
nppes_provider_stateOK	-0.1221106	0.2509726	-0.487	0.626579	
nppes_provider_stateOR	-0.1384059	0.2518101	-0.550	0.582565	
nppes_provider_statePA	-0.0147334	0.2509872	-0.059	0.953190	
nppes_provider_statePR	-0.4980393	0.2885439	-1.726	0.084343	.
nppes_provider_stateRI	-0.0879586	0.2532530	-0.347	0.728355	
nppes_provider_stateSC	0.0932014	0.2519991	0.370	0.711496	
nppes_provider_stateSD	-0.4086357	0.2587671	-1.579	0.114301	
nppes_provider_stateTN	0.1660525	0.2520470	0.659	0.510016	
nppes_provider_stateTX	0.1098870	0.2509421	0.438	0.661461	
nppes_provider_stateUT	-0.1769395	0.2533035	-0.699	0.484849	
nppes_provider_stateVA	0.0528949	0.2510698	0.211	0.833139	
nppes_provider_stateVI	0.0249831	0.3125220	0.080	0.936285	
nppes_provider_stateVT	0.0832585	0.2678821	0.311	0.755951	
nppes_provider_stateWA	-0.0474937	0.2509517	-0.189	0.849894	
nppes_provider_stateWI	0.6268786	0.2518439	2.489	0.012806	*
nppes_provider_stateWV	0.0463289	0.2562383	0.181	0.856522	
nppes_provider_stateWY	-0.6932231	0.2642686	-2.623	0.008713	**
nppes_provider_stateXX	0.6771907	0.5016840	1.350	0.177072	
provider_typeanesthesiologist assistants	0.0412220	0.3012726	0.137	0.891168	
provider_typeanesthesiology	0.2808274	0.1290311	2.176	0.029525	*
provider_typeaudiologist	-0.8567072	0.1471669	-5.821	5.86e-09	***
provider_typecardiac electrophysiology	0.1232318	0.1230442	1.002	0.316575	
provider_typecardiac surgery	0.3796339	0.1468344	2.585	0.009726	**
provider_typecardiology	-0.0884797	0.1212340	-0.730	0.465498	
provider_typecertified clinical nurse specialist	-0.3880912	0.1782510	-2.177	0.029467	*
provider_typecertified registered nurse anesthetist (crna)	0.0568447	0.1422697	0.400	0.689485	
provider_typechiropractic	-0.4291527	0.4523424	-0.949	0.342758	
provider_typeclinical psychologist	0.0878887	0.1539462	0.571	0.568065	
provider_typecolorectal surgery (proctology)	0.4590469	0.1687461	2.720	0.006523	**
provider_typecritical care (intensivists)	0.1993551	0.1258951	1.584	0.113310	
provider_typecrna	-0.0573492	0.1310992	-0.437	0.661787	
provider_typedermatology	-0.4527770	0.1235169	-3.666	0.000247	***
provider_typediagnostic radiology	-0.5999289	0.1212636	-4.947	7.54e-07	***
provider_typeemergency medicine	-0.0449311	0.1219867	-0.368	0.712630	
provider_typeendocrinology	-0.2158049	0.1229512	-1.755	0.079227	.
provider_typefamily practice	-0.2533358	0.1209871	-2.094	0.036271	*
provider_typegastroenterology	0.2944147	0.1235280	2.383	0.017156	*
provider_typegeneral practice	-0.1978604	0.1272428	-1.555	0.119953	
provider_typegeneral surgery	0.1582115	0.1235403	1.281	0.200321	
provider_typegeriatric medicine	0.0837522	0.1329516	0.630	0.528732	
provider_typegeriatric psychiatry	-0.2689443	0.2399028	-1.121	0.262267	
provider_typegynecological/oncology	0.8336488	0.2639251	3.159	0.001585	**
provider_typehand surgery	0.1179495	0.1442455	0.818	0.413530	
provider_typehematology	0.4484557	0.1454645	3.083	0.002050	**
provider_typehematology/oncology	0.5712812	0.1235230	4.625	3.75e-06	***
provider_typehospice and palliative care	-0.1756162	0.1571546	-1.117	0.263794	
provider_typehospitalist	0.1639188	0.1526631	1.074	0.282947	
provider_typeindependent diagnostic testing facility	-0.2076324	0.3745613	-0.554	0.579351	
provider_typeinfectious disease	0.0402936	0.1228887	0.328	0.742998	
provider_typeinternal medicine	-0.0635402	0.1204898	-0.527	0.597952	
provider_typeinterventional radiology	-0.4548286	0.1232521	-3.690	0.000224	***
provider_typelicensed clinical social worker	0.4101916	0.2869497	1.429	0.152867	
provider_typemaxillofacial surgery	-1.5344535	0.2622185	-5.852	4.88e-09	***
provider_typemedical oncology	0.5346323	0.1293112	4.134	3.56e-05	***
provider_typenephrology	-0.0189698	0.1249605	-0.152	0.879340	
provider_typeneurology	-0.2761892	0.1226836	-2.251	0.024373	*
provider_typeneuropsychiatry	-0.4483856	0.2646506	-1.694	0.090220	.
provider_typeneurosurgery	1.0904045	0.1335987	8.162	3.34e-16	***
provider_typenuclear medicine	-0.1086982	0.1249660	-0.870	0.384400	
provider_typenurse practitioner	-0.1595925	0.1221333	-1.307	0.191315	
provider_typeobstetrics/gynecology	-0.3139102	0.1503183	-2.088	0.036773	*
provider_typeoccupational therapist	1.1047143	0.4522266	2.443	0.014574	*

