SOA Research Institute

A Two-Part Model of the Individual Costs of Chronic Kidney Disease

Health Care Cost Trends

December 2021

A Two-Part Model of the Individual Costs of Chronic Kidney Disease

AUTHORS

Brian Hartman, ASA PhD
Brigham Young University
Milliman, Inc.

Courtney Larson
Milliman, Inc.

Christopher Kunkel, FSA PhD
Milliman, Inc.

Cason Wight
Brigham Young University

Richard L. Warr, PhD
Brigham Young University

SPONSOR

Health Care Cost Trends Research Committee

Caveat and Disclaimer

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries or its members. The Society of Actuaries makes no representation or warranty to the accuracy of the information.

CONTENTS

A Two-part Model of the Individual Costs of Chronic Kidney Disease

Executive Summary

Chronic Kidney Disease (CKD) affects many lives and has a large impact on health systems around the world. To better understand and predict costs for insurance plan members with CKD in the United States, we built a new model of their individual costs. Our model is the first to explicitly model both the CKD stage transition process and the distribution of costs given those stages. Additionally, it incorporates numerous covariates and comorbidities. We applied the models to two large and rich datasets, one commercial insurance and the other Medicare fee-for-service, totaling about 40 million beneficiary months. We found that the XGBoost models best predict both stage transitions and costs. If XGBoost models are unavailable, a multivariate logistic regression model with regularization to predict stage and a logit- gamma model of the costs given the stage best predicted the member's healthcare costs in the next month.

1 Introduction

CKD is a global health issue, but in this article, we focus on the United States. In the United States, the Centers for Disease Control and Prevention (2019) notes that 37 million people (about 15% of the adult population) are estimated to have CKD. CKD is most common in older populations. Roughly one in 10 people globally are affected by CKD, which can be developed at any age.

In the United States, many patients with end-stage renal disease (ESRD) and chronic kidney disease (CKD) are covered by Medicare. Collins et al. (2010) reviewed the 2009 annual data report from the US Renal Data System (USRDS) and found that in 2007, more than 100,000 patients started ESRD therapy, spending over $35billion (roughly $24 billion covered by Medicare). More than 17,500 transplants were performed, and the total CKD-treatment-receiving population reached over 500,000 patients. Roughly 6% of Medicare expenses were attributed to ESRD treatment alone.

The stage of a CKD patient is determined by the patient's glomerular filtration rate (GFR), which measures how quickly blood is filtered through the kidneys. The GFR determines six stages of CKD. Stage 1 is any GFR higher than 90 ml/min. A GFR of 60 to 89 ml/min corresponds to stage 2, 45 to 59 ml/min to stage 3a, 30 to 44 ml/min to stage 3b, 15 to 29 ml/min to stage 4, and less than 15 ml/min to stage 5 (Johnson et al. 2004). Kidney failure (or end stage renal disease, ESRD) are patients in stage 5 who require dialysis or a transplant.

These GFR stages can be combined with three categories of albuminuria to assess a patient's risk. Healthy kidneys prevent albumin (a protein found in the blood) from passing into urine. Albuminuria is the level of albumin in the urine; more albumin is a sign of unhealthy kidneys. Stage A1 corresponds to less than 30 mg of albumin per gram of urine, A2 to 30 to 300 mg/g, and A3 to any higher level (Levey et al. 2015). The combination of a higher GFR stage and higher albuminuria indicate a higher risk level to the kidneys. For the interested reader, Levey and Coresh (2012) provide a general introduction to CKD and related topics.

In 2017, spending on Medicare beneficiaries with CKD was over $84 billion, and members with ESRD spent an additional $36 billion (Centers for Disease Control and Prevention 2021). On average, the higher the patient's CKD stage, the higher the costs, across all medical services. Our analysis of administrative claims datasets shows that for both the Medicare and commercial data, the average costs per member, per month (PMPM) throughout the study period (2016-2018) (Tables 1 and 2) increase exponentially as stage increases, highlighting the need for accurate models. These results are similar to those found in other studies (Golestaneh et al. 2017; Saran et al. 2021; Gandjour et al. 2020).

Table 1

AVERAGE PMPM MEDICARE CLAIMS BY CKD STAGE, 2016–2018

CKD Stage	Inpatient Claims PMPM ($)	Outpatient Claims PMPM ($)	Professional Claims PMPM ($)	Total Claims PMPM ($)
1	$1,600	$400	$998	$2,998
2	$2,084	$426	$1,064	$3,574
3	$2,707	$490	$1,253	$4,450
4	$3,427	$514	$1,505	$5,446
5	$3,385	$561	$1,520	$5,466
6 (ESRD)	$3,739	$3,459	$1,737	$8,935

Table 2

AVERAGE PMPM COMMERCIAL CLAIMS BY CKD STAGE, 2016–2018

CKD Stage	Inpatient Claims PMPM ($)	Outpatient Claims PMPM ($)	Professional Claims PMPM ($)	Total Claims PMPM ($)
1	$980	$756	$807	$2,543
2	$1,539	$870	$869	$3,277
3	$2,561	$1,117	$1,129	$4,807
4	$3,316	$1,284	$1,377	$5,977
5	$2,777	$1,453	$1,322	$5,553
6 (ESRD)	$5,411	$12,043	$2,110	$19,564

Note: Excludes pharmacy drug costs derived from diagnoses present in administrative claims data.

In early stages, CKD progression can be slowed through improvements in diet, exercise, and certain medical interventions. As the disease progresses and the kidneys can no longer sufficiently filter the blood, patients may require dialysis (assisted blood filtration). In cases of extreme risk or occurrence of ESRD, a kidney transplant may be necessary. The costs associated with these treatments are often large. The first year of a kidney transplant costs, on average, $442,500 in the US (Bentley and Ortner 2020).

To help account for the large impact of CKD on the US health system, we developed a new model which better predicts the healthcare cost of patients with CKD by explicitly modeling both the stage and the costs within that stage. Section 2 describes the current state of the art in CKD modeling around the world and how our models improve stage and cost prediction. Section 3 describes the commercial and Medicare datasets used in our analyses. Section 4 details our models and section 5 provides the results of our analysis. Finally, section 6 presents our conclusions.

2 Literature Review

Kerr et al. (2012) showed that CKD-related costs to the English National Health Service are very large, comprising mostly renal replacement therapy but also excess comorbidities. The literature points to several drivers of these costs. Tangri et al. (2011) found that common lab results such as calcium, albumin, and phosphate levels are good indicators of kidney failure. Along with other demographic information, Damien et al. (2016) and Manns et al. (2019) found comorbidities to be significant drivers for CKD-related costs. Damien et al. (2016) also found treatment type (dialysis, non-dialysis, or transplant) and insurance type (public or private) to have a significant effect on costs. Laliberté et al. (2009) examined patients with hypertension and/or diabetes in a managed-care setting and found that costs spiked after CKD diagnosis.

Chang et al. (2016) performed an observational study that indicated that peritoneal dialysis dominates hemodialysis in cost effectiveness. Khan and Amedia Jr (2008) reviewed the related literature and found that early nephrologist referrals, the use of erythropoiesis-stimulating proteins (ESP), and management of hypertension each have a large impact on outcomes and costs.

Intervention techniques were found to reduce costs. Hopkins et al. (2011) show that multidisciplinary care can reduce costs, mainly by decreasing the length of hospital stays. Similarly, P. M. Chen et al. (2015) found that a multidisciplinary approach improves health outcomes and reduces average costs. Wei et al. (2010) show that multidisciplinary care is associated with higher costs before dialysis, but much lower costs after dialysis. Erickson et al. (2013) show that preemptive prescription of statins is cost-effective for high-risk patients. Manns et al. (2010) found that the cost of a broad one-off screening for CKD was not cost-effective in Canada. In contrast, Kondo et al. (2012) show that in Japan a widespread test was cost-effective.

Several models have been implemented to examine the costs and progression of CKD. Tangri et al. (2011) used a binary prediction model for kidney failure. Generalized linear models have assessed various covariates as well. Honeycutt et al. (2013) used a two-part model to show the specific effects of CKD on costs and its important covariates. They modeled total costs for patients with CKD and those without CKD to compare the effect of CKD on costs. Damien et al. (2016) also used a generalized linear model to understand the impact of cost drivers.

Markov-style models are useful in this scenario due to the common use of the CKD stages in clinical practice. Hoerger et al. (2010) used microsimulation to look at CKD progression, focusing on costs. In Orlando et al. (2011), a simple Markov model was constructed to simulate progression and provide base cases for costs. This model assumes fixed transition probabilities and fixed effects of intervention on costs. Fricks et al. (2016) used Markov reward modeling to simulate transitions, but the costs per year for each stage are also fixed. They note that after healthy patients, patients with transplants have the best health outcomes; however, the donor pool is insufficient for those with ESRD.

Our model simultaneously improves on the current state of the art in both CKD stage modeling and cost modeling. We use many covariates and a regularization technique to focus on the most important predictors of CKD stage and cost. We also compare several different distributional families, Medicare and commercial populations, as well as explicitly model both the stage transitions and the distribution of costs in each stage.

3 Data

We performed these analyses on two different datasets. Milliman's Consolidated Health Cost Source Database (the commercial dataset) and the Medicare 5% Sample Medicare Standard Analytical Files (the Medicare dataset). We limited both datasets to claims from members with at least one CKD diagnosis from 2016 to 2018. After applying this inclusion criteria, the Medicare dataset contains about 11 million beneficiary months and the commercial dataset contains about 28 million member months. We connected each member's monthly enrollment and claims data within each dataset but could not connect patients across datasets. Note that there are conditions for a patient with CKD to qualify for Medicare before age 65 (Centers for Medicare and Medicaid Services 2021).

For each beneficiary month, we have information on age, sex, state, a measure of rurality (IIR_2010 described in Waldorf & Kim (2018)), inpatient, outpatient, and professional costs, and a series of indicators for CKD stage and the presence of about 50 other comorbidities, which were drawn from claims data. The costs are all-cause healthcare spending, not just the CKD-related costs. Using total costs allows us to account for comorbidities directly and enables practitioners to more easily apply our models to their data. The summary statistics of all the indicators used in our models are provided in Table 3. The ICD-10 codes that trigger indicators of comorbidities are available in Appendix B. With the exception of gross hematuria and the two urinary tract infection (UTI) flags, comorbidities are considered chronic; and therefore, members have the indicator in every subsequent month following its first occurrence. All data has been processed through the Milliman Health Cost Guidelines Grouper, which uses continuous stay logic to keep together inpatient claims where intermediate billing records may be generated (for example an inpatient stay that extends across the end of one month into the beginning of the next). For other services, they are recorded in the month indicated in each record's starting date (often the first of the month, or once per week). The rurality indicator is based on member location. Our data cannot differentiate between stages 3a and 3b, so we combined them into stage 3. For ease in defining our model, we will call ESRD stage 6. We also define stage 0 as members who do not yet have a CKD diagnosis but will in the future. In summary, our data encompasses seven stages: 0 (no CKD), 1, 2, 3 (both 3a and 3b), 4, 5 and 6 (ESRD).

Table 3

EXPLORATORY DATA ANALYSIS

	Commercial	Medicare
Total beneficiary months	27,645,493	10,870,922
Continuous Covariates: Mean (SD)		
IRR 2010	0.33 (0.11)	0.35 (0.13)
Age	59.92 (14.15)	75.92 (11.37)
Inpatient claims	867.48 (21,913.51)	1,079.82 (6,471.09)
Outpatient claims	904.21 (7,046.38)	475.27 (1,673.71)
Professional claims	559.30 (2,281.30)	696.41 (1,630.95)
Categorical Covariates: Count (%)		
CKD stage 0 (no CKD yet)	10,060,507 (36.4)	3,501,047 (32.2)
CKD stage 1	1,197,109 (4.3)	173,249 (1.6)
CKD stage 2	3,280,333 (11.9)	765,838 (7)
CKD stage 3	9,837,864 (35.6)	4,489,667 (41.3)
CKD stage 4	1,451,143 (5.2)	883,228 (8.1)
CKD stage 5	220,030 (0.8)	102,743 (0.9)
CKD stage 6 (ESRD)	1,598,507 (5.8)	955,150 (8.8)
Acute myocardial infarction	816,151 (03.0)	613,334 (05.6)
Aspiration and specified bacterial pneumonias and other severe lung infections	377,514 (01.4)	346,112 (03.2)
Asthma	2,956,670 (10.7)	1,913,269 (17.6)
Atrial and ventricular septal defects, patent ductus arteriosus and other congenital heart/circulatory disorders	208,634 (00.8)	131,715 (01.2)
Back pain	4,545,147 (16.4)	2,645,135 (24.3)

Breast (age > 50) and prostate cancer, benign/uncertain brain tumors, and other cancers and tumors	1,644,239 (05.9)	1,140,431 (10.5)
Cardiomegaly	2,158,610 (07.8)	1,686,058 (15.5)
Chronic obstructive pulmonary disease, including bronchiectasis	2,621,373 (09.5)	2,623,010 (24.1)
Colorectal, breast (age < 50), kidney and other cancers	1,194,058 (04.3)	708,144 (06.5)
Congenital renal cyst	147,167 (00.5)	78,210 (00.7)
Congestive heart failure	3,934,178 (14.2)	3,313,255 (30.5)
Diabetes with chronic complications	8,491,807 (30.7)	4,168,533 (38.3)
Diabetes without complication	9,644,788 (34.9)	4,908,413 (45.2)
Diet	881,294 (03.2)	329,187 (03.0)
Gross hematuria	60,418 (00.2)	37,894 (00.3)
Heart infection/inflammation, except rheumatic	415,396 (01.5)	256,521 (02.4)
Hemodialysis	1,002,348 (03.6)	716,230 (06.6)
Hepatic cysts	35,746 (00.1)	9,963 (00.1)
High blood pressure, not diagnosed as hypertension	913,072 (03.3)	392,343 (03.6)
Hyperlipidemia	15,175,366 (54.9)	7,448,121 (68.5)
Hypertension	19,281,458 (69.7)	9,036,831 (83.1)
Hypoplastic left heart syndrome and other severe congenital heart disorders	7,072 (00.0)	1,194 (00.0)
Ischemic or unspecified stroke	900,171 (03.3)	881,535 (08.1)
Kidney transplant	789,791 (02.9)	230,309 (02.1)
Lung, brain, and other severe cancers, including pediatric acute lymphoid leukemia	592,515 (02.1)	364,227 (03.4)
Major congenital heart/circulatory disorders	195,853 (00.7)	120,427 (01.1)
Metastatic cancer	511,958 (01.9)	256,884 (02.4)
Nicotine dependence	5,083,096 (18.4)	3,316,243 (30.5)
Non-Hodgkin's lymphomas and other cancers and tumors	403,712 (01.5)	239,980 (02.2)
Proteinuria	3,259,233 (11.8)	967,955 (08.9)
Renal anemia	2,441,513 (08.8)	1,682,580 (15.5)
Sex = M (%)	15,684,760 (56.7)	5,245,898 (48.3)
Specified heart arrhythmias	3,301,685 (11.9)	2,930,344 (27.0)
Thyroid cancer, melanoma, neurofibromatosis and other cancers and tumors	430,456 (01.6)	239,045 (02.2)
Unstable angina and other acute ischemic heart disease	787,089 (02.8)	619,325 (05.7)
Urinary tract infection with hematuria	36,481 (00.1)	25,071 (00.2)
Urinary tract infection without hematuria	611,394 (02.2)	540,293 (05.0)
Weight: BMI 19.9 or less	202,621 (00.7)	194,772 (01.8)
Weight: BMI 20–24.9	811,211 (02.9)	577,837 (05.3)
Weight: BMI 25–29.9	2,001,557 (07.2)	1,029,930 (09.5)
Weight: BMI 30–34.9	2,402,540 (08.7)	1,092,267 (10.0)
Weight: BMI 35–39.9	1,737,144 (06.3)	730,401 (06.7)
Weight: BMI 40 or greater	1,888,576 (06.8)	663,713 (06.1)
Weight, newborns: <500 g	186 (00.0)	0 (00.0)
Weight, newborns: 500–749 g	1,293 (00.0)	34 (00.0)
Weight, newborns: 750–999 g	1,471 (00.0)	2 (00.0)
Weight, newborns: 1,000–1,499 g	1,750 (00.0)	46 (00.0)
Weight, newborns: 1,500–1,999 g	2,511 (00.0)	35 (00.0)
Weight, newborns: 2,000–2,499 g	2,982 (00.0)	229 (00.0)
Weight, newborns: Other low-birthweight issues	9,298 (00.0)	268 (00.0)
Weight, pediatric: <5th percentile	6,081 (00.0)	430 (00.0)
Weight, pediatric: 5th–95th percentile	43,461 (00.2)	1,212 (00.0)
Weight, pediatric: >95th percentile	17,535 (00.1)	535 (00.0)

4 Model

Our model has two parts, both explicitly modeled. The first part models the CKD stage and the second part models the monthly cost given the CKD stage.

4.1 STAGE MODEL

Our stage model uses the covariate information from the current month, especially the current CKD stage, to predict the CKD stage for the following month. For months in which the data lacks a diagnosis code or otherwise suggest the patients have moved to lower stages, we assigned the earlier, higher stage. Specifically, we model the odds of a CKD patient remaining in their current state or progressing to any worse state. We assume that patients cannot reverse their CKD progression (National Institute of Diabetes and Digestive and Kidney Disease 2021).

We compare two types of models—multivariate logistic regression and extreme gradient boosting ("XGBoost") (T. Chen and Guestrin 2016; T. Chen et al. 2015)—to predict the transition probabilities. Because we have many potential covariates, we simplified our model by using regularization in the logistic regression through the lasso penalty (Tibshirani 1996). The lasso penalty automatically shrinks some coefficients to zero and effectively constrains the other variables.

We set up our two different training and test sets temporally. The first uses all the 2016 data as the training set and the 2017 data as the test set, imagining we are currently in December 2016 and trying to predict CKD stages for next year. The second uses the 2016-2017 data as the training data and the 2018 data as the test data.

To find appropriate values of the logistic regression hyperparameters (in this case, the lasso penalty λ), we performed five-fold cross validation on the training data. Our hyperparameter estimation technique for the XGBoost model is described in Appendix A.

4.2 COST MODEL

The cost model uses covariate information from the current month and next month's CKD stage to predict the member's full (not just CKD-related) healthcare costs for next month. We chose to use next month's stage rather than use the current month's stage, so we could simply multiply the probabilities from the stage model by the distribution of costs in the cost model to get the overall cost prediction for the member. Also, next month's stage has a more direct impact on next month's costs than the current month's stage. We fit three different cost models (1) a logit-gamma model (2) a Tweedie model and (3) an XGBoost model. The gamma model allows only positive values, so we first ran an additional logistic regression model to predict whether the member had any costs for that specific month and then fit the gamma model on only the members with positive costs. We used a log link to enable better coefficient interpretation than the canonical (inverse) link. The Tweedie and XGBoost models can handle both zero and positive values.

5 Results

For both the stage model and the cost model, we discuss both the model selection results (i.e., which model best predicts) and the covariate inference (which covariates are most influential to costs and stage progression).

5.1 STAGE MODEL

The stage model explicitly models the odds of progression of the member through the CKD stages.

5.1.1 MODEL SELECTION

In addition to the XGBoost model, we compared three different multinomial logistic regression models: one with no regularization, one with a lasso penalty equal to the minimum as selected by the cross validation (min), and a third model with the lasso penalty creating the simplest model within one standard error of the minimum λ (1 s.e.). To compare the models, we calculated the categorical cross-entropy (the negative log of the probability of the correct class) in the test set. The smaller the cross-entropy, the better the model fits. The results of the tests are summarized in Table 4.

Table 4

STAGE MODEL COMPARISON BY CROSS-ENTROPY (MINIMUM VALUES IN BOLD)

Data Set	Stage	Prediction Year	Multi	Multi Min	Multi 1 s.e.	XGBoost
Commercial	1	2017	29,633	29,633	29,533	**29,472**
Commercial	1	2018	26,195	26,200	26,372	**25,860**
Commercial	2	2017	78,986	78,995	79,553	**78,108**
Commercial	2	2018	74,439	74,430	74,166	**72,840**
Commercial	3	2017	137,285	137,201	137,665	**131,353**
Commercial	3	2018	125,789	125,770	125,915	**119,264**
Commercial	4	2017	52,612	52,588	52,687	**51,359**
Commercial	4	2018	46,816	46,794	46,593	**45,160**
Commercial	5	2017	19,067	18,929	17,193	**16,243**
Commercial	5	2018	16,026	15,955	15,218	**14,433**
Medicare	1	2017	8,372	8,053	7,289	**7,277**
Medicare	1	2018	7,435	7,257	6,733	**6,658**
Medicare	2	2017	28,298	**28,229**	28,603	28,379
Medicare	2	2018	25,691	25,658	25,696	**25,590**
Medicare	3	2017	76,988	76,943	76,376	**75,346**
Medicare	3	2018	111,191	110,890	87,736	**66,559**
Medicare	4	2017	25,823	25,423	24,432	**23,914**
Medicare	4	2018	22,533	22,498	22,011	**21,552**
Medicare	5	2017	5,925	5,912	5,943	**5,802**
Medicare	5	2018	5,436	5,433	5,533	**5,237**

The XGBoost model outperforms the logistic model in all subsets except one. XGBoost models are generally more difficult to interpret than the logistic models but predict stage transitions more accurately. If you must use the logistic models, a model with regularization is almost always better (90%) than the model without regularization. For the Medicare data, the 1 s.e. penalty is better for stages 1, 3, and 4 while the minimum penalty is better in stages 2 and 5. The optimal model in the commercial data varies with no regularization being optimal for stage 1 with prediction year 2018 and stage 2 with predictive year 2017. The min penalty is optimal for stage 3 (both predictive years) and for stage 4 with predictive year 2017. The 1 s.e. penalty is optimal for all others. In both situations where the model without regularization is better, the multi min model performance is very close, showing that generally the regularized model is preferred if a logistic model is required.