provider_typeophthalmology	-1.2042475	0.1304954	-9.228	< 2e-16	***
provider_typeoptometry	-1.5421370	0.1318118	-11.700	< 2e-16	***
provider_typeoral surgery (dentist only)	-0.5866661	0.2487791	-2.358	0.018367	*
provider_typeorthopedic surgery	0.6605399	0.1228931	5.375	7.68e-08	***
provider_typeosteopathic manipulative medicine	-0.0402887	0.2035235	-0.198	0.843080	
provider_typeotolaryngology	-0.1741344	0.1235978	-1.409	0.158874	
provider_typepain management	0.4544044	0.1325655	3.428	0.000609	***
provider_typepathology	-0.4571625	0.1238240	-3.692	0.000223	***
provider_typepediatric medicine	-0.4240618	0.1498656	-2.830	0.004661	**
provider_typeperipheral vascular disease	0.1288365	0.1742811	0.739	0.459760	
provider_typephysical medicine and rehabilitation	-0.0204309	0.1235365	-0.165	0.868643	
provider_typephysical therapist	0.5128612	0.1764984	2.906	0.003664	**
provider_typephysician assistant	-0.1494984	0.1235302	-1.210	0.226198	
provider_typeplastic and reconstructive surgery	0.4302277	0.1781623	2.415	0.015745	*
provider_typepodiatry	-0.7651742	0.1235825	-6.192	5.98e-10	***
provider_typepreventive medicine	-0.4914346	0.2284692	-2.151	0.031479	*
provider_typepsychiatry	-0.0710871	0.1334405	-0.533	0.594225	
provider_typepulmonary disease	-0.0003801	0.1211156	-0.003	0.997496	
provider_typeradiation oncology	-0.2915978	0.1336215	-2.182	0.029092	*
provider_typeregistered dietician/nutrition professional	-0.9827464	0.2646795	-3.713	0.000205	***
provider_typerheumatology	0.4054310	0.1269044	3.195	0.001400	**
provider_typesleep medicine	0.1067834	0.1955724	0.546	0.585064	
provider_typesports medicine	0.3649507	0.1713757	2.130	0.033212	*
provider_typesurgical oncology	1.4817268	0.2484376	5.964	2.47e-09	***
provider_typethoracic surgery	0.3004527	0.1518055	1.979	0.047797	*
provider_typeurology	0.0712325	0.1282529	0.555	0.578618	
provider_typevascular surgery	0.2675581	0.1227720	2.179	0.029311	*
log(total_unique_benes)	0.5828783	0.0043913	132.734	< 2e-16	***
log(number_of_hcpcs)	0.2634526	0.0048987	53.780	< 2e-16	***
benef_average_age	0.0105587	0.0012824	8.233	< 2e-16	***
log(benef_female)	-0.2017933	0.0204020	-9.891	< 2e-16	***
log(benef_race_white + 0.01)	0.1437499	0.0081334	17.674	< 2e-16	***
log(benef_race_black + 0.01)	0.0242350	0.0036026	6.727	1.74e-11	***
log(benef_race_api + 0.01)	0.0479643	0.0040939	11.716	< 2e-16	***
log(benef_race_hispanic + 0.01)	0.0420662	0.0038620	10.892	< 2e-16	***
log(benef_race_natind + 0.01)	-0.0036366	0.0072323	-0.503	0.615087	
log(benef_dual + 0.01)	0.0975967	0.0074335	13.129	< 2e-16	***
log(benef_cc_afib + 0.01)	-0.2289775	0.0095901	-23.876	< 2e-16	***
log(benef_cc_alzrdsd)	-0.1390085	0.0089609	-15.513	< 2e-16	***
log(benef_cc_asthma)	-0.1608223	0.0091709	-17.536	< 2e-16	***
log(benef_cc_cancer)	0.0865106	0.0095499	9.059	< 2e-16	***
log(benef_cc_chf)	-0.2132997	0.0130491	-16.346	< 2e-16	***
log(benef_cc_ckd)	-0.1138670	0.0120564	-9.445	< 2e-16	***
log(benef_cc_copd)	-0.0109453	0.0110185	-0.993	0.320539	
log(benef_cc_depr)	-0.0647628	0.0112490	-5.757	8.58e-09	***
log(benef_cc_diab)	-0.0479550	0.0128924	-3.720	0.000200	***
log(benef_cc_hyperl)	0.2896621	0.0234233	12.366	< 2e-16	***
log(benef_cc_hypert)	0.1450866	0.0437406	3.317	0.000910	***
log(benef_cc_ihd)	0.1830120	0.0190720	9.596	< 2e-16	***
log(benef_cc_ost)	-0.0326632	0.0091077	-3.586	0.000336	***
log(benef_cc_raoa)	0.1632097	0.0172954	9.437	< 2e-16	***
log(benef_cc_schiot + 0.01)	-0.1195346	0.0074893	-15.961	< 2e-16	***
log(benef_cc_strk)	0.1302987	0.0086203	15.115	< 2e-16	***
benef_average_risk_score	0.2457659	0.0069176	35.528	< 2e-16	***
log(ambulatory_services + 0.01)	0.0754848	0.0011811	63.912	< 2e-16	***
log(anesthesia + 0.01)	0.0746724	0.0062196	12.006	< 2e-16	***
log(category_iii + 0.01)	0.0325295	0.0036972	8.798	< 2e-16	***
log(drugs_administered_other_than_oral_method + 0.01)	0.0249341	0.0010912	22.850	< 2e-16	***
log(chemotherapy_drugs + 0.01)	0.0876360	0.0022100	39.655	< 2e-16	***
log(vaccines_or_toxoids + 0.01)	-0.0399222	0.0020182	-19.781	< 2e-16	***
log(psychiatry_services_and_procedures + 0.01)	0.0463513	0.0067184	6.899	5.26e-12	***
log(dialysis_services_and_procedures + 0.01)	0.0051427	0.0043692	1.177	0.239177	
log(gastroenterology_procedures + 0.01)	0.0362473	0.0068427	5.297	1.18e-07	***
log(ophthalmology_services_and_procedures + 0.01)	0.0621903	0.0060483	10.282	< 2e-16	***
log(special_otorhinolaryngologic_services_and_procedures + 0.01)	0.0135380	0.0041683	3.248	0.001163	**
log(cardiovascular_procedures + 0.01)	-0.0073654	0.0014804	-4.975	6.53e-07	***
log(non_invasive_vascular_diagnostic_studies + 0.01)	0.0185540	0.0012753	14.549	< 2e-16	***
log(pulmonary_procedures + 0.01)	0.0020872	0.0026735	0.781	0.434986	
log(allergy_and_clinical_immunology_procedures + 0.01)	0.0311406	0.0035183	8.851	< 2e-16	***
log(endocrinology_services + 0.01)	0.0097512	0.0076037	1.282	0.199698	
log(neurology_and_neuromuscular_procedures + 0.01)	0.0368057	0.0024421	15.072	< 2e-16	***
log(highly_complex_drug_or_biologic_agent_administration + 0.01)	-0.0035660	0.0020481	-1.741	0.081664	.
log(therapeutic_procedures + 0.01)	0.0039500	0.0034128	1.157	0.247108	
log(physical_medicine_and_rehabilitation_evaluations + 0.01)	0.0123267	0.0042239	2.918	0.003520	**
log(moderate_conscious_sedation + 0.01)	0.0169562	0.0033897	5.002	5.67e-07	***
log(other_medicine_services_and_procedures + 0.01)	0.0073072	0.0045969	1.590	0.111929	
log(office_or_other_outpatient_services + 0.01)	0.0022643	0.0012661	1.788	0.073706	.
log(hospital_observation_services + 0.01)	0.0073224	0.0019286	3.797	0.000147	***
log(hospital_inpatient_services + 0.01)	0.0193529	0.0017140	11.291	< 2e-16	***
log(emergency_department_services + 0.01)	0.0794753	0.0026625	29.849	< 2e-16	***
log(critical_care_services + 0.01)	0.0433490	0.0016526	26.231	< 2e-16	***
log(nursing_facility_services + 0.01)	0.0227687	0.0014151	16.090	< 2e-16	***
log(home_services + 0.01)	0.0093211	0.0025764	3.618	0.000297	***
log(prolonged_services + 0.01)	0.0171569	0.0031163	5.506	3.69e-08	***
log(other_care_evaluation_and_management_services + 0.01)	0.0183462	0.0030802	5.956	2.59e-09	***
log(organ_or_disease_oriented_panels + 0.01)	0.0028970	0.0024014	1.206	0.227672	
log(drug_assay + 0.01)	0.0602716	0.0080948	7.446	9.72e-14	***
log(urinalysis_procedures + 0.01)	-0.0373142	0.0023772	-15.697	< 2e-16	***
log(chemistry_procedures + 0.01)	0.0005817	0.0015431	0.377	0.706213	
log(hematology_and_coagulation_procedures + 0.01)	0.0061794	0.0014255	4.335	1.46e-05	***
log(immunology_procedures + 0.01)	0.0024393	0.0027807	0.877	0.380354	
log(microbiology_procedures + 0.01)	-0.0336169	0.0029689	-11.323	< 2e-16	***
log(cytopathology_procedures + 0.01)	-0.0034011	0.0025288	-1.345	0.178650	
log(surgical_pathology_procedures + 0.01)	0.0698497	0.0027819	25.108	< 2e-16	***
log(diagnostic_radiology_procedures + 0.01)	0.0148578	0.0015028	9.886	< 2e-16	***
log(diagnostic_ultrasound_procedures + 0.01)	0.0040394	0.0014949	2.702	0.006892	**
log(radiologic_guidance + 0.01)	0.0269125	0.0026300	10.233	< 2e-16	***
log(breast_and_mammography + 0.01)	-0.0114206	0.0020441	-5.587	2.31e-08	***
log(bone_or_joint_studies + 0.01)	-0.0094072	0.0017058	-5.515	3.50e-08	***
log(radiation_oncology_treatment + 0.01)	0.1546038	0.0080742	19.148	< 2e-16	***
log(nuclear_medicine_procedures + 0.01)	-0.0003536	0.0014416	-0.245	0.806243	

```
log(fine_needle_aspiration_biopsy_procedures + 0.01)                0.0185857  0.0037923   4.901 9.56e-07 ***
log(surgical_procedures_on_the_auditory_system + 0.01)             -0.0050072  0.0038535  -1.299 0.193811
log(surgical_procedures_on_the_cardiovascular_system + 0.01)        0.0303972  0.0012804  23.741  < 2e-16 ***
log(surgical_procedures_on_the_digestive_system + 0.01)             0.0258627  0.0033472   7.727 1.11e-14 ***
log(surgical_procedures_on_the_eye_and_ocular_adnexa + 0.01)        0.1710623  0.0037292  45.871  < 2e-16 ***
log(surgical_procedures_on_genital_system + 0.01)                  -0.0275467  0.0076086  -3.620 0.000294 ***
log(surgical_procedures_on_the_hemic_and_lymphatic_systems + 0.01)  0.0178401  0.0049855   3.578 0.000346 ***
log(surgical_procedures_on_the_integumentary_system + 0.01)         0.0063206  0.0022678   2.787 0.005318 **
log(surgical_procedures_on_the_musculoskeletal_system + 0.01)       0.0047257  0.0026040   1.815 0.069561 .
log(surgical_procedures_on_the_nervous_system + 0.01)               0.0428802  0.0027172  15.781  < 2e-16 ***
log(surgical_procedures_on_the_respiratory_system + 0.01)           0.0032207  0.0029313   1.099 0.271897
log(surgical_procedures_on_the_urinary_system + 0.01)               0.0046308  0.0051234   0.904 0.366068
log(other_surgical_procedures + 0.01)                               0.0417089  0.0143277   2.911 0.003603 **
log(screening_examinations_and_disease_management_training + 0.01) -0.0118195  0.0032885  -3.594 0.000326 ***
log(miscellaneous_diagnostic_and_therapeutic_services + 0.01)       0.0052313  0.0015218   3.437 0.000587 ***
log(initial_services_for_medicare_enrollment + 0.01)                0.0034346  0.0059774   0.575 0.565560
log(gross_and_microscopic_examinations_prostate_biopsy + 0.01)      0.0333433  0.0051247   6.506 7.73e-11 ***
log(counseling_screening_and_prevention_services + 0.01)           -0.0017267  0.0028426  -0.607 0.543556
log(miscellaneous_services + 0.01)                                  0.0329315  0.0008487  38.804  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6131 on 103320 degrees of freedom
Multiple R-squared:  0.5987,  Adjusted R-squared:  0.5978
F-statistic: 679.1 on 227 and 103320 DF,  p-value: < 2.2e-16

[1] 192771.2
                                                                        GVIF Df GVIF^(1/(2*Df))
year                                                                3.049958e+01  5        1.407438
nppes_provider_state                                                1.191106e+02 58        1.042068
provider_type                                                       4.749749e+13 71        1.248287
log(total_unique_benes)                                             3.327034e+00  1        1.824016
log(number_of_hcpcs)                                                4.972237e+00  1        2.229851
benef_average_age                                                   4.000266e+00  1        2.000067
log(benef_female)                                                   4.804715e+00  1        2.191966
log(benef_race_white + 0.01)                                        2.384044e+00  1        1.544035
log(benef_race_black + 0.01)                                        2.703474e+00  1        1.644224
log(benef_race_api + 0.01)                                          2.591417e+00  1        1.609788
log(benef_race_hispanic + 0.01)                                     2.770088e+00  1        1.664358
log(benef_race_natind + 0.01)                                       3.175394e+00  1        1.781964
log(benef_dual + 0.01)                                              6.656169e+00  1        2.579955
log(benef_cc_afib + 0.01)                                           5.027064e+00  1        2.242111
log(benef_cc_alzrdsd)                                               6.834911e+00  1        2.614366
log(benef_cc_asthma)                                                5.705660e+00  1        2.388652
log(benef_cc_cancer)                                                3.050780e+00  1        1.746648
log(benef_cc_chf)                                                   1.041894e+01  1        3.227839
log(benef_cc_ckd)                                                   9.282172e+00  1        3.046666
log(benef_cc_copd)                                                  6.610189e+00  1        2.571029
log(benef_cc_depr)                                                  4.830271e+00  1        2.197788
log(benef_cc_diab)                                                  5.898812e+00  1        2.428747
log(benef_cc_hyperl)                                                3.692601e+00  1        1.921614
log(benef_cc_hypert)                                                3.250890e+00  1        1.803023
log(benef_cc_ihd)                                                   9.747511e+00  1        3.122100
log(benef_cc_ost)                                                   2.524315e+00  1        1.588809
log(benef_cc_raoa)                                                  2.996994e+00  1        1.731183
log(benef_cc_schiot + 0.01)                                         5.371165e+00  1        2.317577
log(benef_cc_strk)                                                  5.949484e+00  1        2.439156
benef_average_risk_score                                            6.205251e+00  1        2.491034
log(ambulatory_services + 0.01)                                     2.068057e+00  1        1.438074
log(anesthesia + 0.01)                                              1.182194e+01  1        3.438305
log(category_iii + 0.01)                                            1.070954e+00  1        1.034869
log(drugs_administered_other_than_oral_method + 0.01)               2.669716e+00  1        1.633927
log(chemotherapy_drugs + 0.01)                                      2.245554e+00  1        1.498517
log(vaccines_or_toxoids + 0.01)                                     2.758722e+00  1        1.660940
log(psychiatry_services_and_procedures + 0.01)                      3.323109e+00  1        1.822940
log(dialysis_services_and_procedures + 0.01)                        6.614825e+00  1        2.571930
log(gastroenterology_procedures + 0.01)                             1.234016e+00  1        1.110863
log(ophthalmology_services_and_procedures + 0.01)                   6.110986e+01  1        7.817280
log(special_otorhinolaryngologic_services_and_procedures + 0.01)    2.402202e+00  1        1.549904
log(cardiovascular_procedures + 0.01)                               6.161754e+00  1        2.482288
log(non_invasive_vascular_diagnostic_studies + 0.01)                2.243873e+00  1        1.497956
log(pulmonary_procedures + 0.01)                                    2.862244e+00  1        1.691817
log(allergy_and_clinical_immunology_procedures + 0.01)              1.170284e+00  1        1.081796
log(endocrinology_services + 0.01)                                  1.213573e+00  1        1.101623
log(neurology_and_neuromuscular_procedures + 0.01)                  2.128772e+00  1        1.459031
log(highly_complex_drug_or_biologic_agent_administration + 0.01)    2.887611e+00  1        1.699297
log(therapeutic_procedures + 0.01)                                  1.424943e+00  1        1.193710
log(physical_medicine_and_rehabilitation_evaluations + 0.01)        1.520258e+00  1        1.232987
log(moderate_conscious_sedation + 0.01)                             1.271115e+00  1        1.127437
log(other_medicine_services_and_procedures + 0.01)                  1.252261e+00  1        1.119045
log(office_or_other_outpatient_services + 0.01)                     5.943152e+00  1        2.437858
log(hospital_observation_services + 0.01)                           7.590800e+00  1        2.755141
log(hospital_inpatient_services + 0.01)                             8.741131e+00  1        2.956540
log(emergency_department_services + 0.01)                           6.339414e+00  1        2.517819
log(critical_care_services + 0.01)                                  2.260305e+00  1        1.503431
log(nursing_facility_services + 0.01)                               1.764562e+00  1        1.328368
log(home_services + 0.01)                                           1.361944e+00  1        1.167024
log(prolonged_services + 0.01)                                      1.092081e+00  1        1.045027
log(other_care_evaluation_and_management_services + 0.01)           1.299457e+00  1        1.139937
log(organ_or_disease_oriented_panels + 0.01)                        2.274932e+00  1        1.508288
log(drug_assay + 0.01)                                              1.120253e+00  1        1.058420
log(urinalysis_procedures + 0.01)                                   3.433172e+00  1        1.852882
log(chemistry_procedures + 0.01)                                    2.258097e+00  1        1.502697
log(hematology_and_coagulation_procedures + 0.01)                   2.096224e+00  1        1.447834
log(immunology_procedures + 0.01)                                   1.675815e+00  1        1.294533
log(microbiology_procedures + 0.01)                                 1.381763e+00  1        1.175484
log(cytopathology_procedures + 0.01)                                3.192999e+00  1        1.786896
log(surgical_pathology_procedures + 0.01)                           9.475655e+00  1        3.078255
log(diagnostic_radiology_procedures + 0.01)                         7.462263e+00  1        2.731714
log(diagnostic_ultrasound_procedures + 0.01)                        3.761808e+00  1        1.939538
log(radiologic_guidance + 0.01)                                     2.114130e+00  1        1.454005
```