5.1.2 COVARIATE INFERENCE USING THE MULTINOMIAL MODEL WITH OPTIMAL REGULARIZATION

Figure 1 plots the preliminary (before accounting for any covariates) odds of transitioning to various stages for someone currently in stage 3. In all the plots, the colors are simply to improve visualization. For a member currently in stage 3, Figure 1 shows that the most likely outcome is that the member will stay in stage 3. If the member does move, they will most likely move to stage 4. They are less likely to jump to stage 6, and least likely to jump to stage 5 in the next month. Similar plots for all stages and both datasets are available in Appendix C. The biggest difference between the preliminary odds of the two patient cohorts is that the members in the commercial dataset are much less likely to jump directly to stage 6 from stage 3. A similar pattern is seen across the other stages. In all cases though, the most likely outcome by a significant margin is that the member will stay in the same stage from one month to the next. Figure 2 shows the model coefficients for the Medicare data in stage 3 at the optimal value of λ (in this case 1 s.e.). Plots for all stages are also available in Appendix C. Any covariate not in the plot was removed from the model by the lasso penalization. Patients receiving hemodialysis are very likely to be in stage 6 next month (it is likely they are already there but have not been diagnosed or coded yet). Some other covariates like renal anemia and proteinuria make it likely that they will move to stage 4, but not to stage 6. Others (gross hematuria and UTI with hematuria) appear to show that the member is much more likely to move to stage 5 than expected, but less likely to move to stage 4. All of these associations are correlations and not necessarily causal. Many could be driven by small sample sizes (see Table 3). The majority of these relationships are similar in the commercial data and across stages.

Figure 1

STAGE MODEL INTERCEPTS FOR THE MEDICARE DATA FOR A MEMBER CURRENTLY IN STAGE 3

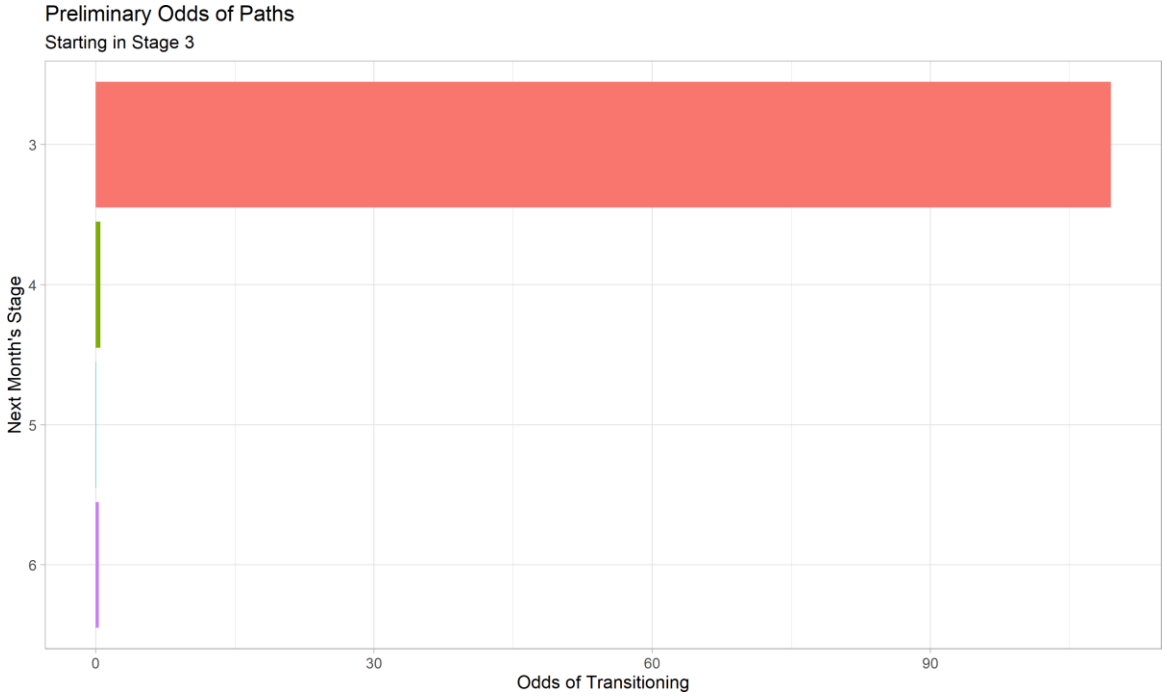
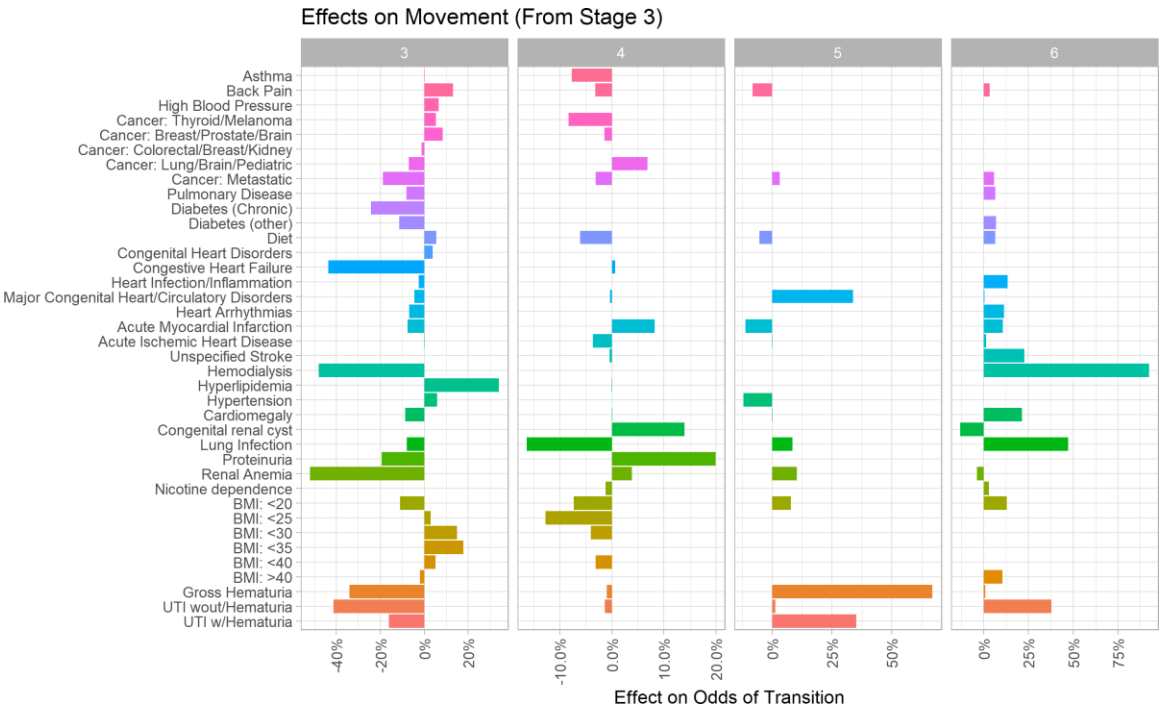


Figure 2
STAGE MODEL COEFFICIENTS FOR THE MEDICARE DATA IN STAGE 3



5.2 COST MODEL

The cost model predicts the total healthcare costs of a member during a certain month, given the CKD stage in that month.

5.2.1 MODEL SELECTION

For the cost models, we compared the logit-gamma and the Tweedie models each with three different levels of regularization (none, min, and 1 s.e.), with the XGBoost model. To compare them, we predicted the costs for each member and then computed the median absolute error. In about half of the cases (13 out of 24), the XGBoost model performed the best. In most other cases, the unregularized logit-gamma model performed the best. The regularized logit-gamma min model matched the unregularized model in many cases, meaning the optimal regularization was no regularization. The complete results are in Table 5.

Table 5
COMPARISON OF HEALTH CARE COST MODELS (MINIMUM ERROR IN BOLD)

			Gamma			Tweedie			
Dataset	Stage	Predictive Year	None	Min	1 s.e.	None	Min	1 s.e.	XGBoost
Commercial	1	2017	823	823	852	1,362	1,362	1,969	**737**
Commercial	1	2018	796	796	834	1,413	1,413	1,982	**638**
Commercial	2	2017	897	897	914	1,483	1,483	2,144	**864**
Commercial	2	2018	855	855	868	1,526	1,526	2,155	**762**
Commercial	3	2017	**1,327**	1,327	1,335	2,028	2,028	2,799	1,359
Commercial	3	2018	**1,297**	1,297	1,309	2,098	2,098	2,803	1,302
Commercial	4	2017	2,173	2,173	2,203	3,027	3,027	3,884	**2,025**

Commercial	4	2018	2,163	2,163	2,196	3,158	3,158	3,888	**1,897**
Commercial	5	2017	**2,448**	2,448	2,505	3,246	3,246	4,222	2,522
Commercial	5	2018	2,403	2,403	2,440	3,354	3,354	4,233	**2,155**
Commercial	6	2017	8,460	8,460	8,626	10,092	10,092	13,181	**7,498**
Commercial	6	2018	8,065	8,065	8,248	10,067	10,067	13,295	**5,216**
Medicare	1	2017	**965**	966	994	1,263	1,263	1,903	1,028
Medicare	1	2018	972	972	983	1,371	1,371	1,916	**846**
Medicare	2	2017	**1,098**	1,098	1,113	1,443	1,443	1,807	1,240
Medicare	2	2018	**1,079**	1,079	1,092	1,538	1,538	1,996	1,104
Medicare	3	2017	**1,389**	1,389	1,393	1,762	1,762	1,962	1,693
Medicare	3	2018	**1,387**	1,387	1,390	1,884	1,884	2,013	1,417
Medicare	4	2017	**2,177**	2,177	2,180	2,597	2,597	2,766	2,569
Medicare	4	2018	2,228	2,228	2,249	2,799	2,799	2,939	**2,121**
Medicare	5	2017	2,447	**2,445**	2,486	2,909	2,909	3,702	3,016
Medicare	5	2018	2,506	2,506	2,570	3,129	3,129	3,714	**2,482**
Medicare	6	2017	**4,075**	4,075	4,129	4,453	4,453	5,035	4,820
Medicare	6	2018	4,239	4,239	4,262	4,809	4,809	5,249	**3,735**

5.2.2 COVARIATE INFERENCE

In the logit-gamma models, essentially all the coefficient values are non-zero; however, some of the smallest values are removed to make the plot easier to read. The logit model (Figure 3) only considers the probability a month has non-zero costs. As expected, the covariates at the top (urinary tract infection, kidney transplant, gross hematuria, and metastatic cancer) are conditions that make it more likely that a member will have costs in the following month. There are not many covariates that significantly reduce the likelihood of a claim. This is consistent with the expectation that the more conditions a member has, the more likely the person is to incur medical costs.

While the logit model predicts the probability of non-zero costs, the gamma model predicts the total cost, given that it is non-zero. As Figure 4 shows, the indicators associated with the highest costs in the following month are cancer, UTI, and gross hematuria. Newborn birthweight is an interesting indicator. Medicare covers all people with ESRD, even newborns (Kirchhoff 2018). Newborns (of any birth weight) with ESRD are likely to have very high costs. Rurality has the most negative impact on cost. Members who live in more rural areas are predicted to have significantly lower costs than those in more urban settings.

Figure 3

COST MODEL COEFFICIENTS FOR THE LOGISTIC REGRESSION MODEL FOR MEDICARE DATA IN STAGE 3

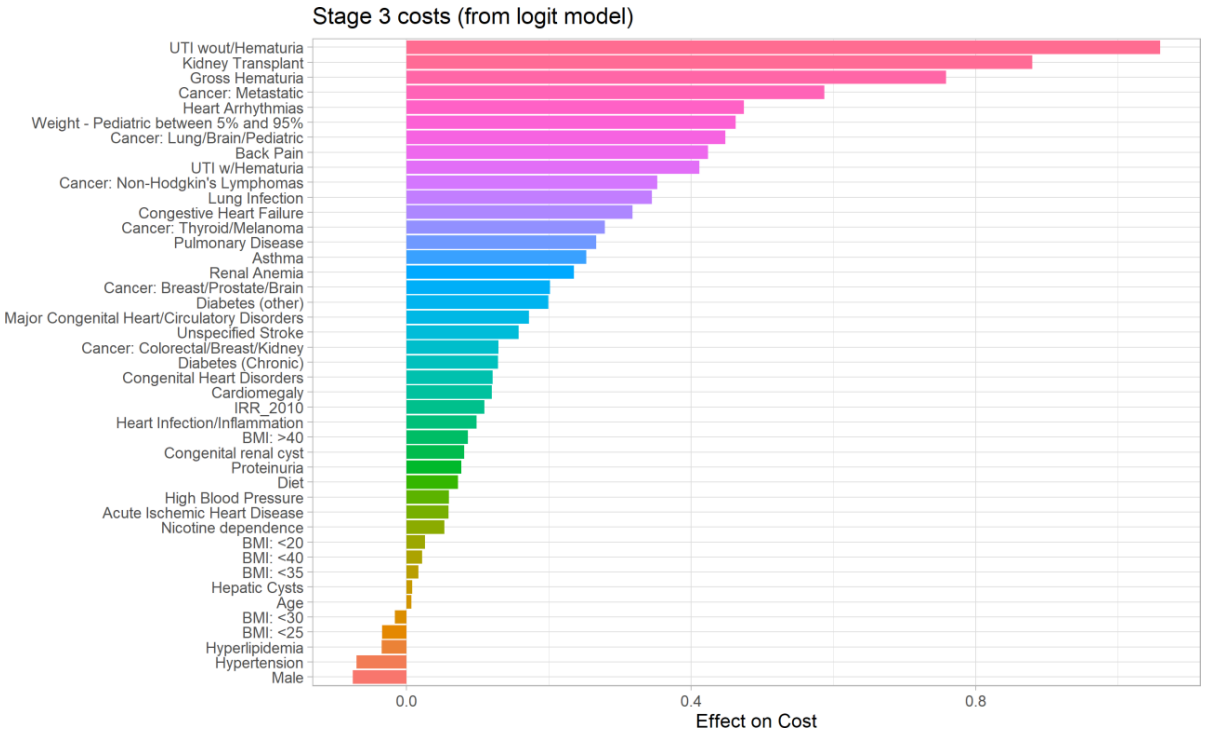
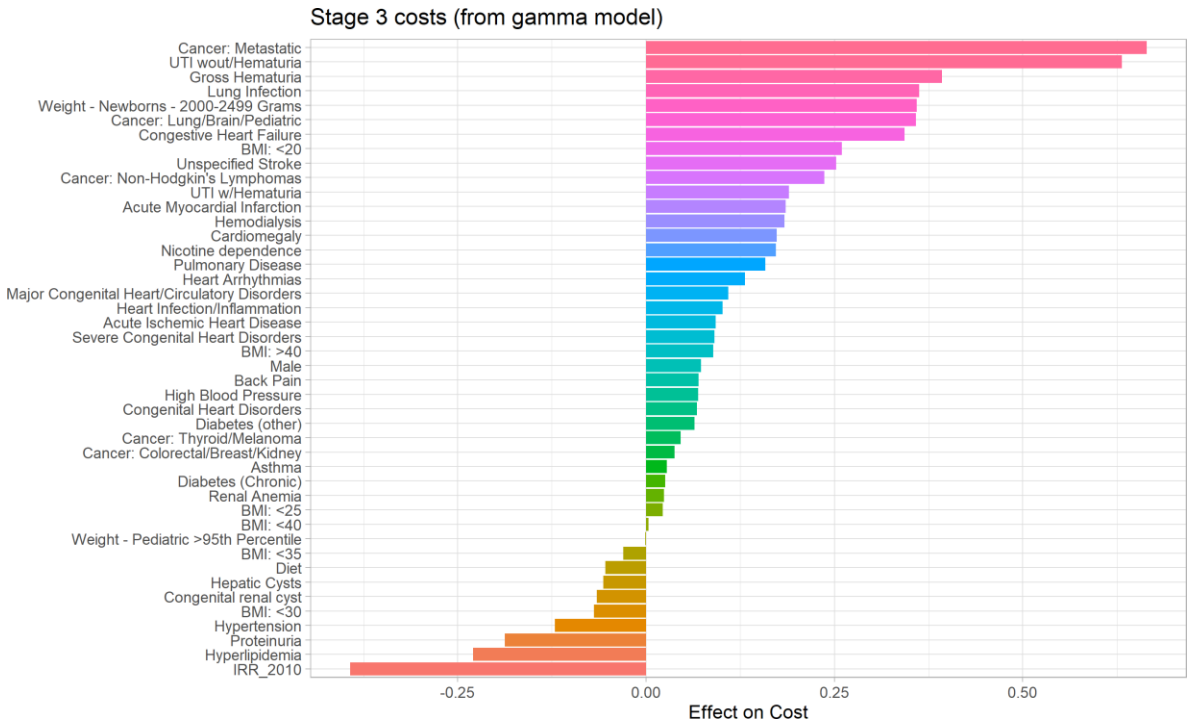


Figure 4

COST MODEL COEFFICIENTS FOR THE GAMMA MODEL FOR MEDICARE DATA IN STAGE 3



6 Conclusion

CKD impacts lives and health care systems around the world. To help those invested in preparing for and mitigating that impact in the United States, we developed a new process that explicitly models both the stage progression and the costs of members with CKD. These models can help those who need to predict and help manage costs from chronic kidney disease to be more efficient and effective. Applying our models to rich datasets of commercial and Medicare members, we found that XGBoost models perform the best in predicting both the stage and the cost of patients with CKD. If XGBoost models are unavailable, multivariate logistic regression models with regularization best predict the stage progression and logit-gamma models best predict the monthly healthcare costs of members with CKD.

7 Limitations

Because of the unique characteristics of the health care system in the United States, care should be taken before applying these models to other health systems. Additionally, other periods of study may yield different results. Also, we used diagnosis codes from claims data and did not use electronic health records to back them up. Finally, the results of some indicators may be unreliable due to small sample sizes.

8 Acknowledgments

This work is funded by a Society of Actuaries Health Care Cost Trends Strategic Research Grant. The authors appreciate Rob Bachler, Deana Bell, Lisa Charron, Gabriela Dieguez, Leah Engel, Mike Hamachek, Austin Levenson, and Karen Schenkenfelder for their work and feedback on the project.

The authors are also grateful to the Project Oversight Group members:

Ken Avner, FSA, MAAA

Joan Barrett, FSA, MAAA

Scott Kelly, FSA, MAAA

Daniel Kurowski

Rhyxian Lim,

George Omondi, ASA, MAAA, FCA

Rebecca Owen, FSA, MAAA, FCA

Alex Ryu, ASA, MAAA

and to the Society of Actuaries for their support of this project:

Achilles Natsis, FSA, MAAA

Erika Schulty

References

Bentley, T. S., and N. J. Ortner. 2020. 2020 US Organ and Tissue Transplants: Cost estimates, discussion, and emerging issues. Milliman research report. Feb. 18.

Centers for Disease Control and Prevention. 2019. Chronic Kidney Disease in the United States. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention. 2021. Chronic Kidney Disease: Common, serious, costly. Infographic. *https://www.cdc.gov/kidneydisease/pdf/CKD-common-serious-costly-h.pdf* (Aug. 23, 2021).

Centers for Medicare and Medicaid Services. 2020. End-Stage Renal Disease (ESRD). *https://www.cms.gov/Medicare/Coordination-of-Benefits-and-Recovery/Coordination-of-Benefits-and-Recovery-Overview/End-Stage-Renal-Disease-ESRD/ESRD* (Aug. 6, 2021).

Chang, Y.-T., J.-S. Hwang, S.-Y. Hung, M.-S. Tsai, J.-L. Wu, J.M. Sung and J.-D. Wang. 2016. Cost-Effectiveness of Hemodialysis and Peritoneal Dialysis: A national cohort study with 14 years follow-up and matched for comorbidities and propensity score. *Scientific Reports* 6(30,266).

Chen, P. M., T. S. Lai, P.Y. Chen, C. F. Lai, S. Y. Yang, V. Wu, C. K. Chiang, T. W. Kao, J. W. Huang, W. C. Chiang et al. 2015. Multidisciplinary Care Program for Advanced Chronic Kidney Disease: Reduces renal replacement and medical costs. *American Journal of Medicine* 128(1): 68–76.

Chen, T., and C. Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York: Association for Computing Machinery Special Interest Group on Management of Data), 785–794.

Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho et al. 2015. Xgboost: Extreme gradient boosting. *R package version 0.4-2*, 1–4.

Collins, A. J., R. N. Foley, C. Herzog, B. M. Chavers, D. Gilbertson, A. Ishani, B. L. Kasiske, J. Liu, L.-W. Mau, M. McBean et al. 2010. Excerpts from the US Renal Data System 2009 Annual Data Report. *American Journal of Kidney Diseases* 55(1): A6–A7.