log(breast_and_mammography + 0.01)	4.314975e+00	1	2.077252
log(bone_or_joint_studies + 0.01)	2.019271e+00	1	1.421011
log(radiation_oncology_treatment + 0.01)	1.391839e+00	1	1.179762
log(nuclear_medicine_procedures + 0.01)	2.207631e+00	1	1.485810
log(fine_needle_aspiration_biopsy_procedures + 0.01)	1.209032e+00	1	1.099560
log(surgical_procedures_on_the_auditory_system + 0.01)	2.422732e+00	1	1.556513
log(surgical_procedures_on_the_cardiovascular_system + 0.01)	2.032705e+00	1	1.425730
log(surgical_procedures_on_the_digestive_system + 0.01)	4.214368e+00	1	2.052893
log(surgical_procedures_on_the_eye_and_ocular_adnexa + 0.01)	1.247418e+01	1	3.531881
log(surgical_procedures_on_genital_system + 0.01)	4.161321e+00	1	2.039932
log(surgical_procedures_on_the_hemic_and_lymphatic_systems + 0.01)	1.206353e+00	1	1.098341
log(surgical_procedures_on_the_integumentary_system + 0.01)	3.532597e+00	1	1.879520
log(surgical_procedures_on_the_musculoskeletal_system + 0.01)	2.832389e+00	1	1.682970
log(surgical_procedures_on_the_nervous_system + 0.01)	1.472657e+00	1	1.213531
log(surgical_procedures_on_the_respiratory_system + 0.01)	2.403793e+00	1	1.550417
log(surgical_procedures_on_the_urinary_system + 0.01)	9.148825e+00	1	3.024702
log(other_surgical_procedures + 0.01)	1.047109e+00	1	1.023284
log(screening_examinations_and_disease_management_training + 0.01)	1.586582e+00	1	1.259596
log(miscellaneous_diagnostic_and_therapeutic_services + 0.01)	3.322483e+00	1	1.822768
log(initial_services_for_medicare_enrollment + 0.01)	1.472483e+00	1	1.213459
log(gross_and_microscopic_examinations_prostate_biopsy + 0.01)	1.195442e+00	1	1.093363
log(counseling_screening_and_prevention_services + 0.01)	2.377121e+00	1	1.541792
log(miscellaneous_services + 0.01)	1.711179e+00	1	1.308120



Model 2 validation

```
m2_train <- lm(log(overcharge) ~ year + nppes_provider_state + provider_type + log(total_unique_benes) +
log(number_of_hcpcs) + benef_average_age + log(benef_female) + log(benef_race_white+0.01) +
log(benef_race_black+0.01) + log(benef_race_api+0.01) + log(benef_race_hispanic+0.01) +
log(benef_race_natind+0.01) + log(benef_dual+0.01) + log(benef_cc_afib+0.01) + log(benef_cc_alzrdsd) +
log(benef_cc_asthma) + log(benef_cc_cancer) + log(benef_cc_chf) + log(benef_cc_ckd) + log(benef_cc_copd) +
log(benef_cc_depr) + log(benef_cc_diab) + log(benef_cc_hyperl) + log(benef_cc_hypert) + log(benef_cc_ihd) +
log(benef_cc_ost) + log(benef_cc_raoa) + log(benef_cc_schiot+0.01) + log(benef_cc_strk) +
benef_average_risk_score + log(ambulatory_services+0.01) + log(anesthesia+0.01) + log(category_iii+0.01) +
log(drugs_administered_other_than_oral_method+0.01) + log(chemotherapy_drugs+0.01) +
log(vaccines_or_toxoids+0.01) + log(psychiatry_services_and_procedures+0.01) +
log(dialysis_services_and_procedures+0.01) + log(gastroenterology_procedures+0.01) +
log(ophthalmology_services_and_procedures+0.01) +
log(special_otorhinolaryngologic_services_and_procedures+0.01) + log(cardiovascular_procedures+0.01) +
log(non_invasive_vascular_diagnostic_studies+0.01) + log(pulmonary_procedures+0.01) +
log(allergy_and_clinical_immunology_procedures+0.01) + log(endocrinology_services+0.01) +
log(neurology_and_neuromuscular_procedures+0.01) +
log(highly_complex_drug_or_biologic_agent_administration+0.01) + log(therapeutic_procedures+0.01) +
log(physical_medicine_and_rehabilitation_evaluations+0.01) + log(moderate_conscious_sedation+0.01) +
log(other_medicine_services_and_procedures+0.01) + log(office_or_other_outpatient_services+0.01) +
log(hospital_observation_services+0.01) + log(hospital_inpatient_services+0.01) +
log(emergency_department_services+0.01) + log(critical_care_services+0.01) +
log(nursing_facility_services+0.01) + log(home_services+0.01) + log(prolonged_services+0.01) +
log(other_care_evaluation_and_management_services+0.01) + log(organ_or_disease_oriented_panels+0.01) +
log(drug_assay+0.01) + log(urinalysis_procedures+0.01) + log(chemistry_procedures+0.01) +
log(hematology_and_coagulation_procedures+0.01) + log(immunology_procedures+0.01) +
log(microbiology_procedures+0.01) + log(cytopathology_procedures+0.01) +
log(surgical_pathology_procedures+0.01) + log(diagnostic_radiology_procedures+0.01) +
log(diagnostic_ultrasound_procedures+0.01) + log(radiologic_guidance+0.01) + log(breast_and_mammography+0.01) +
log(bone_or_joint_studies+0.01) + log(radiation_oncology_treatment+0.01) +
log(nuclear_medicine_procedures+0.01) + log(fine_needle_aspiration_biopsy_procedures+0.01) +
log(surgical_procedures_on_the_auditory_system+0.01) +
log(surgical_procedures_on_the_cardiovascular_system+0.01) +
log(surgical_procedures_on_the_digestive_system+0.01) +
log(surgical_procedures_on_the_eye_and_ocular_adnexa+0.01) + log(surgical_procedures_on_genital_system+0.01) +
log(surgical_procedures_on_the_hemic_and_lymphatic_systems+0.01) +
log(surgical_procedures_on_the_integumentary_system+0.01) +
log(surgical_procedures_on_the_musculoskeletal_system+0.01) +
log(surgical_procedures_on_the_nervous_system+0.01) + log(surgical_procedures_on_the_respiratory_system+0.01) +
log(surgical_procedures_on_the_urinary_system+0.01) + log(other_surgical_procedures+0.01) +
log(screening_examinations_and_disease_management_training+0.01) +
log(miscellaneous_diagnostic_and_therapeutic_services+0.01) +
log(initial_services_for_medicare_enrollment+0.01) +
log(gross_and_microscopic_examinations_prostate_biopsy+0.01) +
log(counseling_screening_and_prevention_services+0.01) + log(miscellaneous_services+0.01), data=df_train)
m2_pred <- predict(m2_train, df_test)
sqrt(mean((df_test$overcharge - exp(m2_pred)) ^ 2))
```

```
[1] 769823.8
```

Model 3

```
temp <- lm(abs(residuals(m2)) ~ fitted(m2))
wt <- c(1/temp$fitted^2)
m3 <- lmer(log(overcharge) ~ year + nppes_provider_state + (1 | provider_type) + log(total_unique_benes) +
log(number_of_hcpcs) + benef_average_age + log(benef_female) + log(benef_race_white+0.01) +
log(benef_race_black+0.01) + log(benef_race_api+0.01) + log(benef_race_hispanic+0.01) +
log(benef_race_natind+0.01) + log(benef_dual+0.01) + log(benef_cc_afib+0.01) + log(benef_cc_alzrdsd) +
log(benef_cc_asthma) + log(benef_cc_cancer) + log(benef_cc_chf) + log(benef_cc_ckd) + log(benef_cc_copd) +
log(benef_cc_depr) + log(benef_cc_diab) + log(benef_cc_hyperl) + log(benef_cc_hypert) + log(benef_cc_ihd) +
log(benef_cc_ost) + log(benef_cc_raoa) + log(benef_cc_schiot+0.01) + log(benef_cc_strk) +
benef_average_risk_score + log(ambulatory_services+0.01) + log(anesthesia+0.01) + log(category_iii+0.01) +
log(drugs_administered_other_than_oral_method+0.01) + log(chemotherapy_drugs+0.01) +
log(vaccines_or_toxoids+0.01) + log(psychiatry_services_and_procedures+0.01) +
log(dialysis_services_and_procedures+0.01) + log(gastroenterology_procedures+0.01) +
log(ophthalmology_services_and_procedures+0.01) +
log(special_otorhinolaryngologic_services_and_procedures+0.01) + log(cardiovascular_procedures+0.01) +
log(non_invasive_vascular_diagnostic_studies+0.01) + log(pulmonary_procedures+0.01) +
log(allergy_and_clinical_immunology_procedures+0.01) + log(endocrinology_services+0.01) +
log(neurology_and_neuromuscular_procedures+0.01) +
log(highly_complex_drug_or_biologic_agent_administration+0.01) + log(therapeutic_procedures+0.01) +
log(physical_medicine_and_rehabilitation_evaluations+0.01) + log(moderate_conscious_sedation+0.01) +
log(other_medicine_services_and_procedures+0.01) + log(office_or_other_outpatient_services+0.01) +
log(hospital_observation_services+0.01) + log(hospital_inpatient_services+0.01) +
log(emergency_department_services+0.01) + log(critical_care_services+0.01) +
log(nursing_facility_services+0.01) + log(home_services+0.01) + log(prolonged_services+0.01) +
log(other_care_evaluation_and_management_services+0.01) + log(organ_or_disease_oriented_panels+0.01) +
log(drug_assay+0.01) + log(urinalysis_procedures+0.01) + log(chemistry_procedures+0.01) +
log(hematology_and_coagulation_procedures+0.01) + log(immunology_procedures+0.01) +
log(microbiology_procedures+0.01) + log(cytopathology_procedures+0.01) +
log(surgical_pathology_procedures+0.01) + log(diagnostic_radiology_procedures+0.01) +
log(diagnostic_ultrasound_procedures+0.01) + log(radiologic_guidance+0.01) + log(breast_and_mammography+0.01) +
log(bone_or_joint_studies+0.01) + log(radiation_oncology_treatment+0.01) +
log(nuclear_medicine_procedures+0.01) + log(fine_needle_aspiration_biopsy_procedures+0.01) +
log(surgical_procedures_on_the_auditory_system+0.01) +
log(surgical_procedures_on_the_cardiovascular_system+0.01) +
log(surgical_procedures_on_the_digestive_system+0.01) +
log(surgical_procedures_on_the_eye_and_ocular_adnexa+0.01) + log(surgical_procedures_on_genital_system+0.01) +
log(surgical_procedures_on_the_hemic_and_lymphatic_systems+0.01) +
log(surgical_procedures_on_the_integumentary_system+0.01) +
log(surgical_procedures_on_the_musculoskeletal_system+0.01) +
log(surgical_procedures_on_the_nervous_system+0.01) + log(surgical_procedures_on_the_respiratory_system+0.01) +
log(surgical_procedures_on_the_urinary_system+0.01) + log(other_surgical_procedures+0.01) +
log(screening_examinations_and_disease_management_training+0.01) +
log(miscellaneous_diagnostic_and_therapeutic_services+0.01) +
log(initial_services_for_medicare_enrollment+0.01) +
log(gross_and_microscopic_examinations_prostate_biopsy+0.01) +
log(counseling_screening_and_prevention_services+0.01) + log(miscellaneous_services+0.01), REML=FALSE,
data=d_model, weights=wt)
summary(m3)
AIC(m3)
ranef(m3)
plot(resid(m3) ~ log(d_model$overcharge))
plot(resid(m3) ~ fitted(m3))
abline(0,0, col='red')
qqnorm(resid(m3))
qqline(resid(m3), col="red")
lev = hat(model.matrix(m3))
plot(lev)
abline(mean(lev)*5,0,col="red")
d_model[lev>mean(lev)*5,]
vif(m3)
```

```
     AIC      BIC   logLik deviance df.resid
193118.6 194636.7 -96400.3 192800.6   103389

Scaled residuals:
     Min       1Q   Median       3Q      Max
-13.8137  -0.5590   0.0233   0.5978   6.7608

Random effects:
 Groups        Name        Variance Std.Dev.
 provider_type (Intercept) 0.2353   0.4851
 Residual                  1.7808   1.3345
Number of obs: 103548, groups:  provider_type, 72

Fixed effects:
                                                                  Estimate Std. Error t value
(Intercept)                                                     11.9415121  0.3197992  37.341
year2013                                                         0.0772310  0.0075399  10.243
year2014                                                         0.1027269  0.0111572   9.207
```

year2015	0.2171948	0.0112559	19.296
year2016	0.3229592	0.0134762	23.965
year2017	0.2119554	0.0129951	16.310
nppes_provider_stateAE	0.1219579	0.6606634	0.185
nppes_provider_stateAK	0.5947145	0.2527266	2.353
nppes_provider_stateAL	-0.1664544	0.2528104	-0.658
nppes_provider_stateAP	0.1958217	0.3696064	0.530
nppes_provider_stateAR	-0.0588599	0.2523245	-0.233
nppes_provider_stateAZ	-0.2143792	0.2511065	-0.854
nppes_provider_stateCA	0.0622050	0.2510761	0.248
nppes_provider_stateCO	-0.1208025	0.2519641	-0.479
nppes_provider_stateCT	0.1133529	0.2514705	0.451
nppes_provider_stateDC	0.0085843	0.2524504	0.034
nppes_provider_stateDE	0.0472481	0.2520533	0.187
nppes_provider_stateFL	-0.0123336	0.2511576	-0.049
nppes_provider_stateGA	0.0839243	0.2514186	0.334
nppes_provider_stateGU	0.2311435	0.4955196	0.466
nppes_provider_stateHI	-0.3515377	0.2557751	-1.374
nppes_provider_stateIA	0.0335478	0.2537605	0.132
nppes_provider_stateID	-0.4148988	0.2560303	-1.621
nppes_provider_stateIL	0.1639382	0.2511294	0.653
nppes_provider_stateIN	0.0292432	0.2521079	0.116
nppes_provider_stateKS	0.0255832	0.2519263	0.102
nppes_provider_stateKY	-0.0187402	0.2528127	-0.074
nppes_provider_stateLA	-0.0363876	0.2524763	-0.144
nppes_provider_stateMA	0.0805359	0.2513179	0.320
nppes_provider_stateMD	0.0045047	0.2512418	0.018
nppes_provider_stateME	0.1663852	0.2583956	0.644
nppes_provider_stateMI	-0.1528332	0.2513399	-0.608
nppes_provider_stateMN	-0.3772606	0.2517648	-1.498
nppes_provider_stateMO	0.0528153	0.2520221	0.210
nppes_provider_stateMP	0.7862846	0.6768403	1.162
nppes_provider_stateMS	0.0119169	0.2537203	0.047
nppes_provider_stateMT	-0.0970443	0.2628698	-0.369
nppes_provider_stateNC	0.1311904	0.2514427	0.522
nppes_provider_stateND	0.0229407	0.2588031	0.089
nppes_provider_stateNE	0.0358837	0.2533699	0.142
nppes_provider_stateNH	0.2077711	0.2582716	0.804
nppes_provider_stateNJ	0.0285384	0.2510868	0.114
nppes_provider_stateNM	-0.1228249	0.2522769	-0.487
nppes_provider_stateNV	0.0685725	0.2514120	0.273
nppes_provider_stateNY	0.0647772	0.2511158	0.258
nppes_provider_stateOH	0.0739567	0.2514577	0.294
nppes_provider_stateOK	-0.1245729	0.2511788	-0.496
nppes_provider_stateOR	-0.1310186	0.2520303	-0.520
nppes_provider_statePA	-0.0091857	0.2511970	-0.037
nppes_provider_statePR	-0.4831447	0.2920765	-1.654
nppes_provider_stateRI	-0.0859552	0.2534760	-0.339
nppes_provider_stateSC	0.0950440	0.2521864	0.377
nppes_provider_stateSD	-0.3993045	0.2592040	-1.541
nppes_provider_stateTN	0.1711096	0.2522553	0.678
nppes_provider_stateTX	0.1131623	0.2511482	0.451
nppes_provider_stateUT	-0.1739650	0.2535768	-0.686
nppes_provider_stateVA	0.0580346	0.2512773	0.231
nppes_provider_stateVI	0.0312512	0.3116901	0.100
nppes_provider_stateVT	0.0672810	0.2686112	0.250
nppes_provider_stateWA	-0.0417192	0.2511601	-0.166
nppes_provider_stateWI	0.6318973	0.2520082	2.507
nppes_provider_stateWV	0.0567236	0.2566531	0.221
nppes_provider_stateWY	-0.6929205	0.2643641	-2.621
nppes_provider_stateXX	0.6837499	0.5005517	1.366
log(total_unique_benes)	0.5777630	0.0044013	131.271
log(number_of_hcpcs)	0.2642716	0.0049222	53.690
benef_average_age	0.0106426	0.0012931	8.230
log(benef_female)	-0.1923242	0.0202974	-9.475
log(benef_race_white + 0.01)	0.1479062	0.0082784	17.867
log(benef_race_black + 0.01)	0.0241795	0.0036151	6.689
log(benef_race_api + 0.01)	0.0475688	0.0041237	11.536
log(benef_race_hispanic + 0.01)	0.0411764	0.0038808	10.610
log(benef_race_natind + 0.01)	-0.0034616	0.0072585	-0.477
log(benef_dual + 0.01)	0.0994678	0.0074448	13.361
log(benef_cc_afib + 0.01)	-0.2325080	0.0096467	-24.102
log(benef_cc_alzrdsd)	-0.1400841	0.0090197	-15.531
log(benef_cc_asthma)	-0.1636728	0.0092071	-17.777
log(benef_cc_cancer)	0.0861905	0.0095812	8.996
log(benef_cc_chf)	-0.2159724	0.0130644	-16.531
log(benef_cc_ckd)	-0.1166440	0.0120919	-9.646
log(benef_cc_copd)	-0.0113617	0.0110579	-1.027
log(benef_cc_depr)	-0.0688159	0.0112821	-6.100
log(benef_cc_diab)	-0.0508232	0.0129148	-3.935
log(benef_cc_hyperl)	0.2920133	0.0236340	12.356
log(benef_cc_hypert)	0.1650620	0.0438629	3.763
log(benef_cc_ihd)	0.1929134	0.0191224	10.088
log(benef_cc_ost)	-0.0341866	0.0091523	-3.735
log(benef_cc_raoa)	0.1645013	0.0173746	9.468
log(benef_cc_schiot + 0.01)	-0.1197353	0.0075211	-15.920
log(benef_cc_strk)	0.1295818	0.0086626	14.959
benef_average_risk_score	0.2512379	0.0068433	36.713
log(ambulatory_services + 0.01)	0.0749931	0.0011665	64.291
log(anesthesia + 0.01)	0.0756121	0.0061077	12.380
log(category_iii + 0.01)	0.0324979	0.0036345	8.942
log(drugs_administered_other_than_oral_method + 0.01)	0.0256415	0.0010782	23.781
log(chemotherapy_drugs + 0.01)	0.0876888	0.0021382	41.011
log(vaccines_or_toxoids + 0.01)	-0.0401644	0.0020278	-19.807
log(psychiatry_services_and_procedures + 0.01)	0.0457047	0.0066796	6.842
log(dialysis_services_and_procedures + 0.01)	0.0032211	0.0043645	0.738
log(gastroenterology_procedures + 0.01)	0.0365174	0.0067748	5.390
log(ophthalmology_services_and_procedures + 0.01)	0.0580359	0.0059624	9.734
log(special_otorhinolaryngologic_services_and_procedures + 0.01)	0.0127278	0.0041638	3.057
log(cardiovascular_procedures + 0.01)	-0.0073179	0.0014891	-4.914
log(non_invasive_vascular_diagnostic_studies + 0.01)	0.0185277	0.0012691	14.599