Damien, P., H. J. Lanham, M. Parthasarathy and N. L. Shah. 2016. Assessing Key Cost Drivers Associated with Caring for Chronic Kidney Disease Patients. *BMC Health Services Research* 16(1): 1–10.

Drawz, P., T. H. Hostetter and M. E. Rosenberg. 2020. Slowing Progression of Chronic Kidney Disease. In *Chronic Renal Disease*. London: Academic Press, 937–959.

Drawz, P. E., and M. E. Rosenberg. 2013. Slowing Progression of Chronic Kidney Disease. *Kidney International Supplements* 3(4): 372–376.

Erickson, K. F., S. Japa, D. K. Owens, G. M. Chertow, A. M. Garber and J. D. Goldhaber-Fiebert. 2013. Cost-Effectiveness of Statins for Primary Cardiovascular Prevention in Chronic Kidney Disease. *Journal of the American College of Cardiology* 61(12): 1250–1258.

Fricks, R. B., A. Bobbio and K. S. Trivedi. 2016. Reliability Models of Chronic Kidney Disease. Paper presented at Reliability and Maintainability Symposium (RAMS), Tucson, AZ, Jan. 25–28, 2016, pp. 1–6.

Gandjour, A., W. Armsen, W. Wehmeyer, J. Multmeier and U. Tschulena. 2020. Costs of Patients with Chronic Kidney Disease in Germany. *PloS ONE* 15(4): e0231375.

Golestaneh, L., P. J. Alvarez, N. L. Reaven, S. E. Funk, K. J. McGaughey, A. Romero, M. S. Brenner and M. Onuigbo. 2017. All-Cause Costs Increase Exponentially with Increased Chronic Kidney Disease Stage. Supplement, *American Journal of Managed Care* 23(10): S163–S172.

Hoerger, T. J., J. S. Wittenborn, J. E. Segel, N. R. Burrows, K. Imai, P. Eggers, M. E. Pavkov, R. Jordan, S. M. Hailpern, A. C. Schoolwerth et al. 2010. A Health Policy Model of CKD: 1. Model construction, assumptions, and validation of health consequences. *American Journal of Kidney Diseases* 55(3): 452–462.

Honeycutt, A. A., J. E. Segel, X. Zhuo, T. J. Hoerger, K. Imai and D. Williams. 2013. Medical Costs of CKD in the Medicare Population. *Journal of the American Society of Nephrology* 24(9): 1478–1483.

Hopkins, R. B., A. X. Garg, A. Levin, A. Molzahn, C. Rigatto, J. Singer, G. Soltys, S. Soroka, P. S. Parfrey, B. J. Barrett et al. 2011. Cost-Effectiveness Analysis of a Randomized Trial Comparing Care Models for Chronic Kidney Disease. *Clinical Journal of the American Society of Nephrology* 6(6): 1248–1257.

Johnson, C. A., A. S. Levey, J. Coresh, A. Levin and J. G. L. Eknoyan. 2004. Clinical Practice Guidelines for Chronic Kidney Disease in Adults: Part 1. Definition, disease stages, evaluation, treatment, and risk factors. *American Family Physician* 70(5): 869–876.

Kerr, M., B. Bray, J. Medcalf, D. J. O'Donoghue and B. Matthews. 2012. Estimating the Financial Cost of Chronic Kidney Disease to the NHS in England. Supplement, *Nephrology Dialysis Transplantation* 27(3), iii73–iii80.

Khan, S., and C. A. Amedia Jr. 2008. Economic Burden of Chronic Kidney Disease. *Journal of Evaluation in Clinical Practice* 14(3): 422–434.

Kirchhoff, S. M. 2018. Medicare Coverage of End-Stage Renal Disease (ESRD). CRS Report R45290. Congressional Research Service, Aug. 16, *https://fas.org/sgp/crs/misc/R45290.pdf.*

Kondo, M., K. Yamagata, S.-L. Hoshi, C. Saito, K. Asahi, T. Moriyama, K. Tsuruya, H. Yoshida, K. Iseki and T. Watanabe. 2012. Cost-Effectiveness of Chronic Kidney Disease Mass Screening Test in Japan. *Clinical and Experimental Nephrology* 16(2): 279–291.

Laliberté, F., B. K. Bookhart, F. Vekeman, M. Corral, M. S. Duh, R. A. Bailey, C. T. Piech and P. Lefebvre. 2009. Direct All-Cause Health Care Costs Associated with Chronic Kidney Disease in Patients with Diabetes and Hypertension: A managed care perspective. *Journal of Managed Care Pharmacy* 15(4): 312–322.

Levey, A. S., C. Becker and L. A. Inker. 2015. Glomerular Filtration Rate and Albuminuria for Detection and Staging of Acute and Chronic Kidney Disease in Adults: A systematic review. *Journal of the American Medical Association* 313(8): 837–846.

Levey, A. S., and J. Coresh. 2012. Chronic Kidney Disease. *The Lancet* 379(9811): 165–180.

Manns, B., B. Hemmelgarn, M. Tonelli, F. Au, T. C. Chiasson, J. Dong and S. Klarenbach. 2010. Population Based Screening for Chronic Kidney Disease: Cost effectiveness study. *BMJ* 341 (Nov. 8).

Manns, B., B. Hemmelgarn, M. Tonelli, F. Au, H. So, R. Weaver, A. E. Quinn and S. Klarenbach. 2019. The Cost of Care for People with Chronic Kidney Disease. *Canadian Journal of Kidney Health and Disease* 6, doi: 10.1177/2054358119835521.

National Institute of Diabetes and Digestive and Kidney Diseases. 2021. Slow Progression and Reduce Complications. *Health Information*, https://www.niddk.nih.gov/health-information/professionals/clinic*al-tools-patient-management/kidney-disease/identify-manage-patients/manage-ckd/slow-progression-reduce-complications* (Aug. 19, 2021).

Orlando, L. A., E. J. Belasco, U. D. Patel and D. B. Matchar. 2011. The Chronic Kidney Disease Model: A general purpose model of disease progression and treatment. *BMC Medical Informatics and Decision Making* 11(1): 41.

Saran, R., A. Pearson, A. Tilea, V. Shahinian, J. Bragg-Gresham, M. Heung, D. W. Hutton, D. Steffick, K. Zheng, H. Morgenstern et al. 2021. Burden and Cost of Caring for US Veterans With CKD: Initial findings from the VA Renal Information System (VA-REINS). *American Journal of Kidney Diseases* 77(3): 397–405.

Tangri, N., L. A. Stevens, J. Griffith, H. Tighiouart, O. Djurdjev, D. Naimark, A. Levin and A. S. Levey. 2011. A Predictive Model for Progression of Chronic Kidney Disease to Kidney Failure. *Journal of the American Medical Association* 305(15): 1553–1559.

Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B (Methodological)* 58(1): 267–288.

Waldorf, B., and A. Kim. 2018. The Index of Relative Rurality (IRR): US county data for 2000 and 2010. Version 1.0. Purdue University Research Repository, *https://purr.purdue.edu/publications/2960/1*.

Wei, S.-Y., Y.-Y. Chang, L.-W. Mau, M.-Y. Lin, H.-C. Chiu, J.-C. Tsai, C.-J. Huang, H.-C. Chen and S.-J. Hwang. 2010. Chronic Kidney Disease Care Program Improves Quality of Pre-End-Stage Renal Disease Care and Reduces Medical Costs. *Nephrology* 15(1): 108–115.

Appendix A. Xgboost Hyperparameter Settings

For both the cost and stage models, we optimized four different hyperparameters (T. Chen et al. 2015), listed here with the range of each in brackets.

- nrounds [100, 500]—the number of trees in each model (each one built sequentially)
- max depth [1, 10]—maximum number of splits in each tree
- eta [0.1, 0.5]—step size shrinkage, used to prevent overfitting
- lambda [1, 10]—L2 regularization, also to prevent overfitting

We randomly selected 200 sets of the listed hyperparameters listed and compared them using fourfold cross validation on the training set.

Appendix B. ICD-10 Codes by Indicator Flag

Table B.1

LIST OF CORRESPONDING ICD-10 CODES BY INDICATOR

Category Description	ICD-10 Codes
CKD stage 1	N181
CKD stage 2	N182
CKD stage 3	N183
CKD stage 4	N184
CKD stage 5	N185
CKD stage 6 (ESRD)	N186
Acute myocardial infarction	I2101, I2102, I2109, I2111, I2119, I2121, I2129, I213, I214, I219, I21A1, I21A9, I220, I221, I222, I228, I229, I234, I235, I511, I512
Aspiration and specified bacterial pneumonias and other severe lung infections	A065, A481, A5004, B380, B381, B382, B390, B391, B392, B400, B401, B402, B410, B664, B671, J150, J151, J1520, J15211, J15212, J1529, J155, J156, J158, J690, J691, J698, J850, J851, J852, J853, J860, J869, P230, P231, P232, P233, P234, P235, P236, P238, P239, P2401, P2411, P2421, P2431, P2481
Asthma	J4520, J4521, J4522, J4530, J4531, J4532, J4540, J4541, J4542, J4550, J4551, J4552, J45901, J45902, J45909, J45990, J45991, J45998, J410, J411, J418, J42, J440, J441
Atrial and ventricular septal defects, patent ductus arteriosus and other congenital heart/circulatory disorders	Q206, Q209, Q210, Q211, Q214, Q218, Q219, Q221, Q222, Q223, Q249, Q250, Q265, Q266, Q270, Q271, Q272, Q2730, Q2731, Q2732, Q2733, Q2734, Q2739, Q274, Q278, Q279, Q280, Q281, Q288, Q289, Q893
Back pain	M545
Breast (age 50+) and prostate cancer, benign/uncertain brain tumors and other cancers and tumors	C4A0, C4A10, C4A11, C4A111, C4A112, C4A12, C4A121, C4A122, C4A20, C4A21, C4A22, C4A30, C4A31, C4A39, C4A4, C4A51, C4A52, C4A59, C4A60, C4A61, C4A62, C4A70, C4A71, C4A72, C4A8, C4A9, C510, C511, C512, C518, C519, C52, C530, C531, C538, C539, C540, C541, C542, C543, C548, C549, C55, C577, C578, C579, C61, C661, C662, C669, C670, C671, C672, C673, C674, C675, C676, C677, C678, C679, C680, C681, C688, C689, C6900, C6901, C6902, C6910, C6911, C6912, C6920, C6921, C6922, C6930, C6931, C6932, C6940, C6941, C6942, C6950, C6951, C6952, C6960, C6961, C6962, C6980, C6981, C6982, C6990, C6991, C6992, C760, C761, C762, C763, C7640, C7641, C7642, C7650, C7651, C7652, C768, C7A00, C7A010, C7A011, C7A012, C7A019, C7A020, C7A021, C7A022, C7A023, C7A024, C7A025, C7A026, C7A029, C7A090, C7A091, C7A092, C7A093, C7A094, C7A095, C7A096, C7A098, C7A1, C7A8, C802, C8100, C8101, C8102, C8103, C8104, C8105, C8106, C8107, C8108, C8109, C8110, C8111, C8112, C8113, C8114, C8115, C8116, C8117, C8118, C8119, C8120, C8121, C8122, C8123, C8124, C8125, C8126, C8127, C8128, C8129, C8130, C8131, C8132, C8133, C8134, C8135, C8136, C8137, C8138, C8139, C8140, C8141, C8142, C8143, C8144, C8145, C8146, C8147, C8148, C8149, C8170, C8171, C8172, C8173, C8174, C8175, C8176, C8177, C8178, C8179, C8190, C8191, C8192, C8193, C8194, C8195, C8196, C8197, C8198, C8199, D1802, D320, D321, D329, D330, D331, D332, D333, D334, D337, D339, D352, D353, D354, D420, D421, D429, D430, D431, D432, D433, D434, D438, D439, D443, D444, D445, D446, D447, D496, Q851, Q858, Q859, C50011, C50012, C50019, C50021, C50022, C50029, C50111, C50112, C50119, C50121, C50122, C50129, C50211, C50212, C50219, C50221, C50222, C50229, C50311, C50312, C50319, C50321, C50322, C50329, C50411, C50412, C50419, C50421, C50422, C50429, C50511, C50512, C50519, C50521, C50522, C50529, C50611, C50612, C50619, C50621, C50622, C50629, C50811, C50812, C50819, C50821, C50822, C50829, C50911, C50912, C50919, C50921, C50922, C50929
Cardiomegaly	I517
Chronic obstructive pulmonary disease, including bronchiectasis	J410, J411, J418, J42, J430, J431, J432, J438, J439, J440, J441, J449, J470, J471, J479, J982, J983

Colorectal, breast (age < 50), kidney and other cancers	C01, C020, C021, C022, C023, C024, C028, C029, C030, C031, C039, C040, C041, C048, C049, C050, C051, C052, C058, C059, C060, C061, C062, C0680, C0689, C069, C07, C080, C081, C089, C090, C091, C098, C099, C100, C101, C102, C103, C104, C108, C109, C110, C111, C112, C113, C118, C119, C12, C130, C131, C132, C138, C139, C140, C142, C148, C180, C181, C182, C183, C184, C185, C186, C187, C188, C189, C19, C20, C210, C211, C212, C218, C260, C261, C269, C300, C301, C310, C311, C312, C313, C318, C319, C320, C321, C322, C323, C328, C329, C37, C381, C382, C383, C388, C390, C399, C50011, C50012, C50019, C50021, C50022, C50029, C50111, C50112, C50119, C50121, C50122, C50129, C50211, C50212, C50219, C50221, C50222, C50229, C50311, C50312, C50319, C50321, C50322, C50329, C50411, C50412, C50419, C50421, C50422, C50429, C50511, C50512, C50519, C50521, C50522, C50529, C50611, C50612, C50619, C50621, C50622, C50629, C50811, C50812, C50819, C50821, C50822, C50829, C50911, C50912, C50919, C50921, C50922, C50929, C561, C562, C569, C5700, C5701, C5702, C5710, C5711, C5712, C5720, C5721, C5722, C573, C574, C58, C641, C642, C649, C651, C652, C659
Congenital renal cyst	Q6100, Q6101, Q6102
Congestive heart failure	A3681, B3324, I0981, I110, I130, I132, I2601, I2602, I2609, I270, I271, I2720, I2721, I2722, I2723, I2724, I2729, I2781, I2783, I2789, I279, I280, I281, I288, I289, I420, I421, I422, I423, I424, I425, I426, I427, I428, I429, I43, I501, I5020, I5021, I5022, I5023, I5030, I5031, I5032, I5033, I5040, I5041, I5042, I5043, I50810, I50811, I50812, I50813, I50814, I5082, I5083, I5084, I5089, I509, I514, I515
Diabetes with chronic complications	E0821, E0822, E0829, E08311, E08319, E083211, E083212, E083213, E083219, E083291, E083292, E083293, E083299, E083311, E083312, E083313, E083319, E083391, E083392, E083393, E083399, E083411, E083412, E083413, E083419, E083491, E083492, E083493, E083499, E083511, E083512, E083513, E083519, E083521, E083522, E083523, E083529, E083531, E083532, E083533, E083539, E083541, E083542, E083543, E083549, E083551, E083552, E083553, E083559, E083591, E083592, E083593, E083599, E0836, E0837X1, E0837X2, E0837X3, E0837X9, E0839, E0840, E0841, E0842, E0843, E0844, E0849, E0851, E0852, E0859, E08610, E08618, E08620, E08621, E08622, E08628, E08630, E08638, E08649, E0865, E0869, E088, E0921, E0922, E0929, E09311, E09319, E093211, E093212, E093213, E093219, E093291, E093292, E093293, E093299, E093311, E093312, E093313, E093319, E093391, E093392, E093393, E093399, E093411, E093412, E093413, E093419, E093491, E093492, E093493, E093499, E093511, E093512, E093513, E093519, E093521, E093522, E093523, E093529, E093531, E093532, E093533, E093539, E093541, E093542, E093543, E093549, E093551, E093552, E093553, E093559, E093591, E093592, E093593, E093599, E0936, E0937X1, E0937X2, E0937X3, E0937X9, E0939, E0940, E0941, E0942, E0943, E0944, E0949, E0951, E0952, E0959, E09610, E09618, E09620, E09621, E09622, E09628, E09630, E09638, E09649, E0965, E0969, E098, E1021, E1022, E1029, E10311, E10319, E103211, E103212, E103213, E103219, E103291, E103292, E103293, E103299, E103311, E103312, E103313, E103319, E103391, E103392, E103393, E103399, E103411, E103412, E103413, E103419, E103491, E103492, E103493, E103499, E103511, E103512, E103513, E103519, E103521, E103522, E103523, E103529, E103531, E103532, E103533, E103539, E103541, E103542, E103543, E103549, E103551, E103552, E103553, E103559, E103591, E103592, E103593, E103599, E1036, E1037X1, E1037X2, E1037X3, E1037X9, E1039, E1040, E1041, E1042, E1043, E1044, E1049, E1051, E1052, E1059, E10610, E10618, E10620, E10621, E10622, E10628, E10630, E10638, E10649, E1065, E1069, E108, E1121, E1122, E1129, E11311, E11319, E113211, E113212, E113213, E113219, E113291, E113292, E113293, E113299, E113311, E113312, E113313, E113319, E113391, E113392, E113393, E113399, E113411, E113412, E113413, E113419, E113491, E113492, E113493, E113499, E113511, E113512, E113513, E113519, E113521, E113522, E113523, E113529, E113531, E113532, E113533, E113539, E113541, E113542, E113543, E113549, E113551, E113552, E113553, E113559, E113591, E113592, E113593, E113599, E1136, E1137X1, E1137X2, E1137X3, E1137X9, E1139, E1140, E1141, E1142, E1143, E1144, E1149, E1151, E1152, E1159, E11610, E11618, E11620, E11621, E11622, E11628, E11630, E11638, E11649, E1165, E1169, E118, E1321, E1322, E1329, E13311, E13319, E133211, E133212, E133213, E133219, E133291, E133292, E133293, E133299, E133311, E133312, E133313, E133319, E133391, E133392, E133393, E133399, E133411, E133412,

	E133413, E133419, E133491, E133492, E133493, E133499, E133511, E133512, E133513, E133519, E133521, E133522, E133523, E133529, E133531, E133532, E133533, E133539, E133541, E133542, E133543, E133549, E133551, E133552, E133553, E133559, E133591, E133592, E133593, E133599, E1336, E1337X1, E1337X2, E1337X3, E1337X9, E1339, E1340, E1341, E1342, E1343, E1344, E1349, E1351, E1352, E1359, E13610, E13618, E13620, E13621, E13622, E13628, E13630, E13638, E13649, E1365, E1369, E138
Diabetes without complication	E089, E099, E109, E119, E139, Z794
Diet	Z713, Z724
Gross hematuria	R310
Heart infection/ inflammation, except rheumatic	A0102, A3282, A381, A3950, A3951, A3952, A3953, A5054, A5200, A5201, A5202, A5203, A5206, A5209, A5483, B2682, B3320, B3321, B3322, B3323, B376, B5881, D8685, I300, I301, I308, I309, I310, I311, I312, I313, I314, I318, I319, I32, I330, I339, I400, I401, I408, I409, I41
Hemodialysis	Z992
Hepatic cysts	Q446
High blood pressure, not diagnosed as hypertension	R030
Hyperlipidemia	E782, E783, E784, E7841, E7849, E785
Hypertension	I10
Hypoplastic left heart syndrome and other severe congenital heart disorders	Q204, Q224, Q226, Q228, Q229, Q234
Ischemic or unspecified stroke	I6300, I63011, I63012, I63013, I63019, I6302, I63031, I63032, I63033, I63039, I6309, I6310, I63111, I63112, I63113, I63119, I6312, I63131, I63132, I63133, I63139, I6319, I6320, I63211, I63212, I63213, I63219, I6322, I63231, I63232, I63233, I63239, I6329, I6330, I63311, I63312, I63313, I63319, I63321, I63322, I63323, I63329, I63331, I63332, I63333, I63339, I63341, I63342, I63343, I63349, I6339, I6340, I63411, I63412, I63413, I63419, I63421, I63422, I63423, I63429, I63431, I63432, I63433, I63439, I63441, I63442, I63443, I63449, I6349, I6350, I63511, I63512, I63513, I63519, I63521, I63522, I63523, I63529, I63531, I63532, I63533, I63539, I63541, I63542, I63543, I63549, I6359, I636, I638, I6381, I6389, I639
Kidney transplant	Z940
Lung, brain and other severe cancers, including pediatric acute lymphoid leukemia	C153, C154, C155, C158, C159, C160, C161, C162, C163, C164, C165, C166, C168, C169, C170, C171, C172, C173, C178, C179, C220, C221, C222, C223, C224, C227, C228, C229, C23, C240, C241, C248, C249, C250, C251, C252, C253, C254, C257, C258, C259, C33, C3400, C3401, C3402, C3410, C3411, C3412, C342, C3430, C3431, C3432, C3480, C3481, C3482, C3490, C3491, C3492, C384, C450, C451, C452, C457, C459, C480, C481, C482, C488, C700, C701, C709, C710, C711, C712, C713, C714, C715, C716, C717, C718, C719, C720, C721, C7220, C7221, C7222, C7230, C7231, C7232, C7240, C7241, C7242, C7250, C7259, C729, C7400, C7401, C7402, C7410, C7411, C7412, C7490, C7491, C7492, C751, C752, C753, C9000, C9001, C9002, C9010, C9011, C9012, C9020, C9021, C9022, C9210, C9211, C9212, C9220, C9221, C9222, C9230, C9231, C9232, C9290, C9291, C9292, C92Z0, C92Z1, C92Z2, C9310, C9311, C9312, C9330, C9331, C9332, C9390, C9391, C9392, C93Z0, C93Z1, C93Z2, C9430, C9431, C9432, C9480, C9481, C9482, C9100, C9101, C9102, C9500, C9501, C9502
Major congenital heart/circulatory disorders	P2930, P2938, Q200, Q201, Q202, Q203, Q205, Q208, Q212, Q213, Q220, Q225, Q230, Q231, Q232, Q233, Q238, Q239, Q240, Q241, Q242, Q243, Q244, Q245, Q246, Q248, Q251, Q2521, Q2529, Q253, Q2540, Q2541, Q2542, Q2543, Q2544, Q2545, Q2546, Q2547, Q2548, Q2549, Q255, Q256, Q2571, Q2572, Q2579, Q258, Q259, Q260, Q261, Q262, Q263, Q264, Q268, Q269
Metastatic cancer	C770, C771, C772, C773, C774, C775, C778, C779, C7800, C7801, C7802, C781, C782, C7830, C7839, C784, C785, C786, C787, C7880, C7889, C7900, C7901, C7902, C7910, C7911, C7919, C792, C7931, C7932, C7940, C7949, C7951, C7952, C7960, C7961, C7962, C7970, C7971, C7972, C7981, C7982, C7989, C799, C7B00, C7B01, C7B02, C7B03, C7B04, C7B09, C7B1, C7B8, C800, C9100, C9101, C9102, C9200, C9201, C9202, C9240,

	C9241, C9242, C9250, C9251, C9252, C9260, C9261, C9262, C92A0, C92A1, C92A2, C9300, C9301, C9302, C9400, C9401, C9402, C9420, C9421, C9422, C9440, C9441, C9442, C9500, C9501, C9502, E883
Nicotine dependence	F17200, F17201, F17203, F17208, F17209, F17210, F17211, F17213, F17218, F17219, F17220, F17221, F17223, F17228, F17229, F17290, F17291, F17293, F17298, F17299, T65211A, T65211D, T65211S, T65212A, T65212D, T65212S, T65213A, T65213D, T65213S, T65214A, T65214D, T65214S, T65221A, T65221D, T65221S, T65222A, T65222D, T65222S, T65223A, T65223D, T65223S, T65224A, T65224D, T65224S, T65291A, T65291D, T65291S, T65292A, T65292D, T65292S, T65293A, T65293D, T65293S, T65294A, T65294D, T65294S, Z122, Z716, Z720, Z87891
Non-Hodgkin's lymphomas and other cancers and tumors	C380, C4000, C4001, C4002, C4010, C4011, C4012, C4020, C4021, C4022, C4030, C4031, C4032, C4080, C4081, C4082, C4090, C4091, C4092, C410, C411, C412, C413, C414, C419, C460, C461, C462, C463, C464, C4650, C4651, C4652, C467, C469, C470, C4710, C4711, C4712, C4720, C4721, C4722, C473, C474, C475, C476, C478, C479, C490, C4910, C4911, C4912, C4920, C4921, C4922, C493, C494, C495, C496, C498, C499, C49A0, C49A1, C49A2, C49A3, C49A4, C49A5, C49A9, C8200, C8201, C8202, C8203, C8204, C8205, C8206, C8207, C8208, C8209, C8210, C8211, C8212, C8213, C8214, C8215, C8216, C8217, C8218, C8219, C8220, C8221, C8222, C8223, C8224, C8225, C8226, C8227, C8228, C8229, C8230, C8231, C8232, C8233, C8234, C8235, C8236, C8237, C8238, C8239, C8240, C8241, C8242, C8243, C8244, C8245, C8246, C8247, C8248, C8249, C8250, C8251, C8252, C8253, C8254, C8255, C8256, C8257, C8258, C8259, C8260, C8261, C8262, C8263, C8264, C8265, C8266, C8267, C8268, C8269, C8280, C8281, C8282, C8283, C8284, C8285, C8286, C8287, C8288, C8289, C8290, C8291, C8292, C8293, C8294, C8295, C8296, C8297, C8298, C8299, C8300, C8301, C8302, C8303, C8304, C8305, C8306, C8307, C8308, C8309, C8310, C8311, C8312, C8313, C8314, C8315, C8316, C8317, C8318, C8319, C8330, C8331, C8332, C8333, C8334, C8335, C8336, C8337, C8338, C8339, C8350, C8351, C8352, C8353, C8354, C8355, C8356, C8357, C8358, C8359, C8370, C8371, C8372, C8373, C8374, C8375, C8376, C8377, C8378, C8379, C8380, C8381, C8382, C8383, C8384, C8385, C8386, C8387, C8388, C8389, C8390, C8391, C8392, C8393, C8394, C8395, C8396, C8397, C8398, C8399, C8400, C8401, C8402, C8403, C8404, C8405, C8406, C8407, C8408, C8409, C8410, C8411, C8412, C8413, C8414, C8415, C8416, C8417, C8418, C8419, C8440, C8441, C8442, C8443, C8444, C8445, C8446, C8447, C8448, C8449, C8460, C8461, C8462, C8463, C8464, C8465, C8466, C8467, C8468, C8469, C8470, C8471, C8472, C8473, C8474, C8475, C8476, C8477, C8478, C8479, C8490, C8491, C8492, C8493, C8494, C8495, C8496, C8497, C8498, C8499, C84A0, C84A1, C84A2, C84A3, C84A4, C84A5, C84A6, C84A7, C84A8, C84A9, C84Z0, C84Z1, C84Z2, C84Z3, C84Z4, C84Z5, C84Z6, C84Z7, C84Z8, C84Z9, C8510, C8511, C8512, C8513, C8514, C8515, C8516, C8517, C8518, C8519, C8520, C8521, C8522, C8523, C8524, C8525, C8526, C8527, C8528, C8529, C8580, C8581, C8582, C8583, C8584, C8585, C8586, C8587, C8588, C8589, C8590, C8591, C8592, C8593, C8594, C8595, C8596, C8597, C8598, C8599, C860, C861, C862, C863, C864, C865, C866, C880, C882, C883, C884, C888, C889, C9030, C9031, C9032, C9110, C9111, C9112, C9130, C9131, C9132, C9140, C9141, C9142, C9150, C9151, C9152, C9160, C9161, C9162, C9190, C9191, C9192, C91A0, C91A1, C91A2, C91Z0, C91Z1, C91Z2, C9510, C9511, C9512, C9590, C9591, C9592, C960, C9620, C9621, C9622, C9629, C964, C965, C966, C969, C96A, C96Z, D151, C7400, C7401, C7402, C7410, C7411, C7412, C7490, C7491, C7492
Proteinuria	R809, R800, R801, R802, R803, R808
Renal anemia	D631
Specified heart arrhythmias	I442, I470, I471, I472, I479, I480, I481, I482, I483, I484, I4891, I4892, I492, I495
Thyroid cancer, melanoma, neurofibromatosis and other cancers and tumors	C430, C4310, C4311, C43111, C43112, C4312, C43121, C43122, C4320, C4321, C4322, C4330, C4331, C4339, C434, C4351, C4352, C4359, C4360, C4361, C4362, C4370, C4371, C4372, C438, C439, C600, C601, C602, C608, C609, C6200, C6201, C6202, C6210, C6211, C6212, C6290, C6291, C6292, C6300, C6301, C6302, C6310, C6311, C6312, C632, C637, C638, C639, C73, C750, C754, C755, C758, C759, C801, D030, D0310, D0311, D03111,

	D03112, D0312, D03121, D03122, D0320, D0321, D0322, D0330, D0339, D034, D0351, D0352, D0359, D0360, D0361, D0362, D0370, D0371, D0372, D038, D039, E340, Q8500, Q8501, Q8502, Q8503, Q8509
Unstable angina and other acute ischemic heart disease	I200, I230, I231, I232, I233, I236, I237, I238, I240, I241, I248, I249, I25110, I25700, I25710, I25720, I25730, I25750, I25760, I25790
Urinary tract infection with hematuria	N3001, N3011, N3021, N3031, N3041, N3081, N3091
Urinary tract infection without hematuria	N3000, N3010, N3020, N3030, N3040, N3080, N3090, N341, N342, A5401, A5601, A5903, B3741, D8684, N10, A0225, N110, N111, N760, N761, N390
Weight: BMI 19.9 or less	Z681
Weight: BMI 20–24.9	Z6820, Z6821, Z6822, Z6823, Z6824
Weight: BMI 25–29.9	Z6825, Z6826, Z6827, Z6828, Z6829
Weight: BMI 30–34.9	Z6830, Z6831, Z6832, Z6833, Z6834
Weight: BMI 35–39.9	Z6835, Z6836, Z6837, Z6838, Z6839
Weight: BMI 40 or greater	Z6841, Z6842, Z6843, Z6844, Z6845
Weight, newborns: < 500 g	P0501, P0511, P0701
Weight, newborns: 500–749 g	P0502, P0512, P0702, P0721, P0722, P0723 P0503, P0513, P0703, P0724, P0725
Weight, newborns: 750–999 g	P0500, P0509, P0510, P0519, P052, P059, P0700, P0710, P0730, P0738, P0739, Z3830, Z3831, Z384, Z385
Weight, newborns: 1,000–1,499 g	P0504, P0505, P0514, P0515, P0714, P0715, P0720, P0726, P0731
Weight, newborns: 1,500–1,999 g	P0506, P0507, P0516, P0517, P0716, P0717, P0732, P0733, P0734, P0735
Weight, newborns: 2,000–2,499 g	P0508, P0518, P0718, P0736, P0737, Z3861, Z3862, Z3863, Z3864, Z3865, Z3866, Z3868, Z3869, Z387, Z388, Q894
Weight, newborns: Other low-birthweight issues	P0500, P0509, P0510, P0519, P052, P059, P0700, P0710, P0730, P0738, P0739, Z3830, Z3831, Z384, Z385
Weight, pediatric: < 5th percentile	Z6851
Weight, pediatric: 5th–95th percentile	Z6852, Z6853
Weight, pediatric: > 95th percentile	Z6854

Appendix C. Complete Stage Model Results

C.1 COMMERCIAL: MULTINOMIAL LOGISTIC MODEL

The following pages present results of using the commercial stage transition model starting in stage 1 (Figure C.1), stage 2 (Figure C.2), stage 3 (Figure C.3), stage 4 (Figure C.4) and stage 5 (Figure C.5).