log(pulmonary_procedures + 0.01)	0.0020328	0.0026846	0.757
log(allergy_and_clinical_immunology_procedures + 0.01)	0.0305460	0.0034786	8.781
log(endocrinology_services + 0.01)	0.0103970	0.0076974	1.351
log(neurology_and_neuromuscular_procedures + 0.01)	0.0364681	0.0024367	14.966
log(highly_complex_drug_or_biologic_agent_administration + 0.01)	-0.0034414	0.0020410	-1.686
log(therapeutic_procedures + 0.01)	0.0040413	0.0033552	1.204
log(physical_medicine_and_rehabilitation_evaluations + 0.01)	0.0122350	0.0042513	2.878
log(moderate_conscious_sedation + 0.01)	0.0166696	0.0033272	5.010
log(other_medicine_services_and_procedures + 0.01)	0.0054744	0.0045513	1.203
log(office_or_other_outpatient_services + 0.01)	0.0028088	0.0012722	2.208
log(hospital_observation_services + 0.01)	0.0073145	0.0019427	3.765
log(hospital_inpatient_services + 0.01)	0.0192797	0.0017292	11.149
log(emergency_department_services + 0.01)	0.0784664	0.0026824	29.252
log(critical_care_services + 0.01)	0.0435287	0.0016604	26.215
log(nursing_facility_services + 0.01)	0.0225093	0.0014356	15.679
log(home_services + 0.01)	0.0093101	0.0026328	3.536
log(prolonged_services + 0.01)	0.0170674	0.0031109	5.486
log(other_care_evaluation_and_management_services + 0.01)	0.0172317	0.0030803	5.594
log(organ_or_disease_oriented_panels + 0.01)	0.0034843	0.0023742	1.468
log(drug_assay + 0.01)	0.0619454	0.0077831	7.959
log(urinalysis_procedures + 0.01)	-0.0374216	0.0023763	-15.748
log(chemistry_procedures + 0.01)	0.0001005	0.0015326	0.066
log(hematology_and_coagulation_procedures + 0.01)	0.0059861	0.0014157	4.228
log(immunology_procedures + 0.01)	0.0027641	0.0027655	0.999
log(microbiology_procedures + 0.01)	-0.0335111	0.0029455	-11.377
log(cytopathology_procedures + 0.01)	-0.0040297	0.0025144	-1.603
log(surgical_pathology_procedures + 0.01)	0.0677080	0.0027507	24.615
log(diagnostic_radiology_procedures + 0.01)	0.0159396	0.0014953	10.660
log(diagnostic_ultrasound_procedures + 0.01)	0.0039228	0.0014887	2.635
log(radiologic_guidance + 0.01)	0.0265974	0.0026013	10.225
log(breast_and_mammography + 0.01)	-0.0108684	0.0020352	-5.340
log(bone_or_joint_studies + 0.01)	-0.0095141	0.0017008	-5.594
log(radiation_oncology_treatment + 0.01)	0.1502765	0.0077488	19.393
log(nuclear_medicine_procedures + 0.01)	-0.0007630	0.0014343	-0.532
log(fine_needle_aspiration_biopsy_procedures + 0.01)	0.0181876	0.0037772	4.815
log(surgical_procedures_on_the_auditory_system + 0.01)	-0.0054524	0.0038815	-1.405
log(surgical_procedures_on_the_cardiovascular_system + 0.01)	0.0306522	0.0012745	24.051
log(surgical_procedures_on_the_digestive_system + 0.01)	0.0246313	0.0033282	7.401
log(surgical_procedures_on_the_eye_and_ocular_adnexa + 0.01)	0.1752124	0.0037864	46.275
log(surgical_procedures_on_genital_system + 0.01)	-0.0260352	0.0075083	-3.468
log(surgical_procedures_on_the_hemic_and_lymphatic_systems + 0.01)	0.0181111	0.0049012	3.695
log(surgical_procedures_on_the_integumentary_system + 0.01)	0.0051556	0.0022629	2.278
log(surgical_procedures_on_the_musculoskeletal_system + 0.01)	0.0052271	0.0025986	2.012
log(surgical_procedures_on_the_nervous_system + 0.01)	0.0419748	0.0026811	15.656
log(surgical_procedures_on_the_respiratory_system + 0.01)	0.0027522	0.0029210	0.942
log(surgical_procedures_on_the_urinary_system + 0.01)	0.0045609	0.0050818	0.898
log(other_surgical_procedures + 0.01)	0.0424547	0.0141587	2.998
log(screening_examinations_and_disease_management_training + 0.01)	-0.0119340	0.0032883	-3.629
log(miscellaneous_diagnostic_and_therapeutic_services + 0.01)	0.0049723	0.0015202	3.271
log(initial_services_for_medicare_enrollment + 0.01)	0.0039324	0.0060469	0.650
log(gross_and_microscopic_examinations_prostate_biopsy + 0.01)	0.0345632	0.0050734	6.813
log(counseling_screening_and_prevention_services + 0.01)	-0.0020761	0.0028749	-0.722
log(miscellaneous_services + 0.01)	0.0329870	0.0008393	39.303

Correlation matrix not shown by default, as p = 157 > 12.
Use print(x, correlation=TRUE) or
    vcov(x)   if you need it

[1] 193118.6
$provider_type

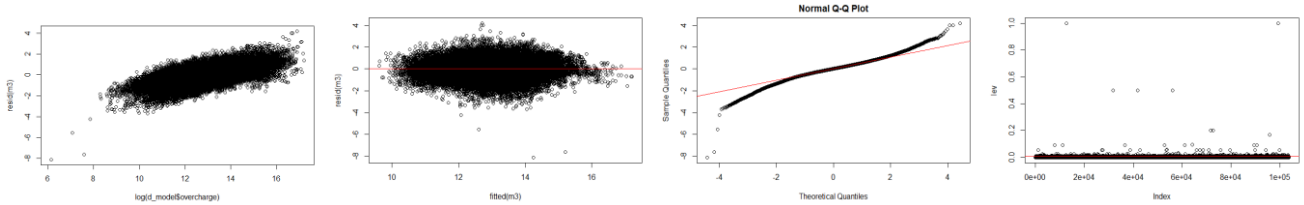
	(Intercept)
allergy/immunology	0.03136011
anesthesiologist assistants	0.04616785
anesthesiology	0.30120087
audiologist	-0.79116560
cardiac electrophysiology	0.15276704
cardiac surgery	0.38573623
cardiology	-0.05811840
certified clinical nurse specialist	-0.32771169
certified registered nurse anesthetist (crna)	0.07219437
chiropractic	-0.20739797
clinical psychologist	0.11772002
colorectal surgery (proctology)	0.45754773
critical care (intensivists)	0.23065984
crna	-0.04290741
dermatology	-0.41125543
diagnostic radiology	-0.56597032
emergency medicine	-0.00507126
endocrinology	-0.18875168
family practice	-0.22571156
gastroenterology	0.33176874
general practice	-0.16908663
general surgery	0.19063986
geriatric medicine	0.10985439
geriatric psychiatry	-0.18251416
gynecological/oncology	0.69944820
hand surgery	0.12742565
hematology	0.47806413
hematology/oncology	0.59520374
hospice and palliative care	-0.15534833
hospitalist	0.19135626
independent diagnostic testing facility	-0.11095159
infectious disease	0.06988871
internal medicine	-0.03121090
interventional radiology	-0.42662239
licensed clinical social worker	0.33885434
maxillofacial surgery	-1.19233833
medical oncology	0.55526972
nephrology	0.01223272
neurology	-0.24632719
neuropsychiatry	-0.33626772

neurosurgery	1.11328026
nuclear medicine	-0.06487812
nurse practitioner	-0.12981674
obstetrics/gynecology	-0.27898845
occupational therapist	0.60617050
ophthalmology	-1.16068524
optometry	-1.49143953
oral surgery (dentist only)	-0.45703429
orthopedic surgery	0.67505398
osteopathic manipulative medicine	0.01152265
otolaryngology	-0.13414879
pain management	0.47920823
pathology	-0.39512693
pediatric medicine	-0.36555683
peripheral vascular disease	0.16012449
physical medicine and rehabilitation	0.01114764
physical therapist	0.50475159
physician assistant	-0.11776099
plastic and reconstructive surgery	0.42106763
podiatry	-0.73176967
preventive medicine	-0.39796224
psychiatry	-0.03259453
pulmonary disease	0.02892163
radiation oncology	-0.23379388
registered dietician/nutrition professional	-0.75674131
rheumatology	0.42893932
sleep medicine	0.13536322
sports medicine	0.36360765
surgical oncology	1.27939593
thoracic surgery	0.32137491
urology	0.09379586
vascular surgery	0.29394012

with conditional variances for “provider_type”