Figure C.1

COMMERCIAL STAGE TRANSITION MODEL STARTING IN STAGE 1

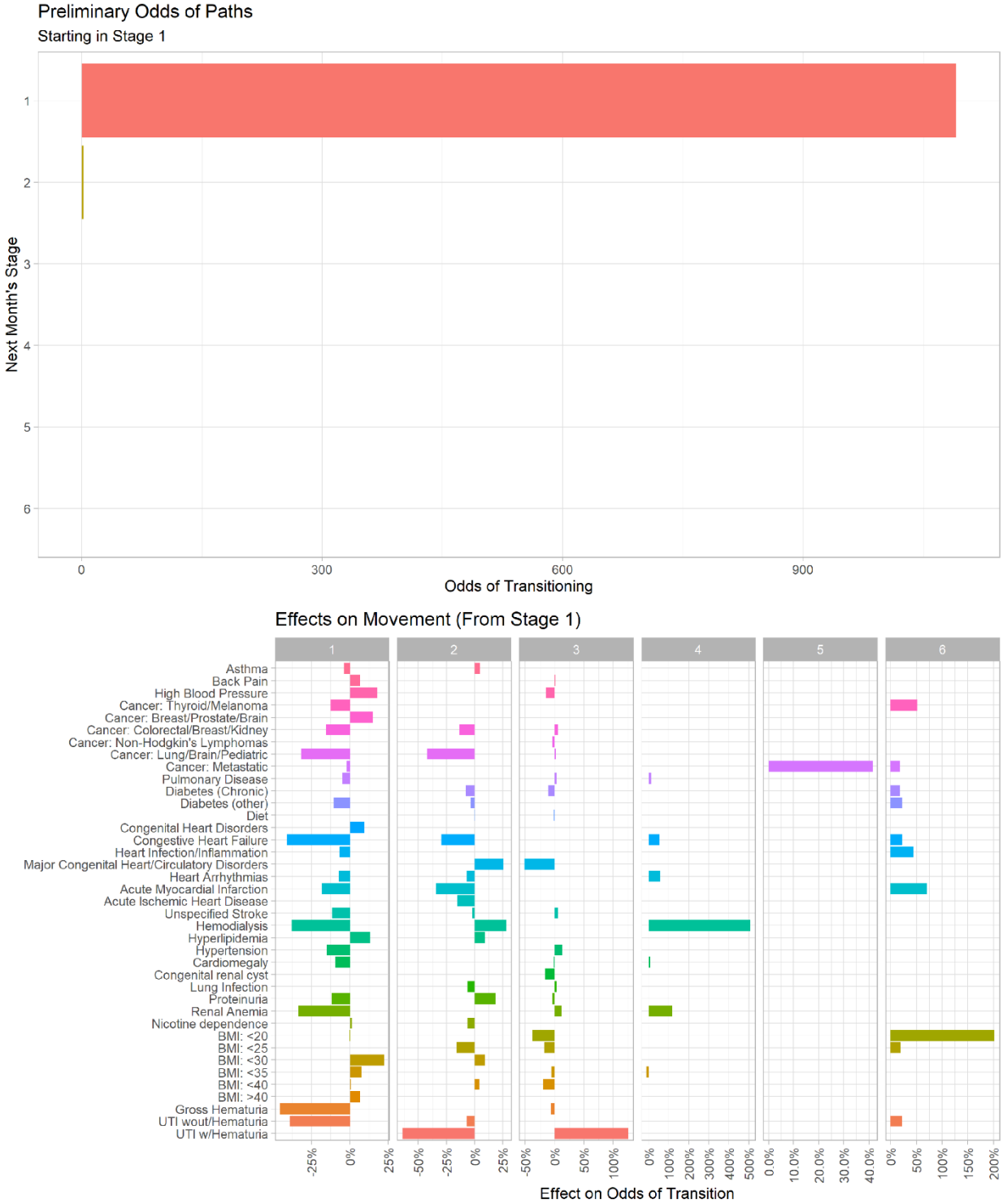


Figure C.2

COMMERCIAL STAGE TRANSITION MODEL STARTING IN STAGE 2

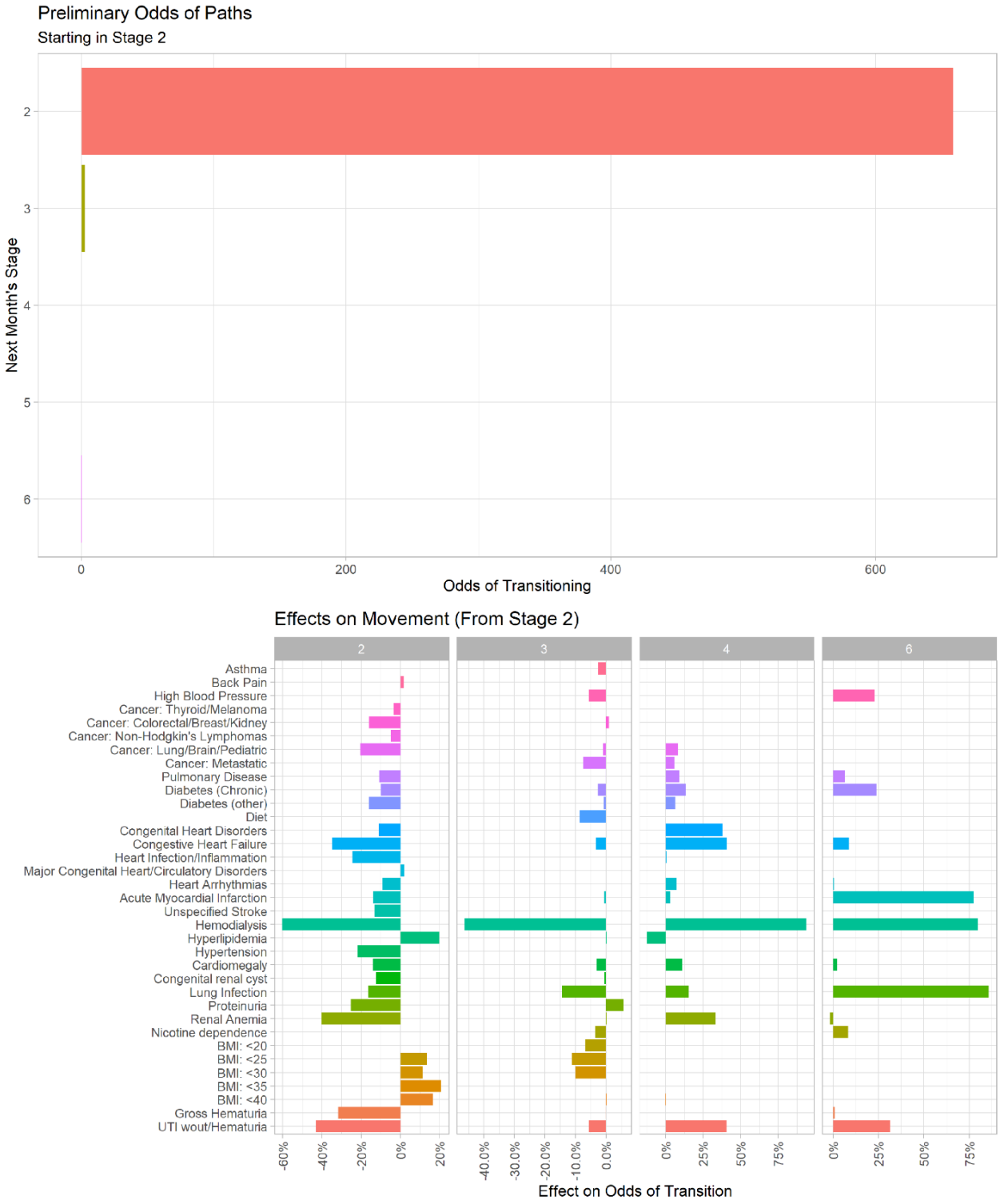


Figure C.3
COMMERCIAL STAGE TRANSITION MODEL STARTING IN STAGE 3

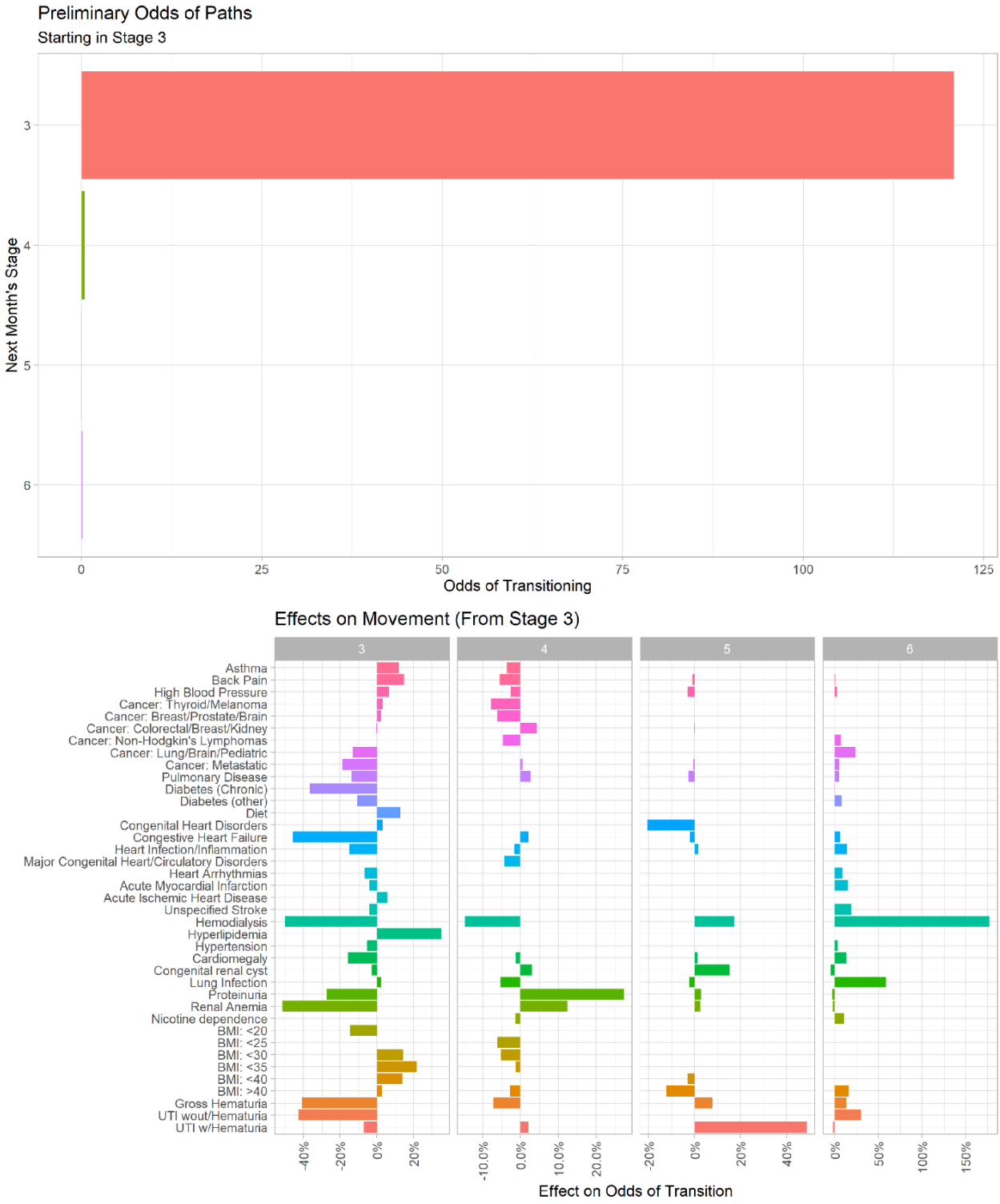


Figure C.4

COMMERCIAL STAGE TRANSITION MODEL STARTING IN STAGE 4

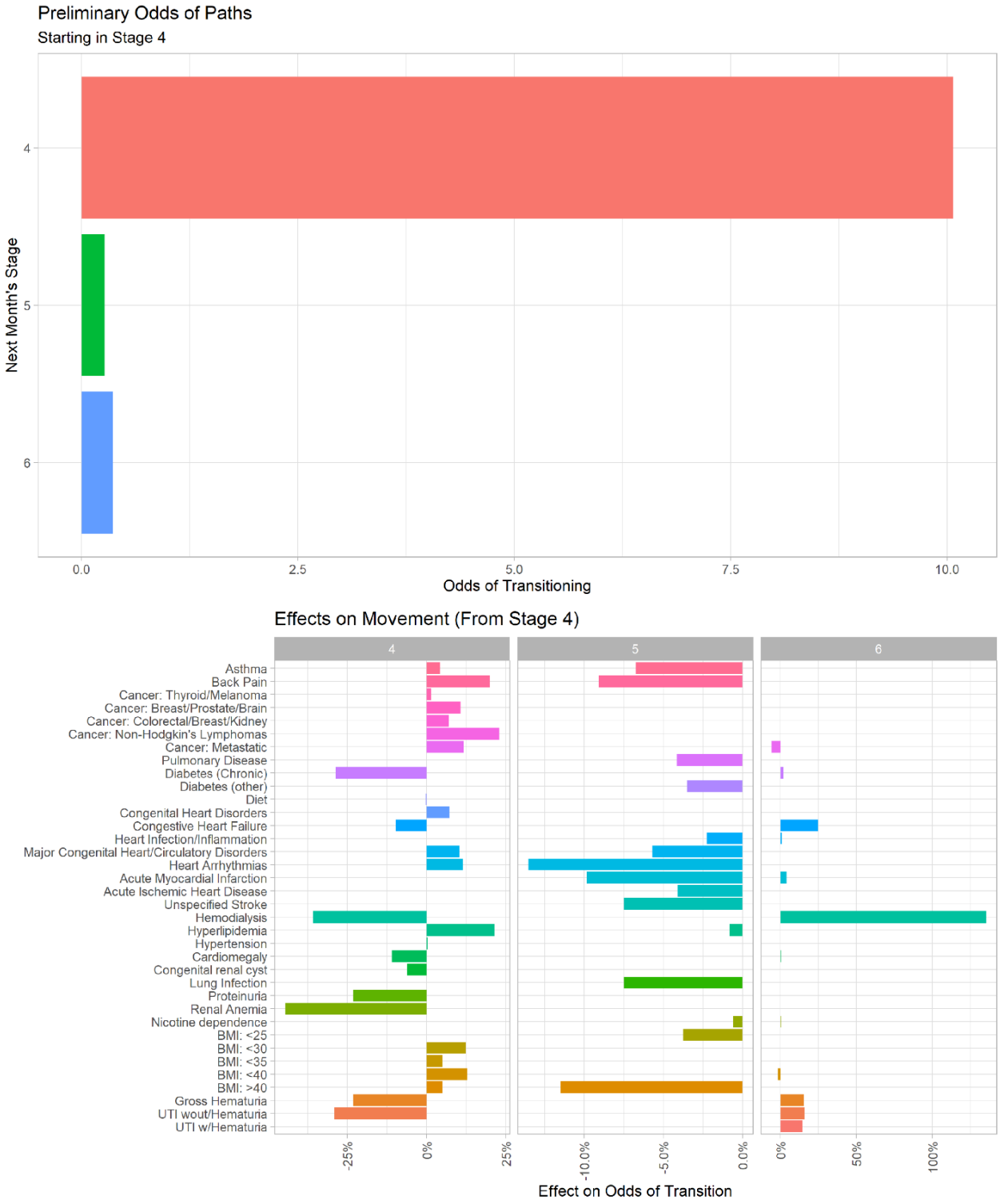
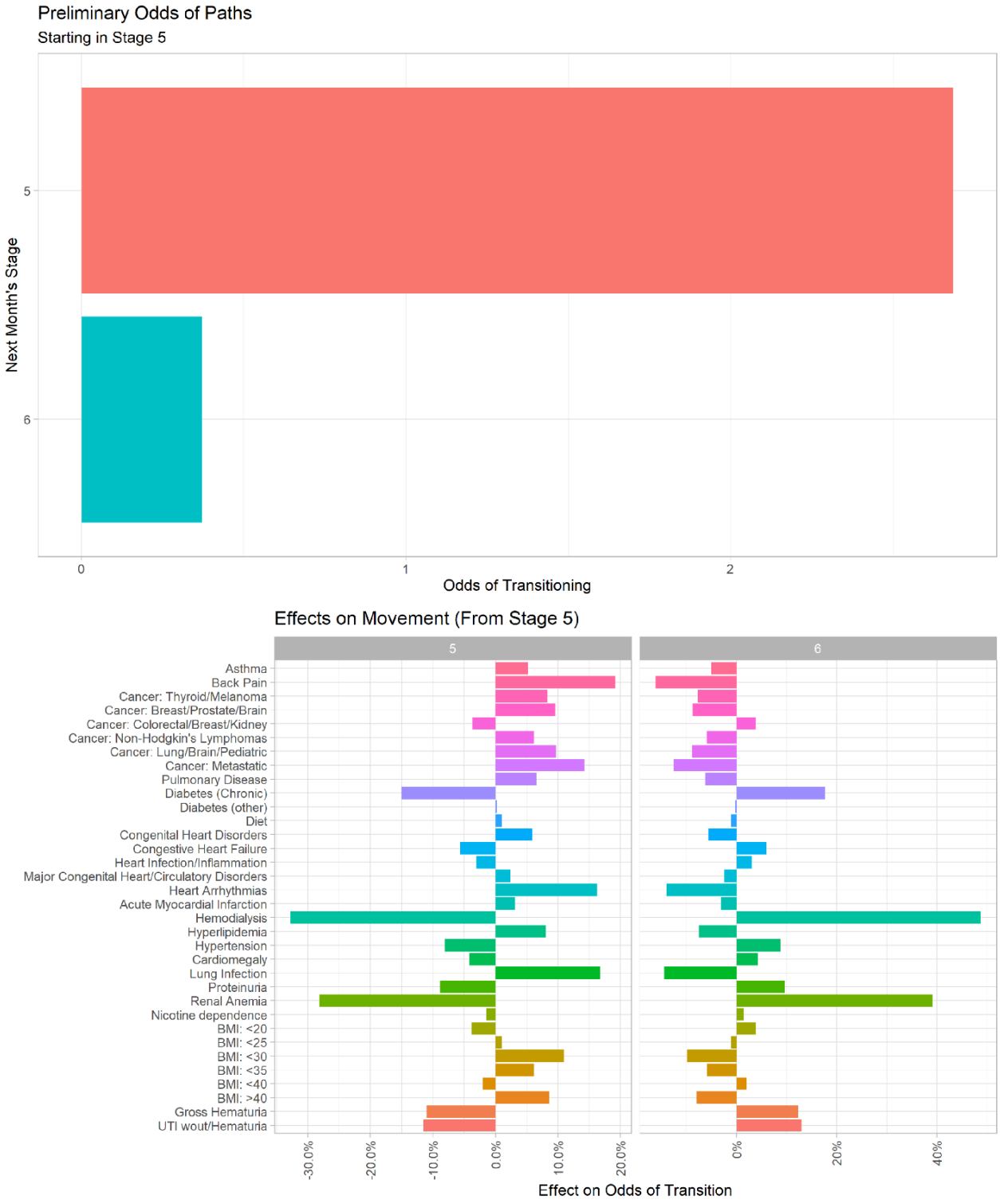


Figure C.5

COMMERCIAL STAGE TRANSITION MODEL STARTING IN STAGE 5



C.2 MEDICARE: MULTINOMIAL LOGISTIC MODEL

The following pages present results of using the Medicare stage transition model starting in stage 1 (Figure C.6), stage 2 (Figure C.7), stage 3 (Figure C.8), stage 4 (Figure C.9) and stage 5 (Figure C.10).

Figure C.6

MEDICARE STAGE TRANSITION MODEL STARTING IN STAGE 1

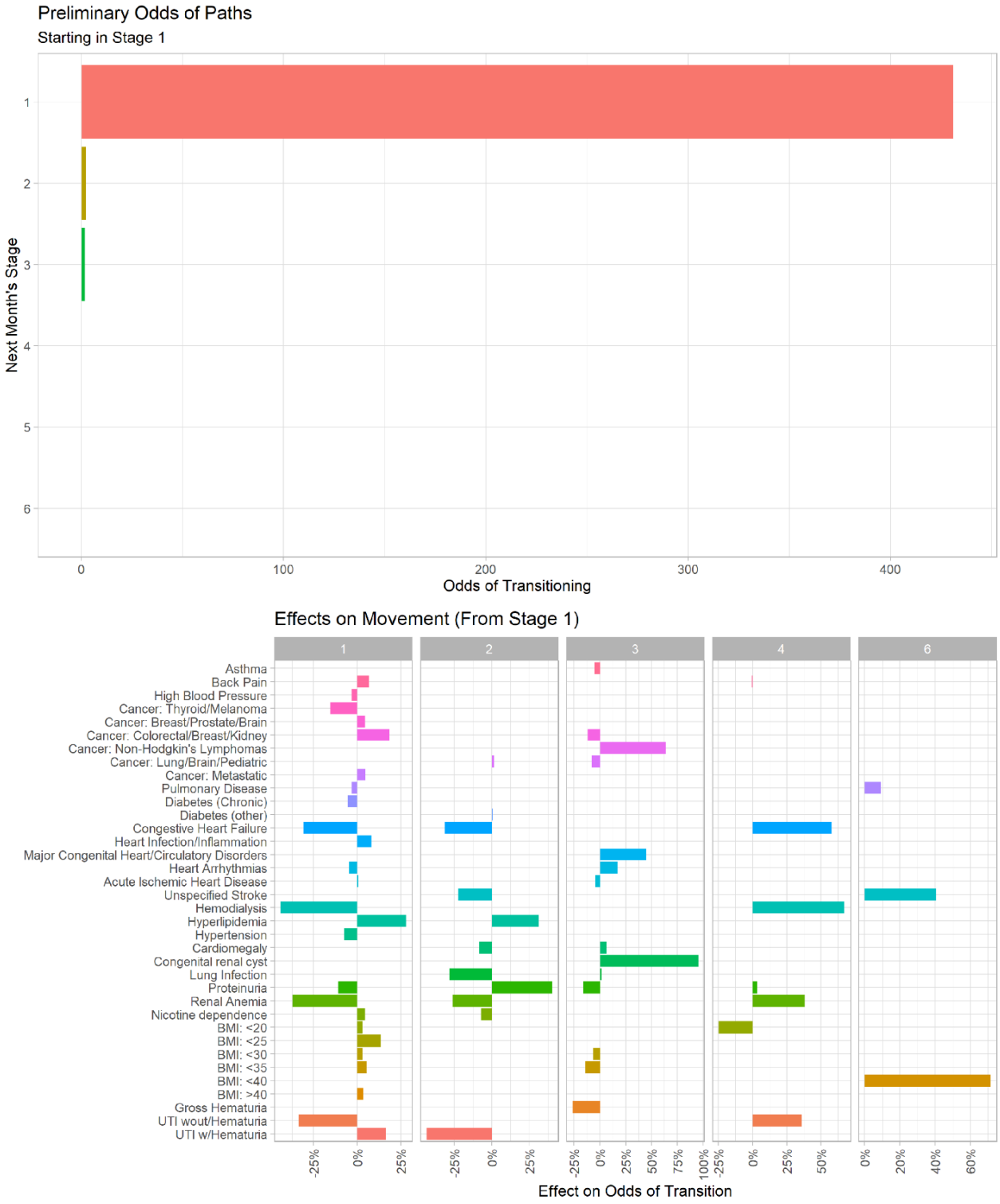


Figure C.7

MEDICARE STAGE TRANSITION MODEL STARTING IN STAGE 2

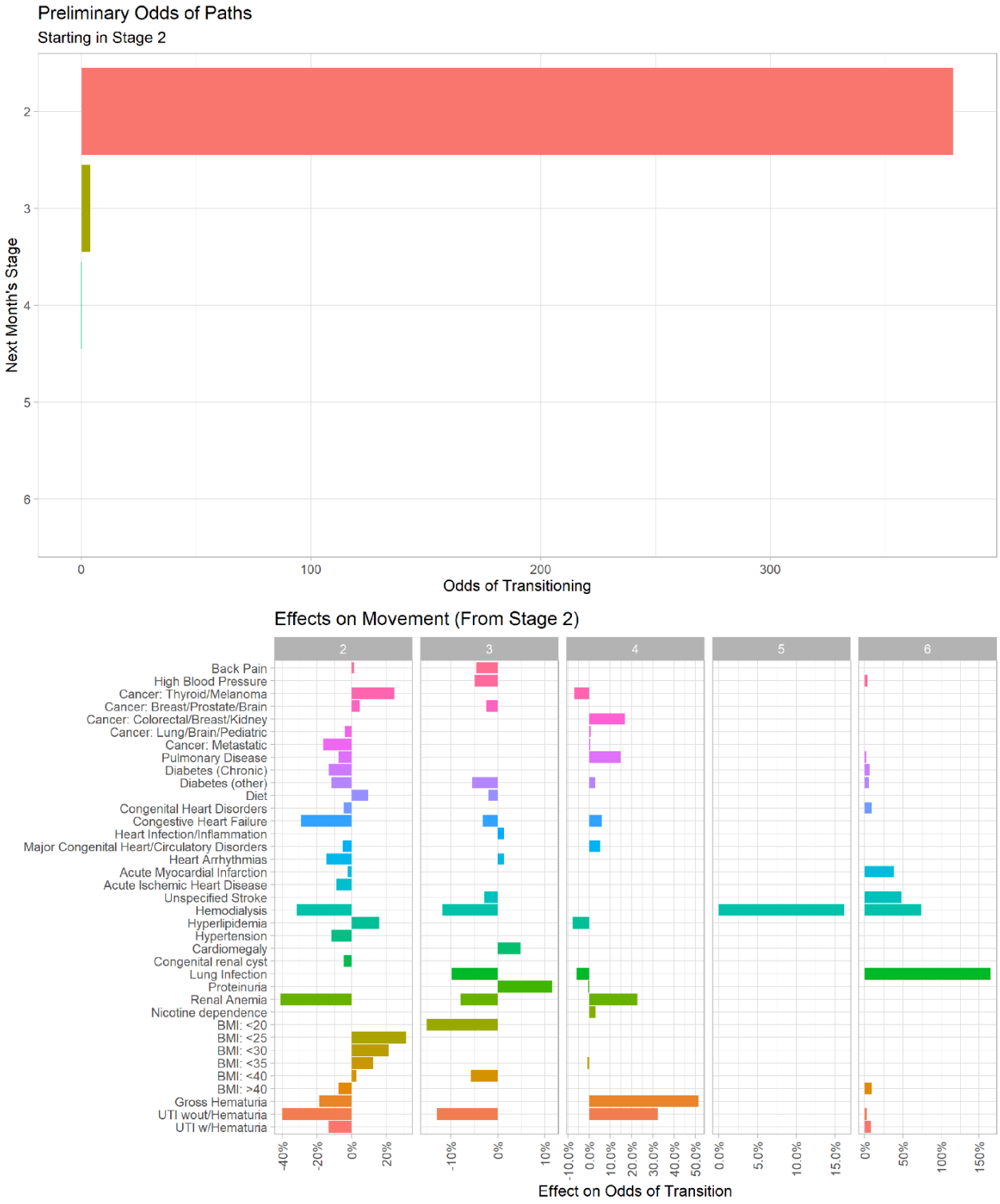


Figure C.8

MEDICARE STAGE TRANSITION MODEL STARTING IN STAGE 3

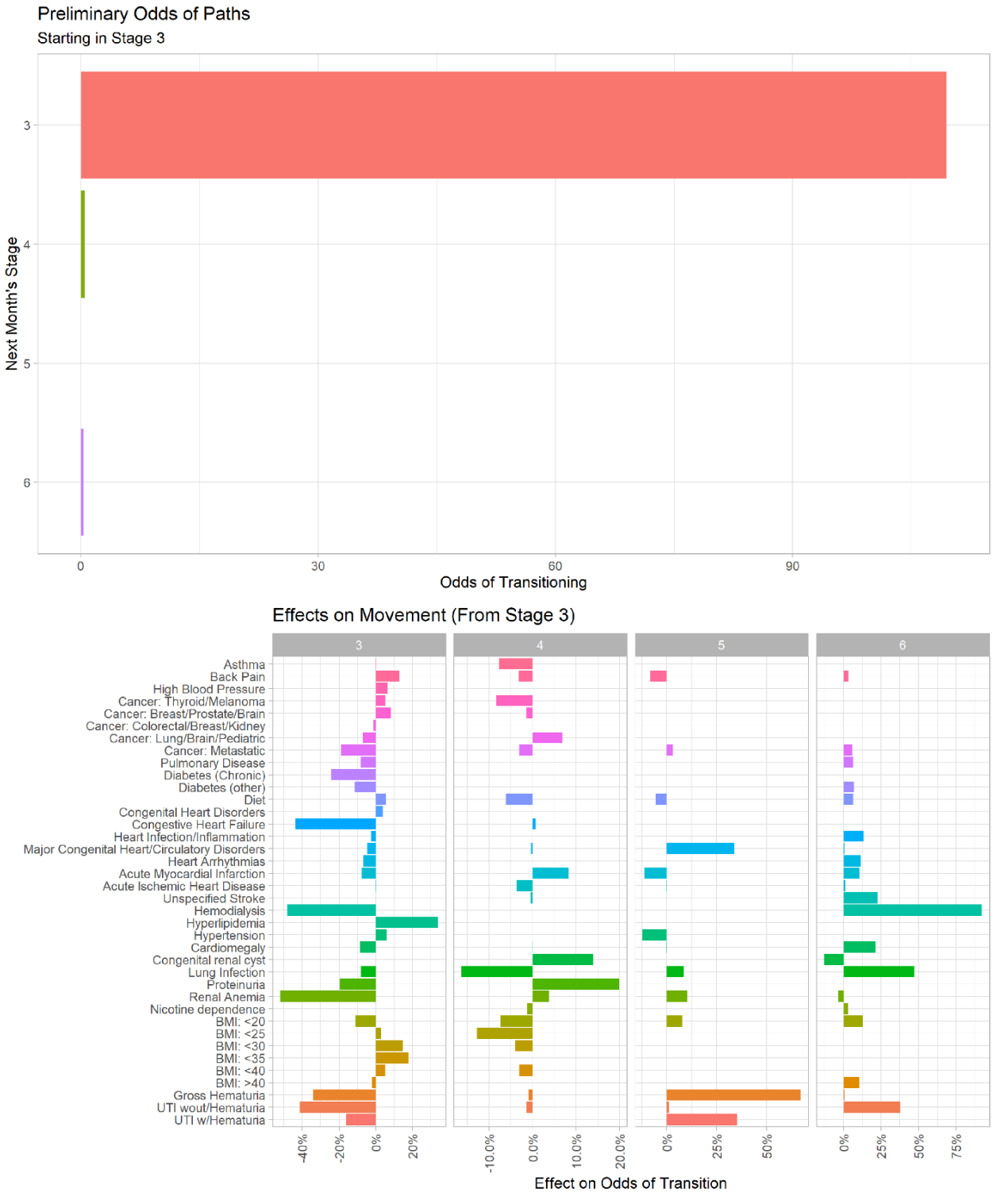


Figure C.9

MEDICARE STAGE TRANSITION MODEL STARTING IN STAGE 4

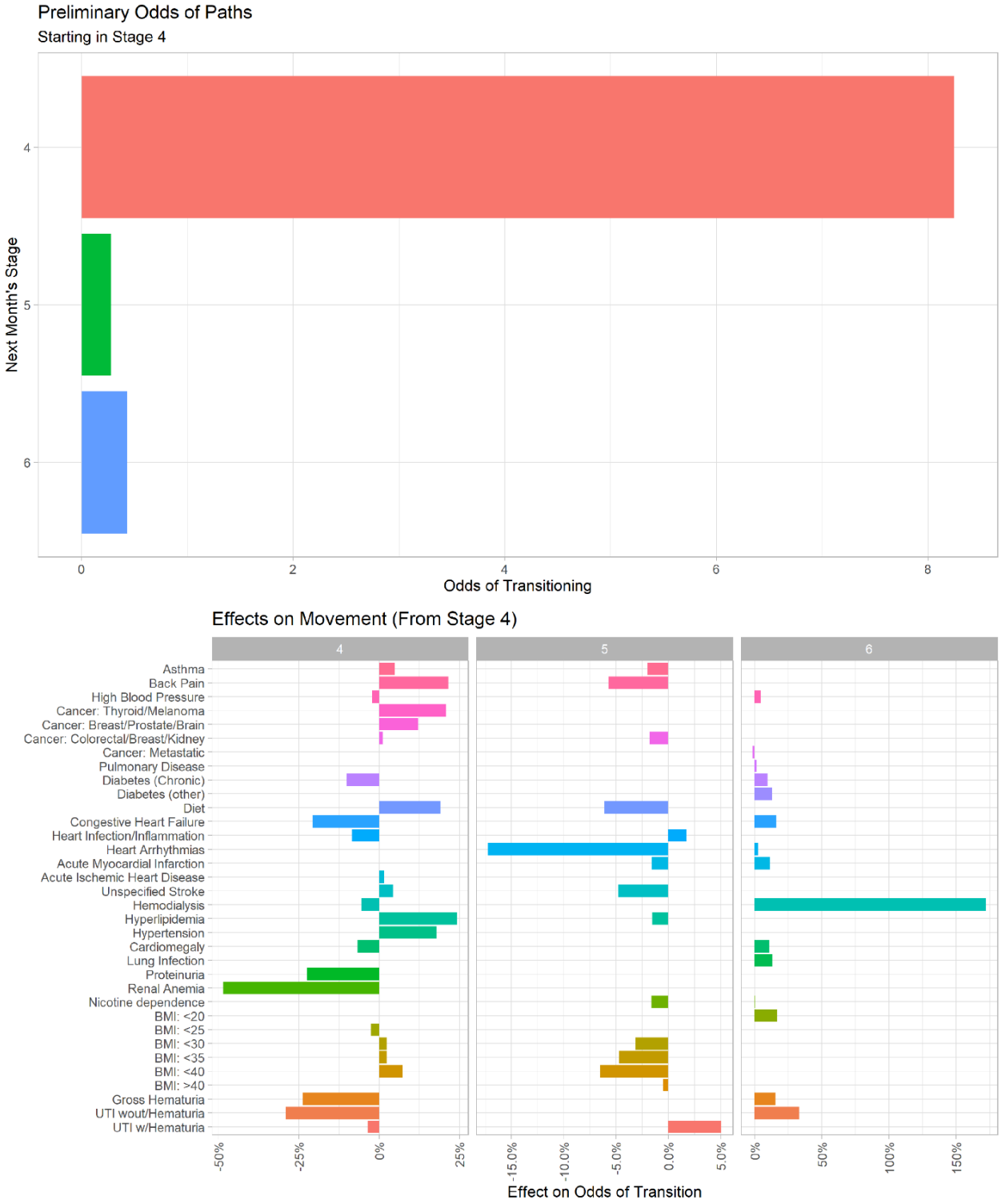
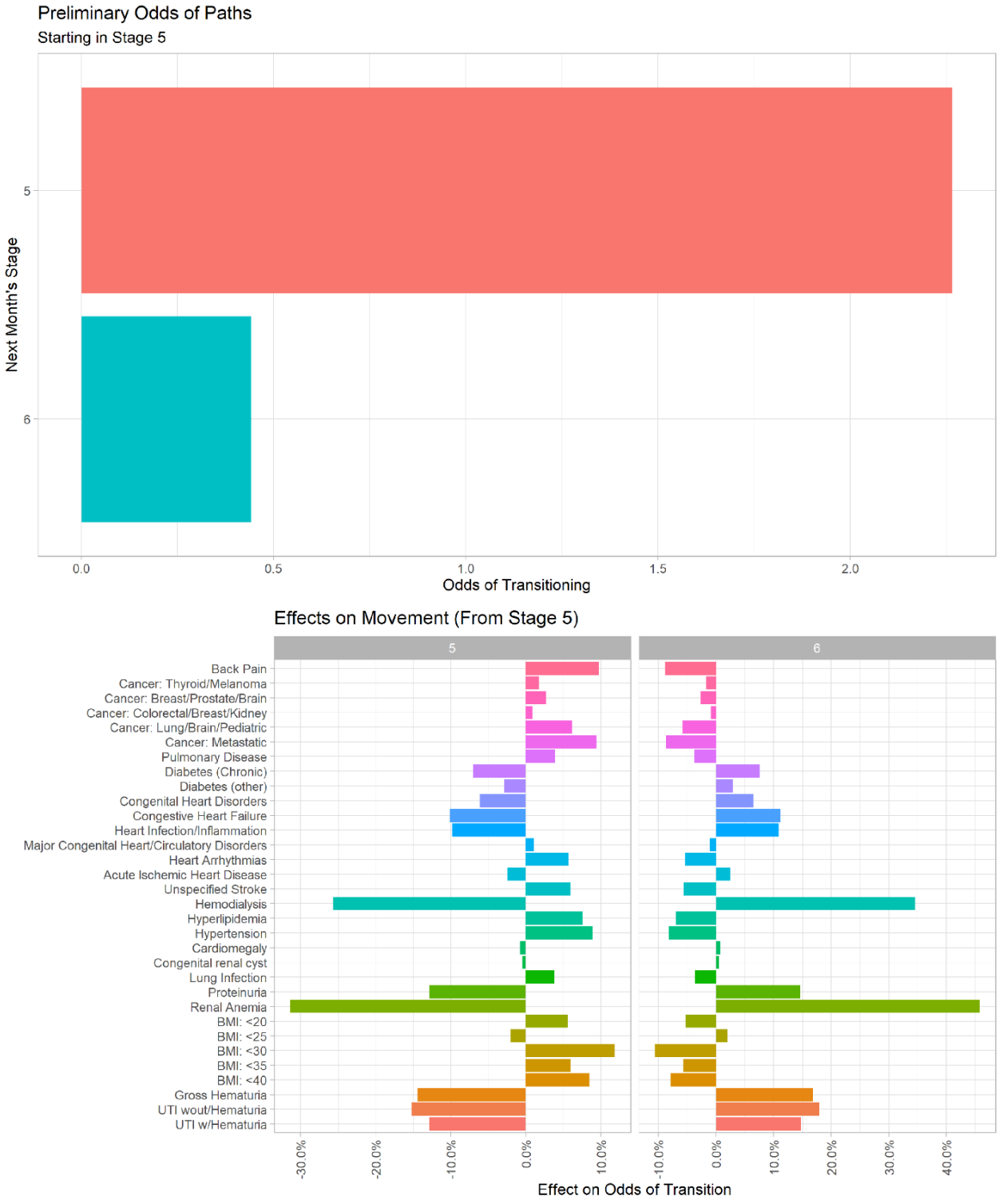


Figure C.10

MEDICARE STAGE TRANSITION MODEL STARTING ON STAGE 5-



Appendix D. Complete Cost Model Results

D.1 COMMERCIAL

D.1.1 LOGISTIC REGRESSION (ZERO/NONZERO)

The following pages present results of using the commercial logistic regression cost model for stage 1 (Figure D.1), stage 2 (Figure D.2), stage 3 (Figure D.3), stage 4 (Figure D.4), stage 5 (Figure D.5) and stage 6 (Figure D.6).