	GVIF	Df	GVIF^(1/(2*Df))
year	30.767993	5	1.408672
nppes_provider_state	86.623046	58	1.039211
log(total_unique_benes)	1.686980	1	1.298838
log(number_of_hcpcs)	2.121618	1	1.456577
benef_average_age	3.254637	1	1.804061
log(benef_female)	1.890663	1	1.375014
log(benef_race_white + 0.01)	2.300752	1	1.516823
log(benef_race_black + 0.01)	2.595503	1	1.611057
log(benef_race_api + 0.01)	2.384724	1	1.544255
log(benef_race_hispanic + 0.01)	2.659278	1	1.630729
log(benef_race_natind + 0.01)	2.828607	1	1.681846
log(benef_dual + 0.01)	5.588017	1	2.363899
log(benef_cc_afib + 0.01)	3.339791	1	1.827509
log(benef_cc_alzrdsd)	5.050361	1	2.247301
log(benef_cc_asthma)	4.641904	1	2.154508
log(benef_cc_cancer)	1.868485	1	1.366925
log(benef_cc_chf)	6.366762	1	2.523244
log(benef_cc_ckd)	7.149241	1	2.673806
log(benef_cc_copd)	4.246919	1	2.060805
log(benef_cc_depr)	3.778763	1	1.943904
log(benef_cc_diab)	4.689870	1	2.165611
log(benef_cc_hyperl)	3.023518	1	1.738827
log(benef_cc_hypert)	2.774215	1	1.665598
log(benef_cc_ihd)	6.151616	1	2.480245
log(benef_cc_ost)	2.030357	1	1.424906
log(benef_cc_raoa)	2.259051	1	1.503014
log(benef_cc_schiot + 0.01)	3.748691	1	1.936154
log(benef_cc_strk)	3.838915	1	1.959315
benef_average_risk_score	3.563974	1	1.887849
log(ambulatory_services + 0.01)	1.699008	1	1.303460
log(anesthesia + 0.01)	1.153650	1	1.074081
log(category_iii + 0.01)	1.040438	1	1.020019
log(drugs_administered_other_than_oral_method + 0.01)	1.836494	1	1.355173
log(chemotherapy_drugs + 0.01)	1.210185	1	1.100084
log(vaccines_or_toxoids + 0.01)	1.788330	1	1.337285
log(psychiatry_services_and_procedures + 0.01)	1.028862	1	1.014328
log(dialysis_services_and_procedures + 0.01)	1.103794	1	1.050616
log(gastroenterology_procedures + 0.01)	1.026305	1	1.013067
log(ophthalmology_services_and_procedures + 0.01)	1.138004	1	1.066772
log(special_otorhinolaryngologic_services_and_procedures + 0.01)	1.073151	1	1.035930
log(cardiovascular_procedures + 0.01)	1.215513	1	1.102503
log(non_invasive_vascular_diagnostic_studies + 0.01)	1.354247	1	1.163721
log(pulmonary_procedures + 0.01)	1.218955	1	1.104063
log(allergy_and_clinical_immunology_procedures + 0.01)	1.026144	1	1.012988
log(endocrinology_services + 0.01)	1.027064	1	1.013442
log(neurology_and_neuromuscular_procedures + 0.01)	1.090619	1	1.044327
log(highly_complex_drug_or_biologic_agent_administration + 0.01)	1.601551	1	1.265524
log(therapeutic_procedures + 0.01)	1.153856	1	1.074177
log(physical_medicine_and_rehabilitation_evaluations + 0.01)	1.137475	1	1.066525
log(moderate_conscious_sedation + 0.01)	1.218665	1	1.103932
log(other_medicine_services_and_procedures + 0.01)	1.034292	1	1.017002
log(office_or_other_outpatient_services + 0.01)	1.503213	1	1.226056
log(hospital_observation_services + 0.01)	2.565432	1	1.601697
log(hospital_inpatient_services + 0.01)	2.577510	1	1.605463
log(emergency_department_services + 0.01)	1.146423	1	1.070711
log(critical_care_services + 0.01)	1.123607	1	1.060003
log(nursing_facility_services + 0.01)	1.324694	1	1.150954
log(home_services + 0.01)	1.147684	1	1.071300
log(prolonged_services + 0.01)	1.044353	1	1.021936
log(other_care_evaluation_and_management_services + 0.01)	1.191771	1	1.091683
log(organ_or_disease_oriented_panels + 0.01)	2.012194	1	1.418518
log(drug_assay + 0.01)	1.098828	1	1.048250
log(urinalysis_procedures + 0.01)	1.663079	1	1.289604
log(chemistry_procedures + 0.01)	1.935854	1	1.391350
log(hematology_and_coagulation_procedures + 0.01)	1.525495	1	1.235109
log(immunology_procedures + 0.01)	1.419688	1	1.191507

log(microbiology_procedures + 0.01)	1.262130	1	1.123446
log(cytopathology_procedures + 0.01)	1.112887	1	1.054934
log(surgical_pathology_procedures + 0.01)	1.105158	1	1.051265
log(diagnostic_radiology_procedures + 0.01)	1.273801	1	1.128628
log(diagnostic_ultrasound_procedures + 0.01)	1.410126	1	1.187487
log(radiologic_guidance + 0.01)	1.668611	1	1.291747
log(breast_and_mammography + 0.01)	2.922011	1	1.709389
log(bone_or_joint_studies + 0.01)	1.399713	1	1.183095
log(radiation_oncology_treatment + 0.01)	1.046725	1	1.023096
log(nuclear_medicine_procedures + 0.01)	1.382020	1	1.175594
log(fine_needle_aspiration_biopsy_procedures + 0.01)	1.123331	1	1.059873
log(surgical_procedures_on_the_auditory_system + 0.01)	1.164498	1	1.079119
log(surgical_procedures_on_the_cardiovascular_system + 0.01)	1.569044	1	1.252615
log(surgical_procedures_on_the_digestive_system + 0.01)	1.476922	1	1.215287
log(surgical_procedures_on_the_eye_and_ocular_adnexa + 0.01)	1.174226	1	1.083617
log(surgical_procedures_on_genital_system + 0.01)	1.068377	1	1.033623
log(surgical_procedures_on_the_hemic_and_lymphatic_systems + 0.01)	1.072369	1	1.035552
log(surgical_procedures_on_the_integumentary_system + 0.01)	1.291037	1	1.136238
log(surgical_procedures_on_the_musculoskeletal_system + 0.01)	1.316658	1	1.147457
log(surgical_procedures_on_the_nervous_system + 0.01)	1.201006	1	1.095904
log(surgical_procedures_on_the_respiratory_system + 0.01)	1.446242	1	1.202598
log(surgical_procedures_on_the_urinary_system + 0.01)	1.107677	1	1.052462
log(other_surgical_procedures + 0.01)	1.013388	1	1.006672
log(screening_examinations_and_disease_management_training + 0.01)	1.184338	1	1.088273
log(miscellaneous_diagnostic_and_therapeutic_services + 0.01)	2.557841	1	1.599325
log(initial_services_for_medicare_enrollment + 0.01)	1.364182	1	1.167982
log(gross_and_microscopic_examinations_prostate_biopsy + 0.01)	1.074223	1	1.036447
log(counseling_screening_and_prevention_services + 0.01)	1.810699	1	1.345622
log(miscellaneous_services + 0.01)	1.493703	1	1.222171



Model 3 validation

```
temp <- lm(abs(residuals(m2_train)) ~ fitted(m2_train))
wt <- c(1/temp$fitted^2)
m3_train <- lmer(log(overcharge) ~ year + nppes_provider_state + (1 | provider_type) + log(total_unique_benes)
+ log(number_of_hcpcs) + benef_average_age + log(benef_female) + log(benef_race_white+0.01) +
log(benef_race_black+0.01) + log(benef_race_api+0.01) + log(benef_race_hispanic+0.01) +
log(benef_race_natind+0.01) + log(benef_dual+0.01) + log(benef_cc_afib+0.01) + log(benef_cc_alzrdsd) +
log(benef_cc_asthma) + log(benef_cc_cancer) + log(benef_cc_chf) + log(benef_cc_ckd) + log(benef_cc_copd) +
log(benef_cc_depr) + log(benef_cc_diab) + log(benef_cc_hyperl) + log(benef_cc_hypert) + log(benef_cc_ihd) +
log(benef_cc_ost) + log(benef_cc_raoa) + log(benef_cc_schiot+0.01) + log(benef_cc_strk) +
benef_average_risk_score + log(ambulatory_services+0.01) + log(anesthesia+0.01) + log(category_iii+0.01) +
log(drugs_administered_other_than_oral_method+0.01) + log(chemotherapy_drugs+0.01) +
log(vaccines_or_toxoids+0.01) + log(psychiatry_services_and_procedures+0.01) +
log(dialysis_services_and_procedures+0.01) + log(gastroenterology_procedures+0.01) +
log(ophthalmology_services_and_procedures+0.01) +
log(special_otorhinolaryngologic_services_and_procedures+0.01) + log(cardiovascular_procedures+0.01) +
log(non_invasive_vascular_diagnostic_studies+0.01) + log(pulmonary_procedures+0.01) +
log(allergy_and_clinical_immunology_procedures+0.01) + log(endocrinology_services+0.01) +
log(neurology_and_neuromuscular_procedures+0.01) +
log(highly_complex_drug_or_biologic_agent_administration+0.01) + log(therapeutic_procedures+0.01) +
log(physical_medicine_and_rehabilitation_evaluations+0.01) + log(moderate_conscious_sedation+0.01) +
log(other_medicine_services_and_procedures+0.01) + log(office_or_other_outpatient_services+0.01) +
log(hospital_observation_services+0.01) + log(hospital_inpatient_services+0.01) +
log(emergency_department_services+0.01) + log(critical_care_services+0.01) +
log(nursing_facility_services+0.01) + log(home_services+0.01) + log(prolonged_services+0.01) +
log(other_care_evaluation_and_management_services+0.01) + log(organ_or_disease_oriented_panels+0.01) +
log(drug_assay+0.01) + log(urinalysis_procedures+0.01) + log(chemistry_procedures+0.01) +
log(hematology_and_coagulation_procedures+0.01) + log(immunology_procedures+0.01) +
log(microbiology_procedures+0.01) + log(cytopathology_procedures+0.01) +
log(surgical_pathology_procedures+0.01) + log(diagnostic_radiology_procedures+0.01) +
log(diagnostic_ultrasound_procedures+0.01) + log(radiologic_guidance+0.01) + log(breast_and_mammography+0.01) +
log(bone_or_joint_studies+0.01) + log(radiation_oncology_treatment+0.01) +
log(nuclear_medicine_procedures+0.01) + log(fine_needle_aspiration_biopsy_procedures+0.01) +
log(surgical_procedures_on_the_auditory_system+0.01) +
log(surgical_procedures_on_the_cardiovascular_system+0.01) +
log(surgical_procedures_on_the_digestive_system+0.01) +
log(surgical_procedures_on_the_eye_and_ocular_adnexa+0.01) + log(surgical_procedures_on_genital_system+0.01) +
log(surgical_procedures_on_the_hemic_and_lymphatic_systems+0.01) +
log(surgical_procedures_on_the_integumentary_system+0.01) +
log(surgical_procedures_on_the_musculoskeletal_system+0.01) +
log(surgical_procedures_on_the_nervous_system+0.01) + log(surgical_procedures_on_the_respiratory_system+0.01) +
log(surgical_procedures_on_the_urinary_system+0.01) + log(other_surgical_procedures+0.01) +
log(screening_examinations_and_disease_management_training+0.01) +
log(miscellaneous_diagnostic_and_therapeutic_services+0.01) +
log(initial_services_for_medicare_enrollment+0.01) +
```

```
log(gross_and_microscopic_examinations_prostate_biopsy+0.01) +
log(counseling_screening_and_prevention_services+0.01) + log(miscellaneous_services+0.01), REML=FALSE,
weights=wt, data=df_train)
m3_pred <- predict(m3_train, df_test)
sqrt(mean((df_test$overcharge - exp(m3_pred)) ^ 2))
```

```
[1] 766854.2
```