Figure D.1

COMMERCIAL LOGISTIC REGRESSION COST MODEL FOR STAGE 1

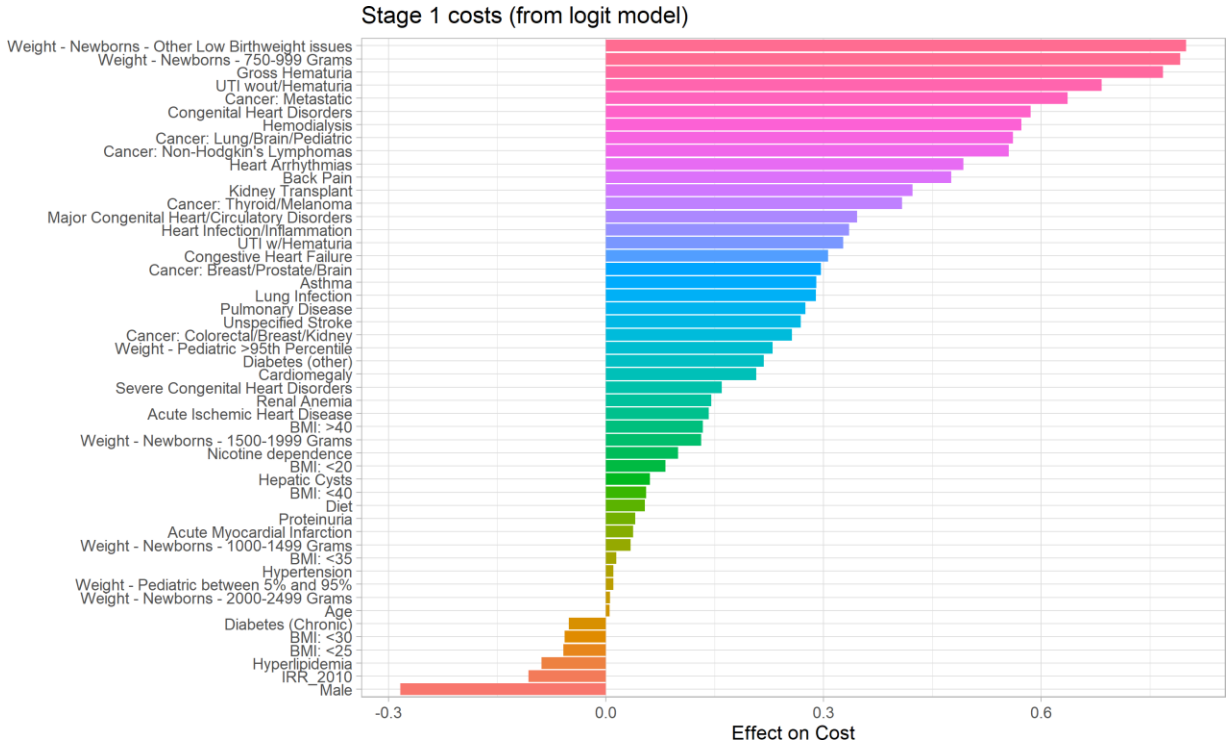


Figure D.2

COMMERCIAL LOGISTIC REGRESSION COST MODEL FOR STAGE 2

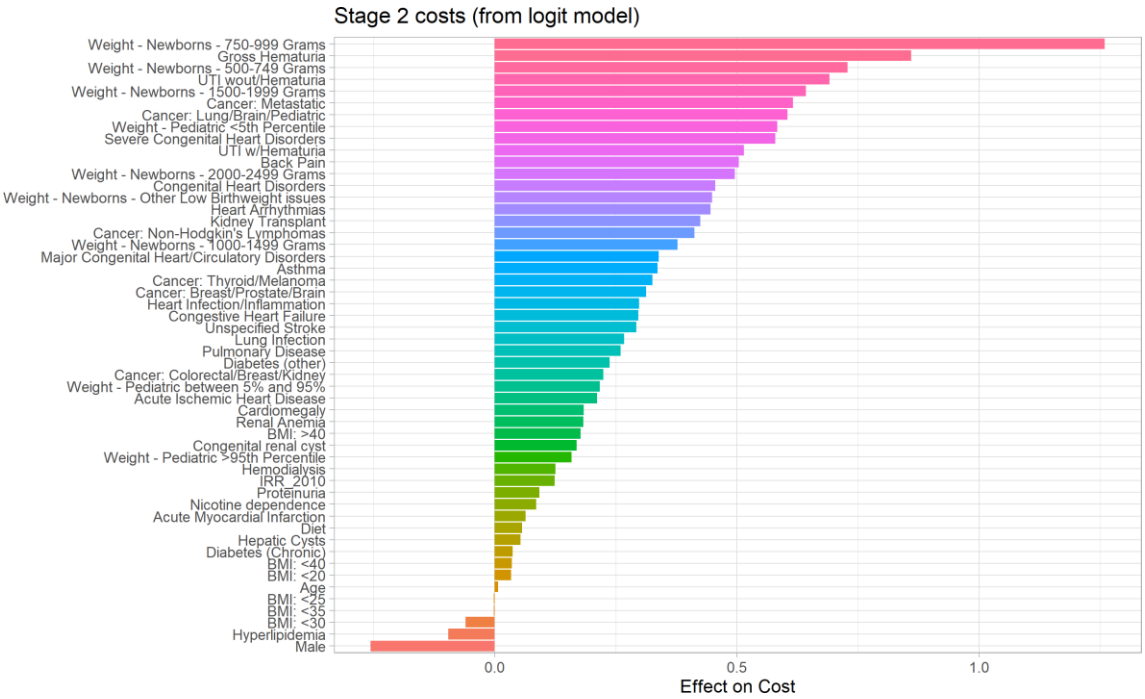


Figure D.3

COMMERCIAL LOGISTIC REGRESSION COST MODEL FOR STAGE 3

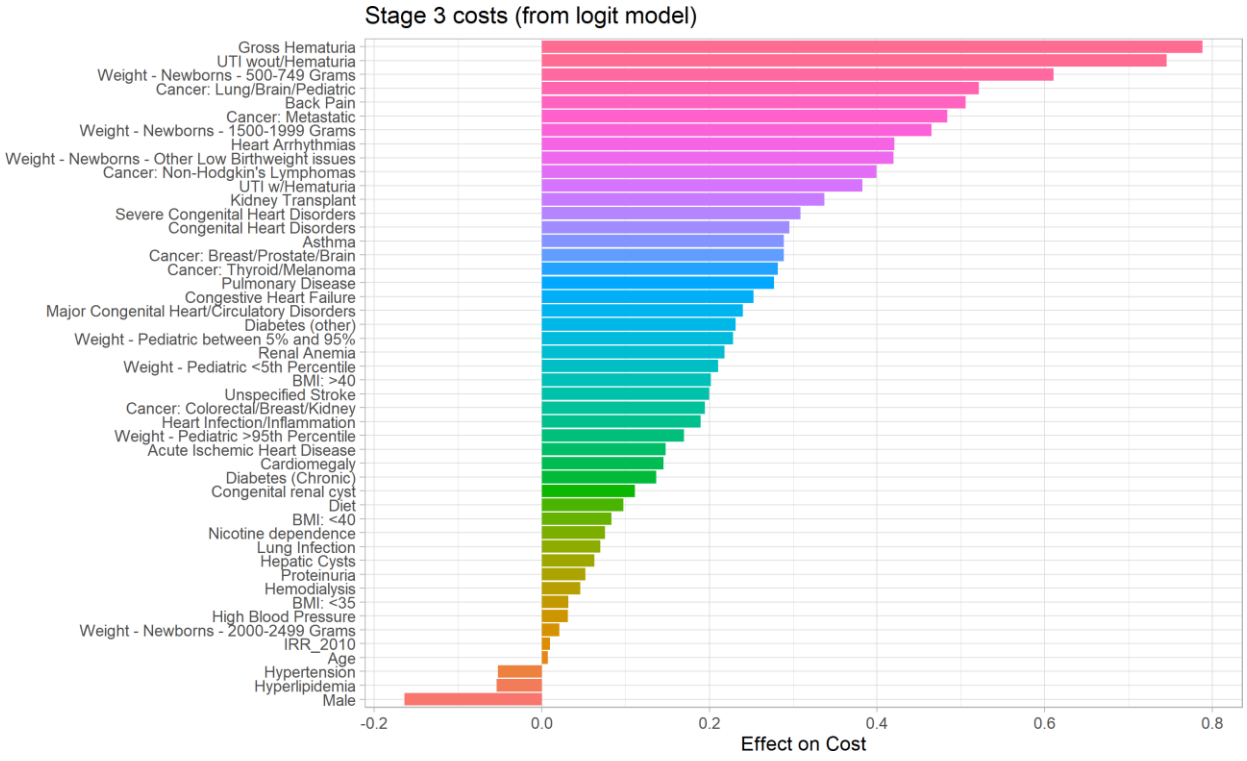


Figure D.4

COMMERCIAL LOGISTIC REGRESSION COST MODEL FOR STAGE 4

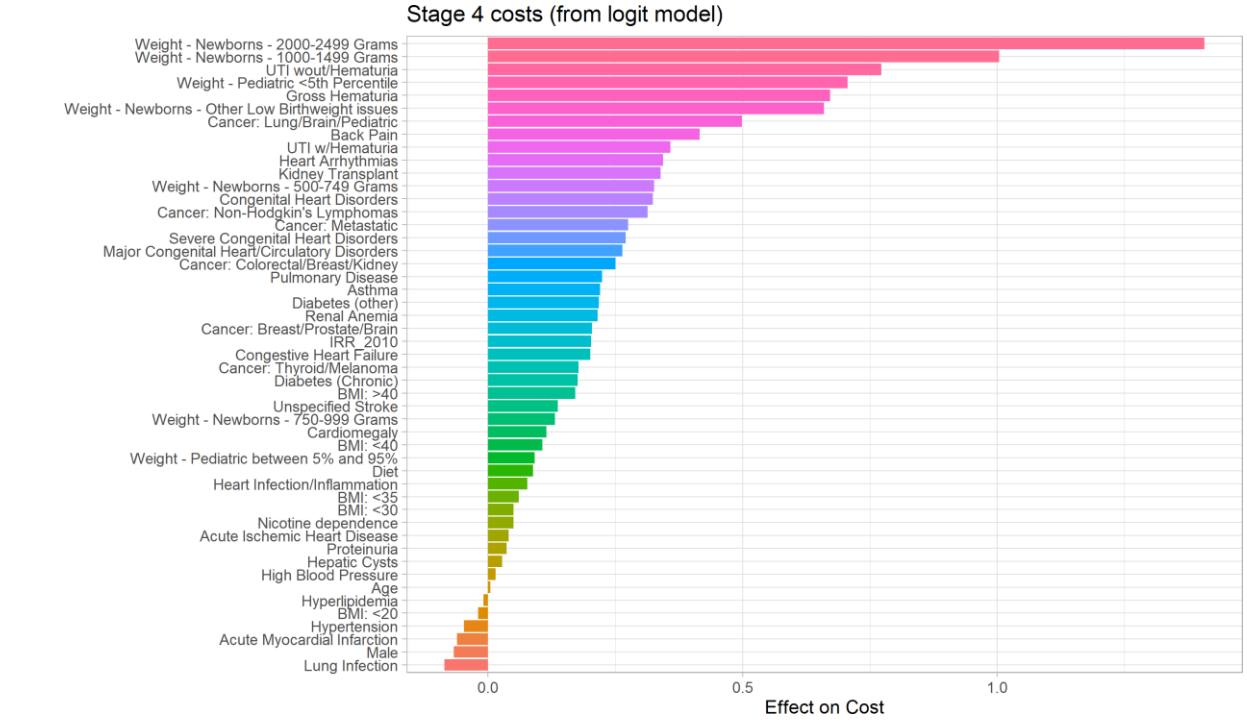


Figure D.5

COMMERCIAL LOGISTIC REGRESSION COST MODEL FOR STAGE 5

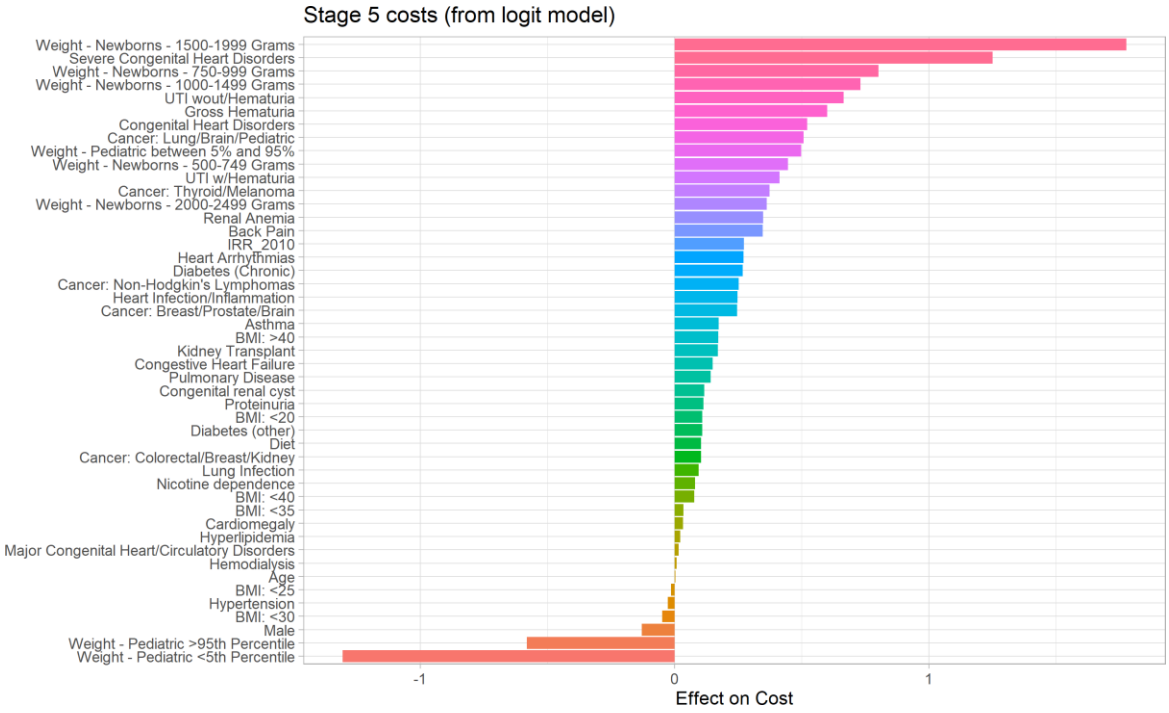
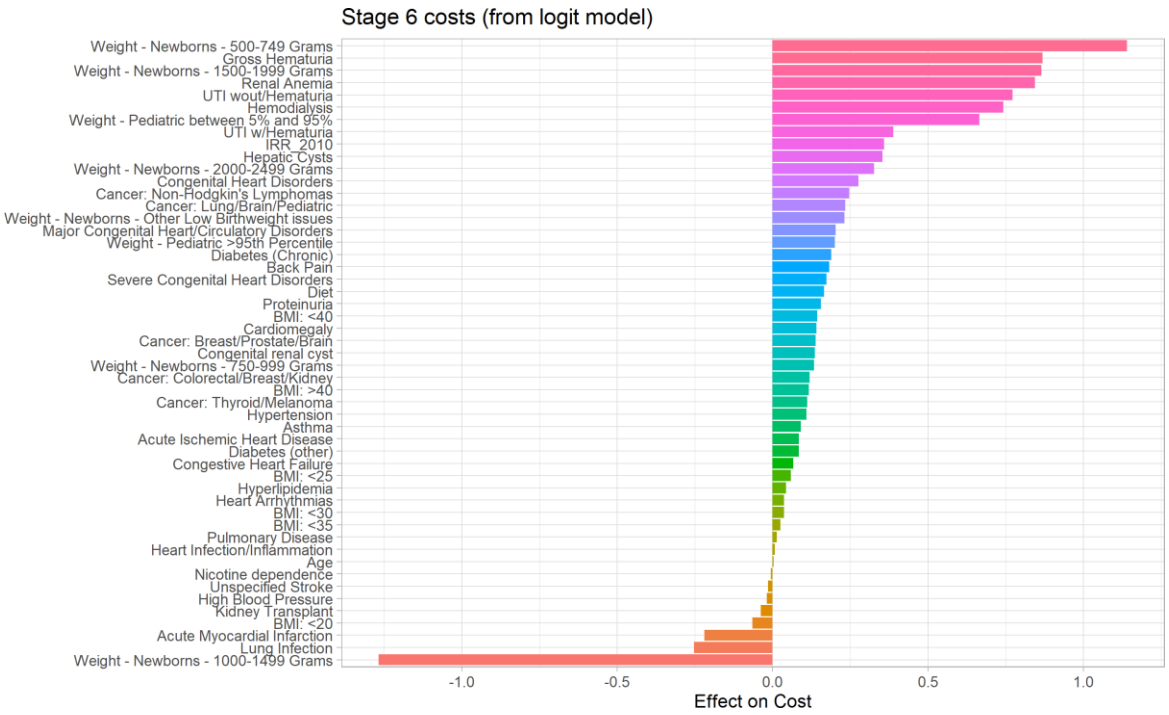


Figure D.6

COMMERCIAL LOGISTIC REGRESSION COST MODEL FOR STAGE 6



D.1.2 GAMMA REGRESSION

The following pages present results of using the commercial gamma regression cost model for stage 1 (Figure D.7), stage 2 (Figure D.8), stage 3 (Figure D.9), stage 4 (Figure D.10), stage 5 (Figure D.11) and stage 6 (Figure D.12).

Figure D.7

COMMERCIAL GAMMA REGRESSION COST MODEL FOR STAGE 1

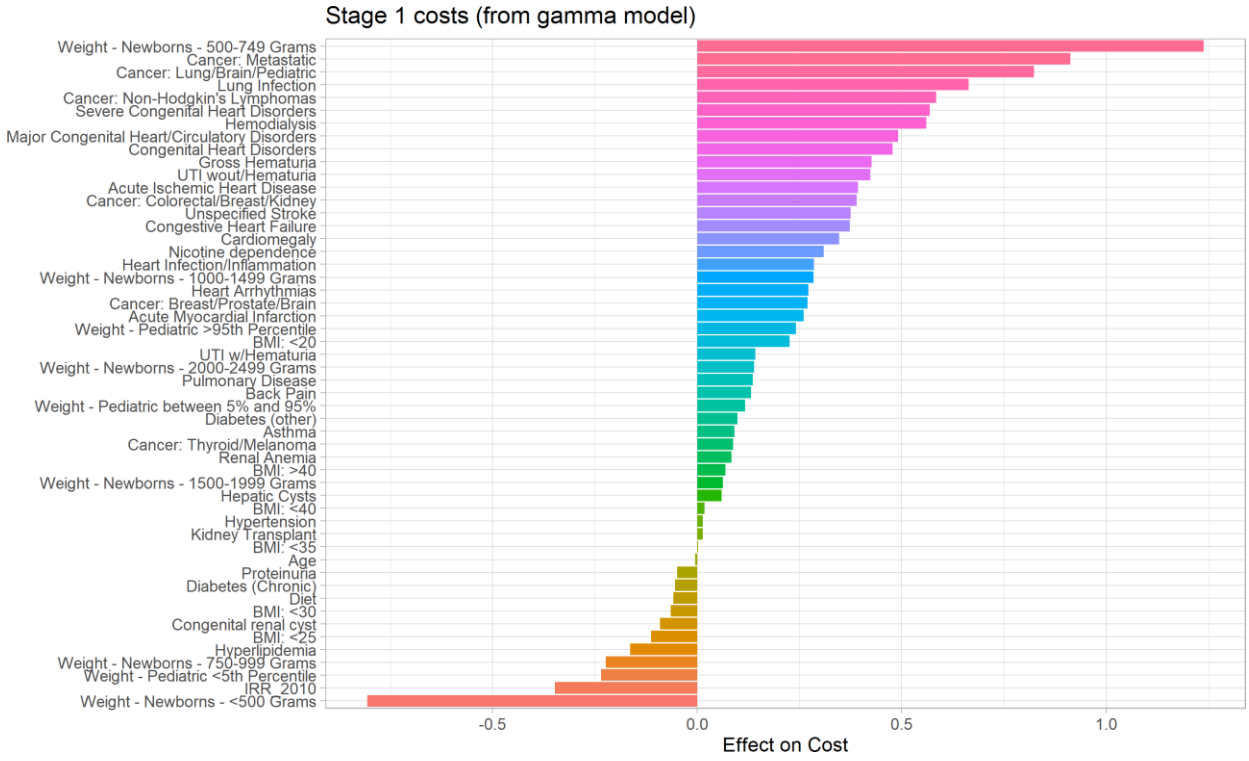


Figure D.8

COMMERCIAL GAMMA REGRESSION COST MODEL FOR STAGE 2

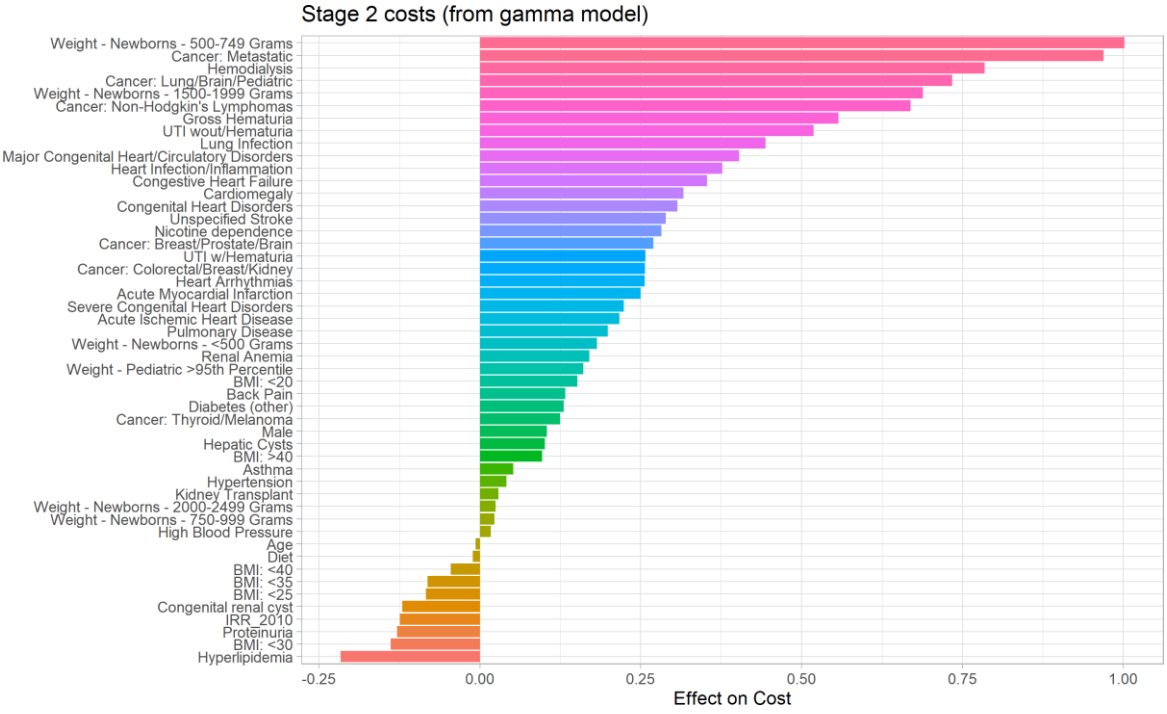


Figure D.9
COMMERCIAL GAMMA REGRESSION COST MODEL FOR STAGE 3

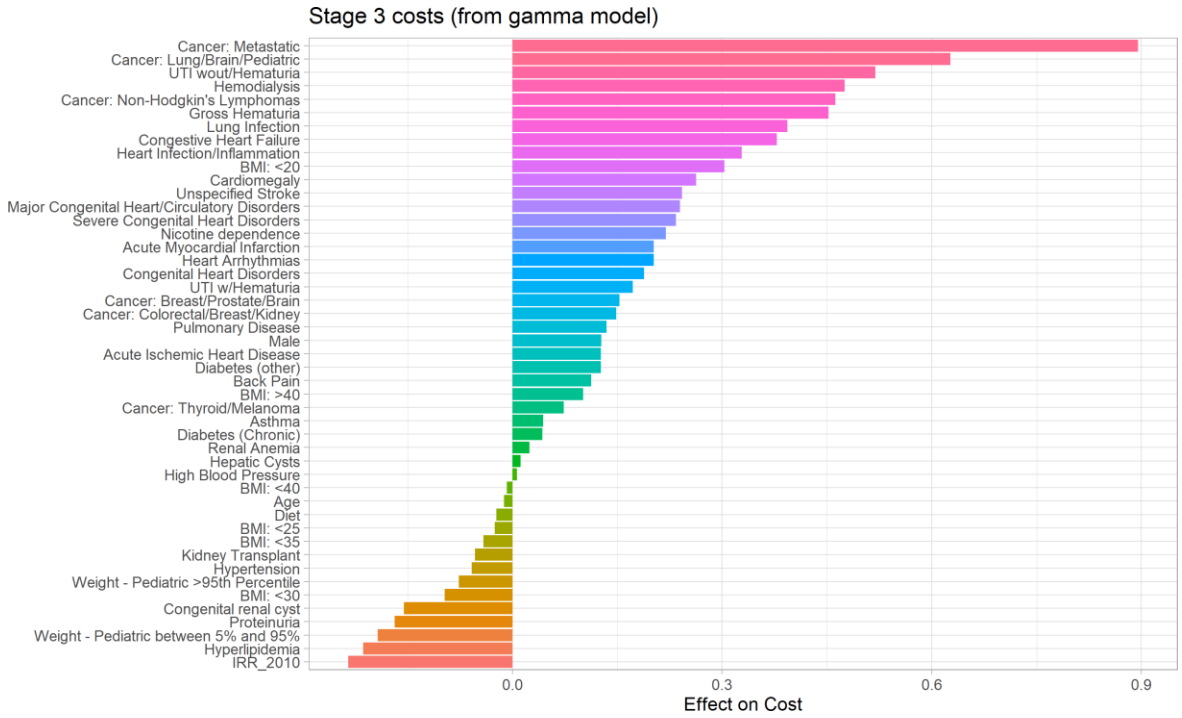


Figure D.10
COMMERCIAL GAMMA REGRESSION COST MODEL FOR STAGE 4

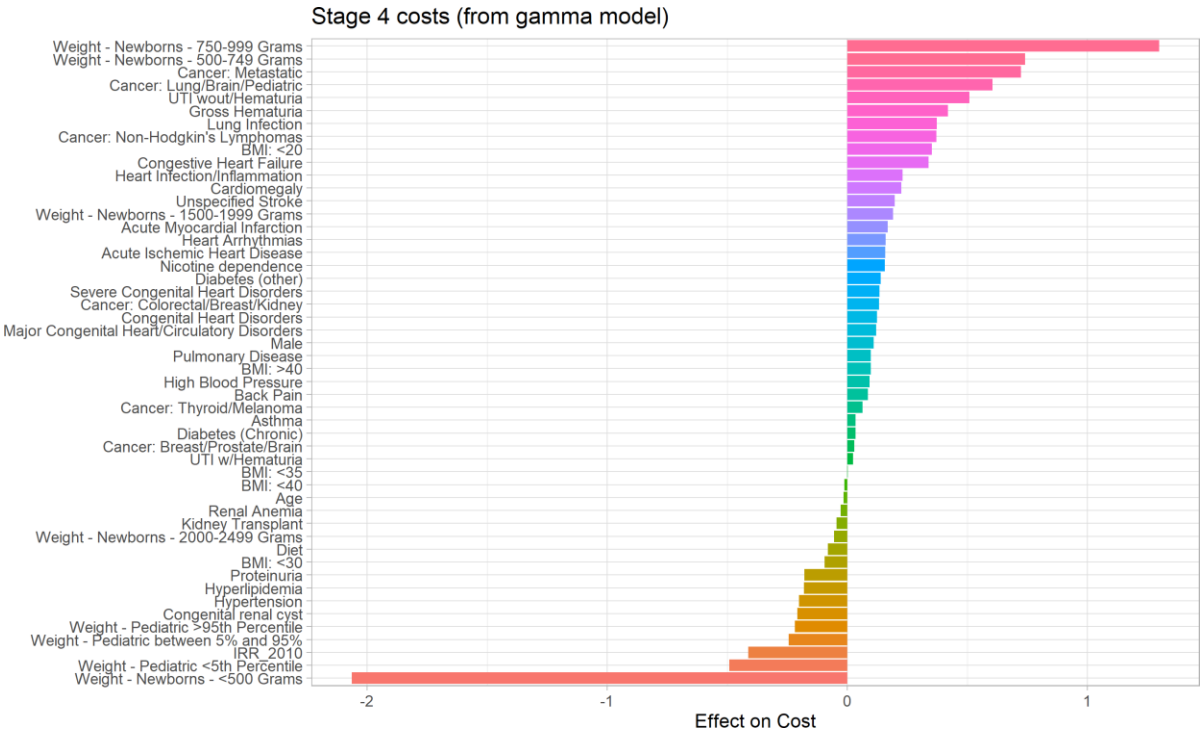


Figure D.11

COMMERCIAL GAMMA REGRESSION COST MODEL FOR STAGE 5

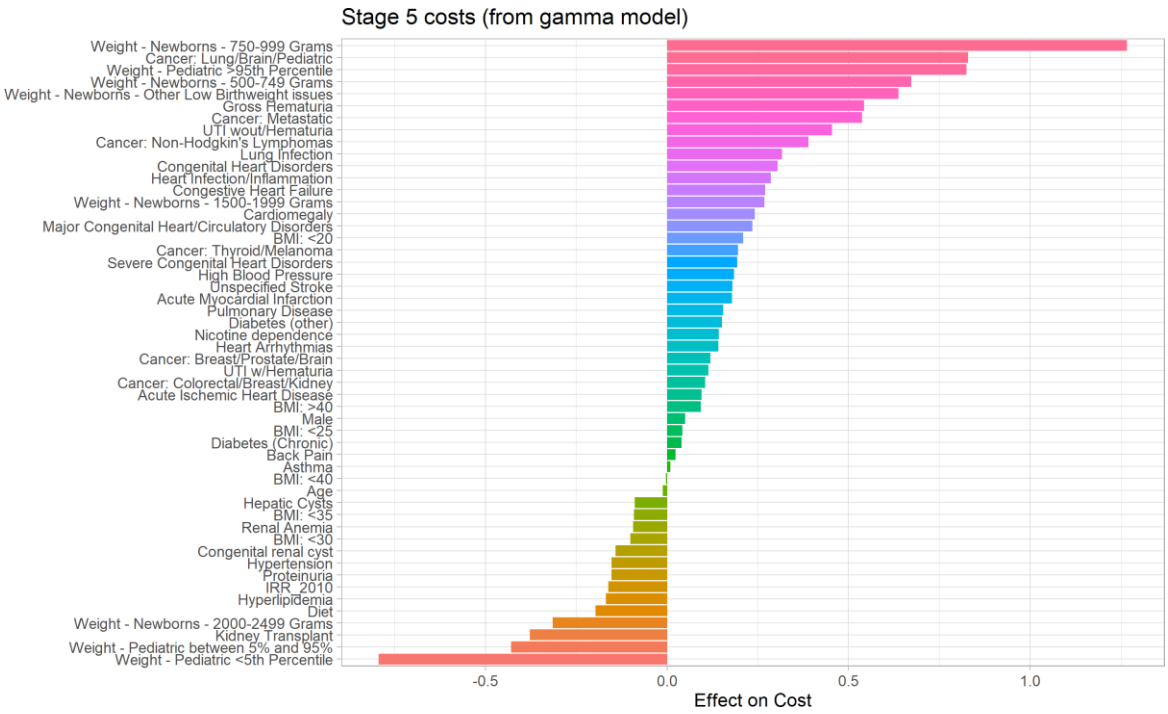
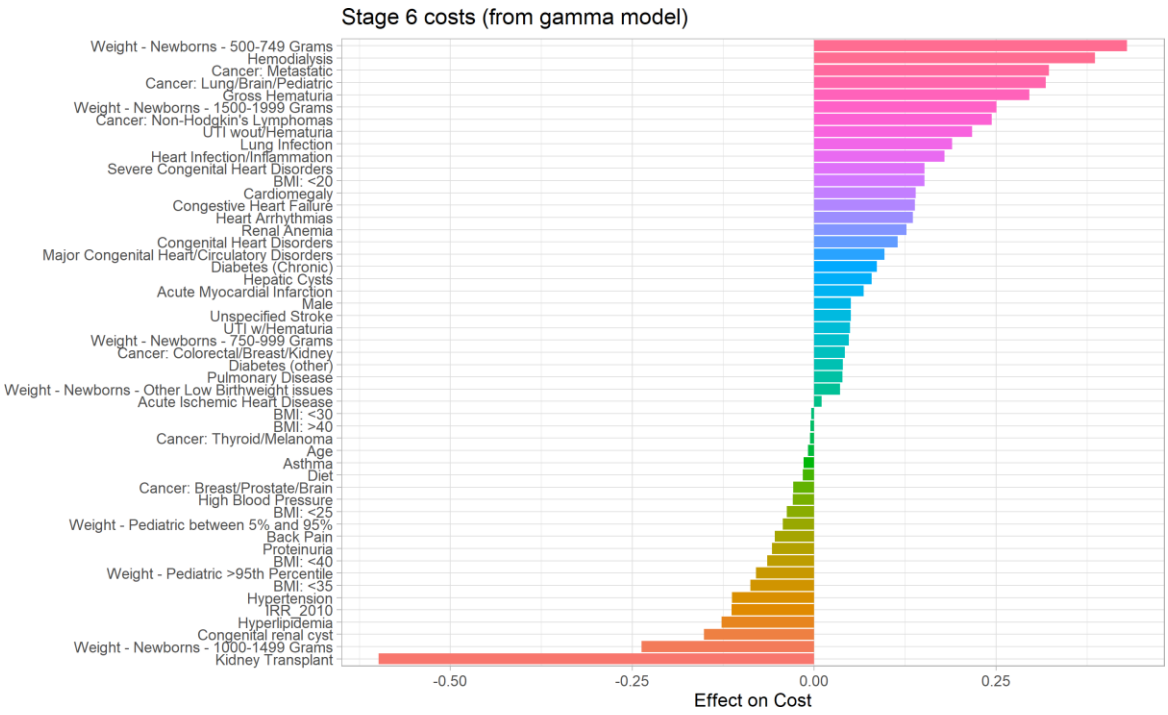


Figure D.12

COMMERCIAL GAMMA REGRESSION COST MODEL FOR STAGE 6



D.1.3 TWEEDIE REGRESSION

The following pages present results of using the commercial Tweedie regression cost model for stage 1 (Figure D.13), stage 2 (Figure D.14), stage 3 (Figure D.15), stage 4 (Figure D.16), stage 5 (Figure D.17) and stage 6 (Figure D.18).