Summary output

```
stargazer(m1, m2, m3, type="text", single.row=TRUE)
```

```
=====================================================================================================================================
                                                                                          Dependent variable:
                                                                     ----------------------------------------------------------------
                                                                                          log(overcharge)
                                                                           linear                  OLS                   linear
                                                                        mixed-effects                                 mixed-effects
                                                                             (1)                   (2)                     (3)
-------------------------------------------------------------------------------------------------------------------------------------
year2013                                                             0.076*** (0.008)        0.077*** (0.008)        0.077*** (0.008)
year2014                                                             0.080*** (0.012)        0.103*** (0.011)        0.103*** (0.011)
year2015                                                             0.198*** (0.013)        0.216*** (0.011)        0.217*** (0.011)
year2016                                                             0.324*** (0.015)        0.321*** (0.013)        0.323*** (0.013)
year2017                                                             0.134*** (0.014)        0.212*** (0.013)        0.212*** (0.013)
nppes_provider_stateAE                                                 0.567 (0.746)           0.114 (0.663)           0.122 (0.661)
nppes_provider_stateAK                                               1.059*** (0.284)         0.591** (0.253)         0.595** (0.253)
nppes_provider_stateAL                                                 0.187 (0.284)          -0.169 (0.253)          -0.166 (0.253)
nppes_provider_stateAP                                                 0.582 (0.419)           0.191 (0.372)           0.196 (0.370)
nppes_provider_stateAR                                                 0.281 (0.284)          -0.059 (0.252)          -0.059 (0.252)
nppes_provider_stateAZ                                                 0.278 (0.282)          -0.222 (0.251)          -0.214 (0.251)
nppes_provider_stateCA                                                0.477* (0.282)           0.059 (0.251)           0.062 (0.251)
nppes_provider_stateCO                                                 0.227 (0.283)          -0.123 (0.252)          -0.121 (0.252)
nppes_provider_stateCT                                                0.502* (0.283)           0.112 (0.251)           0.113 (0.251)
nppes_provider_stateDC                                                 0.407 (0.284)           0.005 (0.252)           0.009 (0.252)
nppes_provider_stateDE                                                 0.415 (0.284)           0.045 (0.252)           0.047 (0.252)
nppes_provider_stateFL                                                 0.424 (0.283)          -0.011 (0.251)          -0.012 (0.251)
nppes_provider_stateGA                                                 0.431 (0.283)           0.083 (0.251)           0.084 (0.251)
nppes_provider_stateGU                                                 0.698 (0.564)           0.216 (0.501)           0.231 (0.496)
nppes_provider_stateHI                                                -0.014 (0.287)          -0.348 (0.255)          -0.352 (0.256)
nppes_provider_stateIA                                                 0.448 (0.285)           0.030 (0.254)           0.034 (0.254)
nppes_provider_stateID                                                -0.016 (0.288)          -0.409 (0.256)          -0.415 (0.256)
nppes_provider_stateIL                                                0.494* (0.282)           0.160 (0.251)           0.164 (0.251)
nppes_provider_stateIN                                                 0.299 (0.284)           0.025 (0.252)           0.029 (0.252)
nppes_provider_stateKS                                                 0.310 (0.283)           0.020 (0.252)           0.026 (0.252)
nppes_provider_stateKY                                                 0.255 (0.284)          -0.024 (0.253)          -0.019 (0.253)
nppes_provider_stateLA                                                 0.248 (0.284)          -0.034 (0.252)          -0.036 (0.252)
nppes_provider_stateMA                                                 0.428 (0.283)           0.079 (0.251)           0.081 (0.251)
nppes_provider_stateMD                                                 0.427 (0.283)           0.002 (0.251)           0.005 (0.251)
nppes_provider_stateME                                                 0.391 (0.291)           0.149 (0.258)           0.166 (0.258)
nppes_provider_stateMI                                                 0.143 (0.283)          -0.154 (0.251)          -0.153 (0.251)
nppes_provider_stateMN                                                 0.119 (0.283)          -0.375 (0.252)          -0.377 (0.252)
nppes_provider_stateMO                                                 0.276 (0.283)           0.047 (0.252)           0.053 (0.252)
nppes_provider_stateMP                                                1.232* (0.747)           0.791 (0.663)           0.786 (0.677)
nppes_provider_stateMS                                                 0.417 (0.285)           0.010 (0.253)           0.012 (0.254)
nppes_provider_stateMT                                                 0.154 (0.296)          -0.097 (0.263)          -0.097 (0.263)
nppes_provider_stateNC                                                0.524* (0.283)           0.126 (0.251)           0.131 (0.251)
nppes_provider_stateND                                                 0.185 (0.291)           0.018 (0.259)           0.023 (0.259)
nppes_provider_stateNE                                                 0.321 (0.285)           0.028 (0.253)           0.036 (0.253)
nppes_provider_stateNH                                                0.507* (0.290)           0.203 (0.258)           0.208 (0.258)
nppes_provider_stateNJ                                                 0.415 (0.282)           0.025 (0.251)           0.029 (0.251)
nppes_provider_stateNM                                                 0.307 (0.284)          -0.132 (0.252)          -0.123 (0.252)
nppes_provider_stateNV                                                0.486* (0.283)           0.068 (0.251)           0.069 (0.251)
nppes_provider_stateNY                                                 0.462 (0.282)           0.064 (0.251)           0.065 (0.251)
nppes_provider_stateOH                                                 0.298 (0.283)           0.070 (0.251)           0.074 (0.251)
nppes_provider_stateOK                                                 0.147 (0.283)          -0.122 (0.251)          -0.125 (0.251)
nppes_provider_stateOR                                                 0.246 (0.283)          -0.138 (0.252)          -0.131 (0.252)
nppes_provider_statePA                                                 0.290 (0.283)          -0.015 (0.251)          -0.009 (0.251)
nppes_provider_statePR                                                -0.186 (0.325)         -0.498* (0.289)         -0.483* (0.292)
nppes_provider_stateRI                                                 0.323 (0.285)          -0.088 (0.253)          -0.086 (0.253)
nppes_provider_stateSC                                                0.475* (0.284)           0.093 (0.252)           0.095 (0.252)
nppes_provider_stateSD                                                -0.106 (0.291)          -0.409 (0.259)          -0.399 (0.259)
nppes_provider_stateTN                                                0.514* (0.284)           0.166 (0.252)           0.171 (0.252)
nppes_provider_stateTX                                                0.467* (0.283)           0.110 (0.251)           0.113 (0.251)
nppes_provider_stateUT                                                 0.083 (0.285)          -0.177 (0.253)          -0.174 (0.254)
nppes_provider_stateVA                                                 0.444 (0.283)           0.053 (0.251)           0.058 (0.251)
nppes_provider_stateVI                                                0.607* (0.352)           0.025 (0.313)           0.031 (0.312)
nppes_provider_stateVT                                                 0.303 (0.302)           0.083 (0.268)           0.067 (0.269)
nppes_provider_stateWA                                                 0.427 (0.282)          -0.047 (0.251)          -0.042 (0.251)
nppes_provider_stateWI                                               0.942*** (0.284)         0.627** (0.252)         0.632** (0.252)
nppes_provider_stateWV                                                 0.243 (0.289)           0.046 (0.256)           0.057 (0.257)
nppes_provider_stateWY                                                 0.027 (0.297)         -0.693*** (0.264)       -0.693*** (0.264)
nppes_provider_stateXX                                                 0.860 (0.564)           0.677 (0.502)           0.684 (0.501)
provider_typeanesthesiologist assistants                                                       0.041 (0.301)
provider_typeanesthesiology                                                                   0.281** (0.129)
provider_typeaudiologist                                                                     -0.857*** (0.147)
provider_typecardiac electrophysiology                                                         0.123 (0.123)
provider_typecardiac surgery                                                                  0.380*** (0.147)
provider_typecardiology                                                                       -0.088 (0.121)
provider_typecertified clinical nurse specialist                                             -0.388** (0.178)
provider_typecertified registered nurse anesthetist (crna)                                     0.057 (0.142)
provider_typechiropractic                                                                     -0.429 (0.452)
provider_typeclinical psychologist                                                             0.088 (0.154)
provider_typecolorectal surgery (proctology)                                                  0.459*** (0.169)
provider_typecritical care (intensivists)                                                      0.199 (0.126)
provider_typecrna                                                                             -0.057 (0.131)
provider_typedermatology                                                                     -0.453*** (0.124)
```

provider_typediagnostic radiology		-0.600*** (0.121)	
provider_typeemergency medicine		-0.045 (0.122)	
provider_typeendocrinology		-0.216* (0.123)	
provider_typefamily practice		-0.253** (0.121)	
provider_typegastroenterology		0.294** (0.124)	
provider_typegeneral practice		-0.198 (0.127)	
provider_typegeneral surgery		0.158 (0.124)	
provider_typegeriatric medicine		0.084 (0.133)	
provider_typegeriatric psychiatry		-0.269 (0.240)	
provider_typegynecological/oncology		0.834*** (0.264)	
provider_typehand surgery		0.118 (0.144)	
provider_typehematology		0.448*** (0.145)	
provider_typehematology/oncology		0.571*** (0.124)	
provider_typehospice and palliative care		-0.176 (0.157)	
provider_typehospitalist		0.164 (0.153)	
provider_typeindependent diagnostic testing facility		-0.208 (0.375)	
provider_typeinfectious disease		0.040 (0.123)	
provider_typeinternal medicine		-0.064 (0.120)	
provider_typeinterventional radiology		-0.455*** (0.123)	
provider_typelicensed clinical social worker		0.410 (0.287)	
provider_typemaxillofacial surgery		-1.534*** (0.262)	
provider_typemedical oncology		0.535*** (0.129)	
provider_typenephrology		-0.019 (0.125)	
provider_typeneurology		-0.276** (0.123)	
provider_typeneuropsychiatry		-0.448* (0.265)	
provider_typeneurosurgery		1.090*** (0.134)	
provider_typenuclear medicine		-0.109 (0.125)	
provider_typenurse practitioner		-0.160 (0.122)	
provider_typeobstetrics/gynecology		-0.314** (0.150)	
provider_typeoccupational therapist		1.105** (0.452)	
provider_typeophthalmology		-1.204*** (0.130)	
provider_typeoptometry		-1.542*** (0.132)	
provider_typeoral surgery (dentist only)		-0.587** (0.249)	
provider_typeorthopedic surgery		0.661*** (0.123)	
provider_typeosteopathic manipulative medicine		-0.040 (0.204)	
provider_typeotolaryngology		-0.174 (0.124)	
provider_typepain management		0.454*** (0.133)	
provider_typepathology		-0.457*** (0.124)	
provider_typepediatric medicine		-0.424*** (0.150)	
provider_typeperipheral vascular disease		0.129 (0.174)	
provider_typephysical medicine and rehabilitation		-0.020 (0.124)	
provider_typephysical therapist		0.513*** (0.176)	
provider_typephysician assistant		-0.149 (0.124)	
provider_typeplastic and reconstructive surgery		0.430** (0.178)	
provider_typepodiatry		-0.765*** (0.124)	
provider_typepreventive medicine		-0.491** (0.228)	
provider_typepsychiatry		-0.071 (0.133)	
provider_typepulmonary disease		-0.0004 (0.121)	
provider_typeradiation oncology		-0.292** (0.134)	
provider_typeregistered dietician/nutrition professional		-0.983*** (0.265)	
provider_typerheumatology		0.405*** (0.127)	
provider_typesleep medicine		0.107 (0.196)	
provider_typesports medicine		0.365** (0.171)	
provider_typesurgical oncology		1.482*** (0.248)	
provider_typethoracic surgery		0.300** (0.152)	
provider_typeurology		0.071 (0.128)	
provider_typevascular surgery		0.268** (0.123)	
log(total_unique_benes)	0.628*** (0.005)	0.583*** (0.004)	0.578*** (0.004)
log(number_of_hcpcs)	0.394*** (0.004)	0.263*** (0.005)	0.264*** (0.005)
benef_average_age	0.021*** (0.001)	0.011*** (0.001)	0.011*** (0.001)
log(benef_female)	-0.231*** (0.021)	-0.202*** (0.020)	-0.192*** (0.020)
log(benef_race_white + 0.01)	0.184*** (0.009)	0.144*** (0.008)	0.148*** (0.008)
log(benef_race_black + 0.01)	-0.001 (0.004)	0.024*** (0.004)	0.024*** (0.004)
log(benef_race_api + 0.01)	0.057*** (0.005)	0.048*** (0.004)	0.048*** (0.004)
log(benef_race_hispanic + 0.01)	0.039*** (0.004)	0.042*** (0.004)	0.041*** (0.004)
log(benef_race_natind + 0.01)	-0.044*** (0.008)	-0.004 (0.007)	-0.003 (0.007)
log(benef_dual + 0.01)	0.106*** (0.008)	0.098*** (0.007)	0.099*** (0.007)
log(benef_cc_afib + 0.01)	-0.344*** (0.011)	-0.229*** (0.010)	-0.233*** (0.010)
log(benef_cc_alzrdsd)	-0.190*** (0.010)	-0.139*** (0.009)	-0.140*** (0.009)
log(benef_cc_asthma)	-0.246*** (0.010)	-0.161*** (0.009)	-0.164*** (0.009)
log(benef_cc_cancer)	0.095*** (0.010)	0.087*** (0.010)	0.086*** (0.010)
log(benef_cc_chf)	-0.320*** (0.014)	-0.213*** (0.013)	-0.216*** (0.013)
log(benef_cc_ckd)	-0.093*** (0.013)	-0.114*** (0.012)	-0.117*** (0.012)
log(benef_cc_copd)	0.057*** (0.012)	-0.011 (0.011)	-0.011 (0.011)
log(benef_cc_depr)	-0.060*** (0.013)	-0.065*** (0.011)	-0.069*** (0.011)
log(benef_cc_diab)	0.009 (0.014)	-0.048*** (0.013)	-0.051*** (0.013)
log(benef_cc_hyperl)	0.280*** (0.026)	0.290*** (0.023)	0.292*** (0.024)
log(benef_cc_hypert)	0.323*** (0.048)	0.145*** (0.044)	0.165*** (0.044)
log(benef_cc_ihd)	0.297*** (0.021)	0.183*** (0.019)	0.193*** (0.019)
log(benef_cc_ost)	-0.103*** (0.010)	-0.033*** (0.009)	-0.034*** (0.009)
log(benef_cc_raoa)	0.362*** (0.019)	0.163*** (0.017)	0.165*** (0.017)
log(benef_cc_schiot + 0.01)	-0.193*** (0.008)	-0.120*** (0.007)	-0.120*** (0.008)
log(benef_cc_strk)	0.220*** (0.009)	0.130*** (0.009)	0.130*** (0.009)
benef_average_risk_score	0.332*** (0.007)	0.246*** (0.007)	0.251*** (0.007)
log(ambulatory_services + 0.01)		0.075*** (0.001)	0.075*** (0.001)
log(anesthesia + 0.01)		0.075*** (0.006)	0.076*** (0.006)
log(category_iii + 0.01)		0.033*** (0.004)	0.032*** (0.004)
log(drugs_administered_other_than_oral_method + 0.01)		0.025*** (0.001)	0.026*** (0.001)
log(chemotherapy_drugs + 0.01)		0.088*** (0.002)	0.088*** (0.002)
log(vaccines_or_toxoids + 0.01)		-0.040*** (0.002)	-0.040*** (0.002)
log(psychiatry_services_and_procedures + 0.01)		0.046*** (0.007)	0.046*** (0.007)
log(dialysis_services_and_procedures + 0.01)		0.005 (0.004)	0.003 (0.004)
log(gastroenterology_procedures + 0.01)		0.036*** (0.007)	0.037*** (0.007)
log(ophthalmology_services_and_procedures + 0.01)		0.062*** (0.006)	0.058*** (0.006)
log(special_otorhinolaryngologic_services_and_procedures + 0.01)		0.014*** (0.004)	0.013*** (0.004)
log(cardiovascular_procedures + 0.01)		-0.007*** (0.001)	-0.007*** (0.001)
log(non_invasive_vascular_diagnostic_studies + 0.01)		0.019*** (0.001)	0.019*** (0.001)
log(pulmonary_procedures + 0.01)		0.002 (0.003)	0.002 (0.003)
log(allergy_and_clinical_immunology_procedures + 0.01)		0.031*** (0.004)	0.031*** (0.003)
log(endocrinology_services + 0.01)		0.010 (0.008)	0.010 (0.008)
log(neurology_and_neuromuscular_procedures + 0.01)		0.037*** (0.002)	0.036*** (0.002)

log(highly_complex_drug_or_biologic_agent_administration + 0.01)		-0.004* (0.002)	-0.003* (0.002)
log(therapeutic_procedures + 0.01)		0.004 (0.003)	0.004 (0.003)
log(physical_medicine_and_rehabilitation_evaluations + 0.01)		0.012*** (0.004)	0.012*** (0.004)
log(moderate_conscious_sedation + 0.01)		0.017*** (0.003)	0.017*** (0.003)
log(other_medicine_services_and_procedures + 0.01)		0.007 (0.005)	0.005 (0.005)
log(office_or_other_outpatient_services + 0.01)		0.002* (0.001)	0.003** (0.001)
log(hospital_observation_services + 0.01)		0.007*** (0.002)	0.007*** (0.002)
log(hospital_inpatient_services + 0.01)		0.019*** (0.002)	0.019*** (0.002)
log(emergency_department_services + 0.01)		0.079*** (0.003)	0.078*** (0.003)
log(critical_care_services + 0.01)		0.043*** (0.002)	0.044*** (0.002)
log(nursing_facility_services + 0.01)		0.023*** (0.001)	0.023*** (0.001)
log(home_services + 0.01)		0.009*** (0.003)	0.009*** (0.003)
log(prolonged_services + 0.01)		0.017*** (0.003)	0.017*** (0.003)
log(other_care_evaluation_and_management_services + 0.01)		0.018*** (0.003)	0.017*** (0.003)
log(organ_or_disease_oriented_panels + 0.01)		0.003 (0.002)	0.003 (0.002)
log(drug_assay + 0.01)		0.060*** (0.008)	0.062*** (0.008)
log(urinalysis_procedures + 0.01)		-0.037*** (0.002)	-0.037*** (0.002)
log(chemistry_procedures + 0.01)		0.001 (0.002)	0.0001 (0.002)
log(hematology_and_coagulation_procedures + 0.01)		0.006*** (0.001)	0.006*** (0.001)
log(immunology_procedures + 0.01)		0.002 (0.003)	0.003 (0.003)
log(microbiology_procedures + 0.01)		-0.034*** (0.003)	-0.034*** (0.003)
log(cytopathology_procedures + 0.01)		-0.003 (0.003)	-0.004 (0.003)
log(surgical_pathology_procedures + 0.01)		0.070*** (0.003)	0.068*** (0.003)
log(diagnostic_radiology_procedures + 0.01)		0.015*** (0.002)	0.016*** (0.001)
log(diagnostic_ultrasound_procedures + 0.01)		0.004*** (0.001)	0.004*** (0.001)
log(radiologic_guidance + 0.01)		0.027*** (0.003)	0.027*** (0.003)
log(breast_and_mammography + 0.01)		-0.011*** (0.002)	-0.011*** (0.002)
log(bone_or_joint_studies + 0.01)		-0.009*** (0.002)	-0.010*** (0.002)
log(radiation_oncology_treatment + 0.01)		0.155*** (0.008)	0.150*** (0.008)
log(nuclear_medicine_procedures + 0.01)		-0.0004 (0.001)	-0.001 (0.001)
log(fine_needle_aspiration_biopsy_procedures + 0.01)		0.019*** (0.004)	0.018*** (0.004)
log(surgical_procedures_on_the_auditory_system + 0.01)		-0.005 (0.004)	-0.005 (0.004)
log(surgical_procedures_on_the_cardiovascular_system + 0.01)		0.030*** (0.001)	0.031*** (0.001)
log(surgical_procedures_on_the_digestive_system + 0.01)		0.026*** (0.003)	0.025*** (0.003)
log(surgical_procedures_on_the_eye_and_ocular_adnexa + 0.01)		0.171*** (0.004)	0.175*** (0.004)
log(surgical_procedures_on_genital_system + 0.01)		-0.028*** (0.008)	-0.026*** (0.008)
log(surgical_procedures_on_the_hemic_and_lymphatic_systems + 0.01)		0.018*** (0.005)	0.018*** (0.005)
log(surgical_procedures_on_the_integumentary_system + 0.01)		0.006*** (0.002)	0.005** (0.002)
log(surgical_procedures_on_the_musculoskeletal_system + 0.01)		0.005* (0.003)	0.005** (0.003)
log(surgical_procedures_on_the_nervous_system + 0.01)		0.043*** (0.003)	0.042*** (0.003)
log(surgical_procedures_on_the_respiratory_system + 0.01)		0.003 (0.003)	0.003 (0.003)
log(surgical_procedures_on_the_urinary_system + 0.01)		0.005 (0.005)	0.005 (0.005)
log(other_surgical_procedures + 0.01)		0.042*** (0.014)	0.042*** (0.014)
log(screening_examinations_and_disease_management_training + 0.01)		-0.012*** (0.003)	-0.012*** (0.003)
log(miscellaneous_diagnostic_and_therapeutic_services + 0.01)		0.005*** (0.002)	0.005*** (0.002)
log(initial_services_for_medicare_enrollment + 0.01)		0.003 (0.006)	0.004 (0.006)
log(gross_and_microscopic_examinations_prostate_biopsy + 0.01)		0.033*** (0.005)	0.035*** (0.005)
log(counseling_screening_and_prevention_services + 0.01)		-0.002 (0.003)	-0.002 (0.003)
log(miscellaneous_services + 0.01)		0.033*** (0.001)	0.033*** (0.001)
Constant	3.525*** (0.316)	12.027*** (0.336)	11.942*** (0.320)
Observations	103,548	103,548	103,548
R2		0.599	
Adjusted R2		0.598	
Log Likelihood	-108,753.600		-96,400.310
Akaike Inf. Crit.	217,693.200		193,118.600
Bayesian Inf. Crit.	218,581.200		194,636.700
Residual Std. Error		0.613 (df = 103320)	
F Statistic		679.124*** (df = 227; 103320)	
Note:			*p<0.1; **p<0.05; ***p<0.01