Figure D.13

COMMERCIAL TWEEDIE REGRESSION COST MODEL FOR STAGE 1

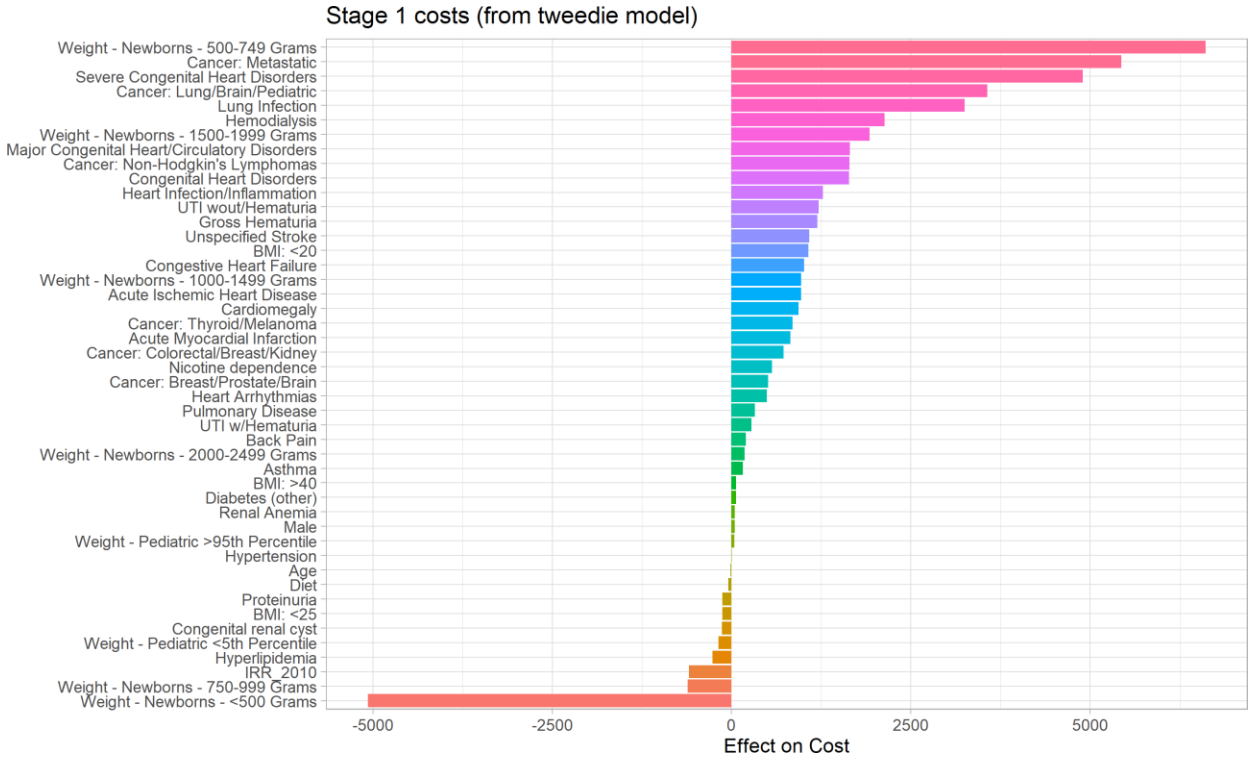


Figure D.14

COMMERCIAL TWEEDIE REGRESSION COST MODEL FOR STAGE 2

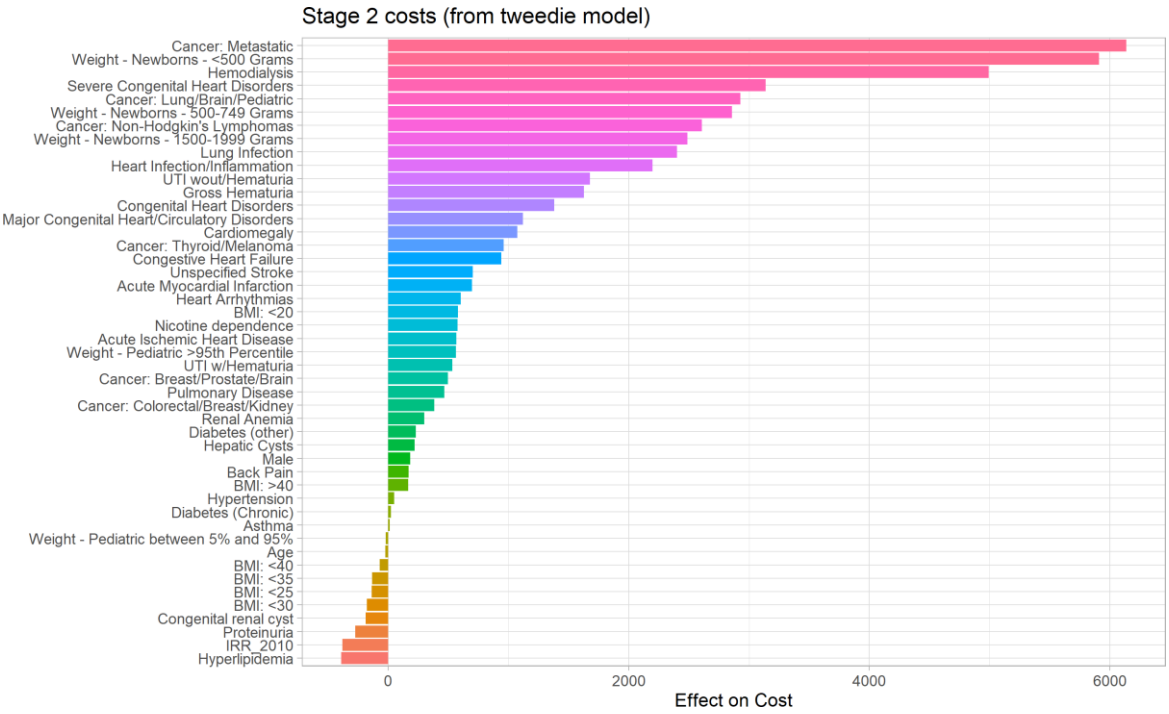


Figure D.15

COMMERCIAL TWEEDIE REGRESSION COST MODEL FOR STAGE 3

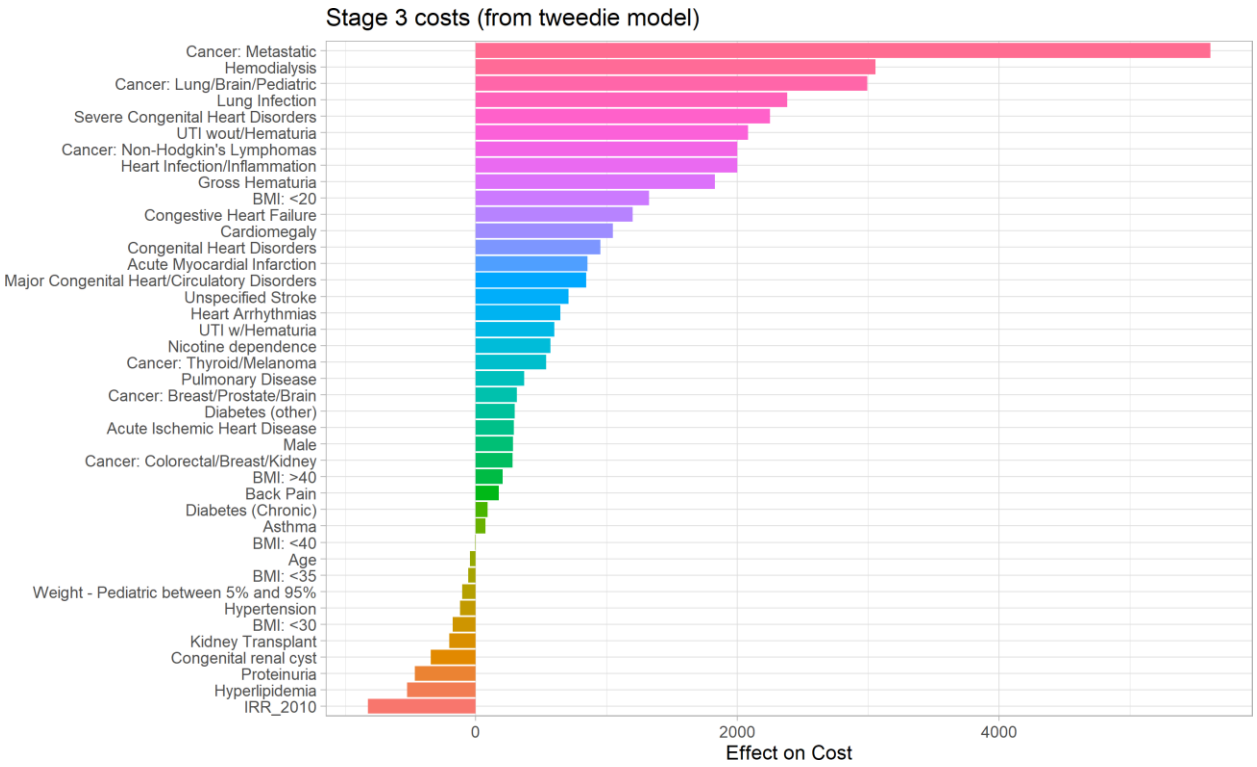


Figure D.16

COMMERCIAL TWEEDIE REGRESSION COST MODEL FOR STAGE 4

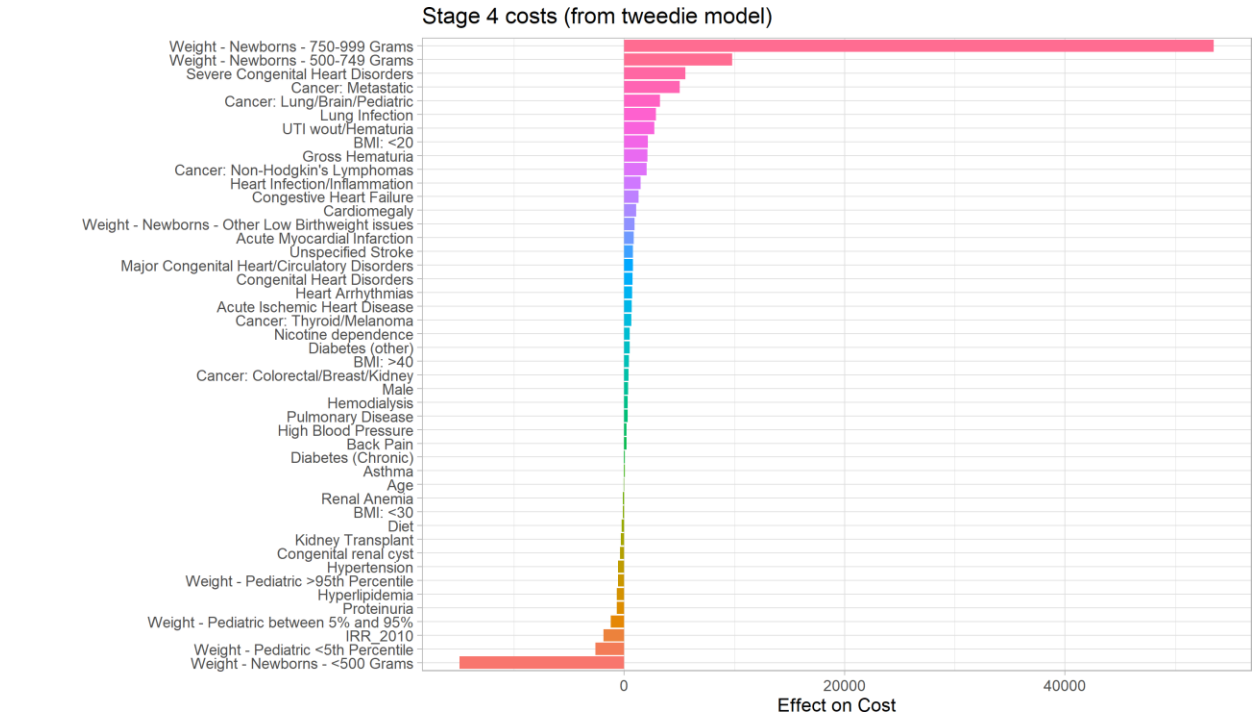


Figure D.17

COMMERCIAL TWEEDIE REGRESSION COST MODEL FOR STAGE 5

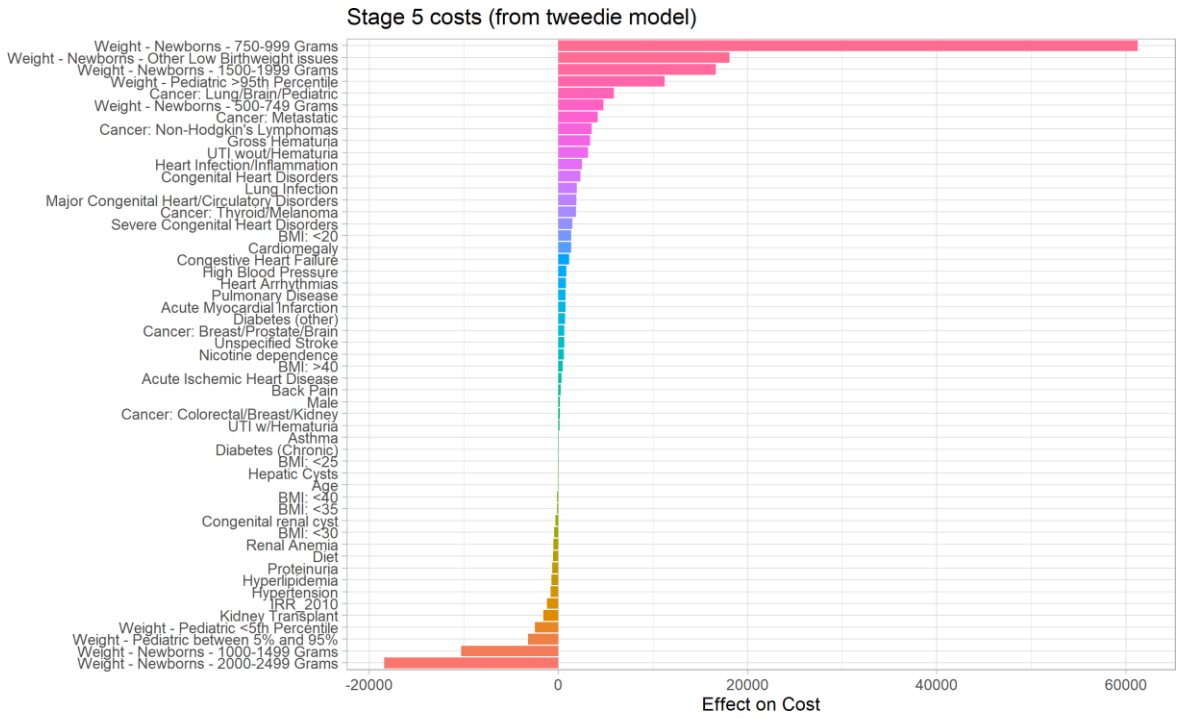
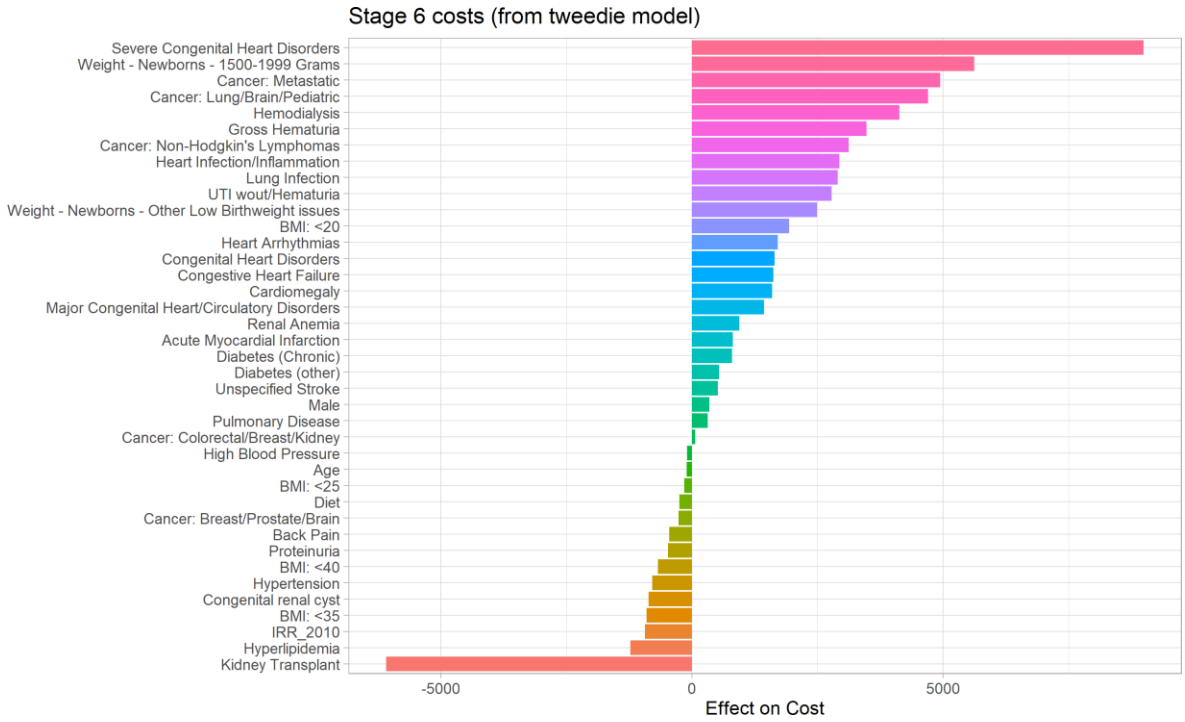


Figure D.18

COMMERCIAL TWEEDIE REGRESSION COST MODEL FOR STAGE 6



D.2 MEDICARE

D.2.1 LOGISTIC REGRESSION (ZERO/NONZERO)

The following pages present results of using the Medicare logistic regression cost model for stage 1 (Figure D.19), stage 2 (Figure D.20), stage 3 (Figure D.21), stage 4 (Figure D.22), stage 5 (Figure D.23) and stage 6 (Figure D.24).

Figure D.19

MEDICARE LOGISTIC REGRESSION COST MODEL FOR STAGE 1

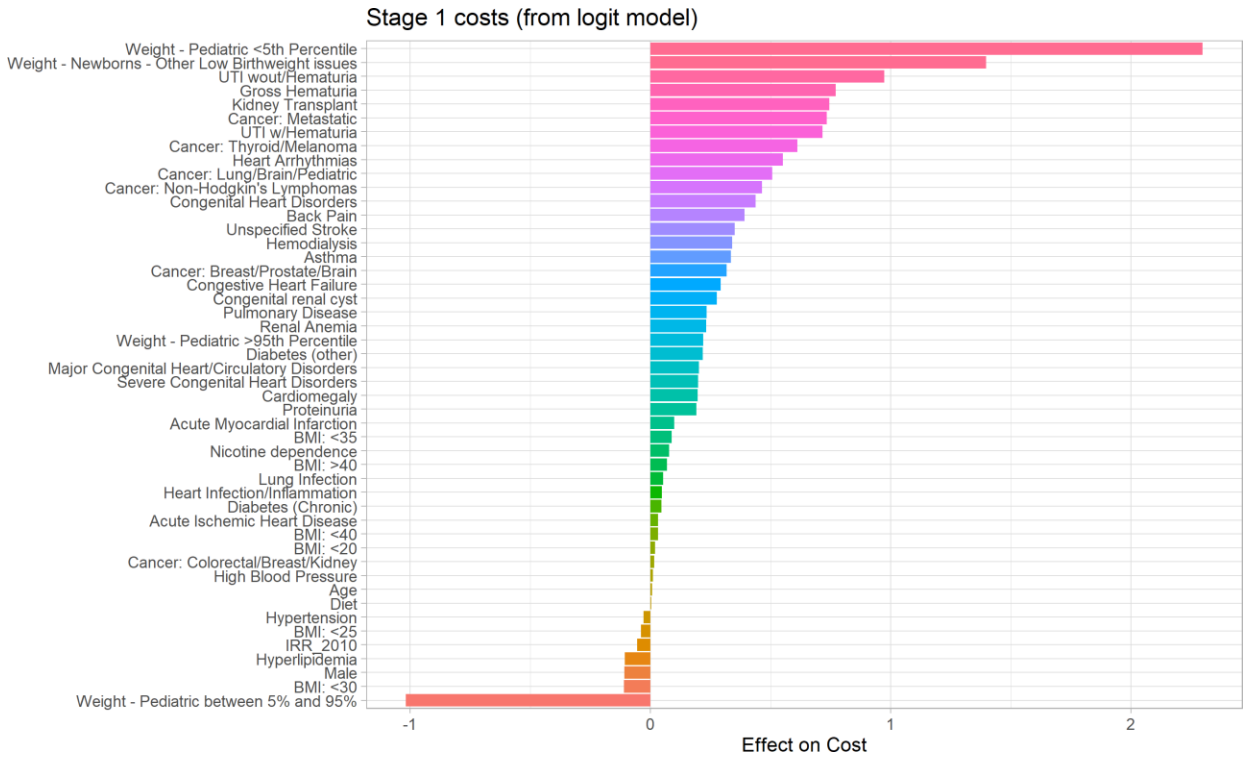


Figure D.20

MEDICARE LOGISTIC REGRESSION COST MODEL FOR STAGE 2

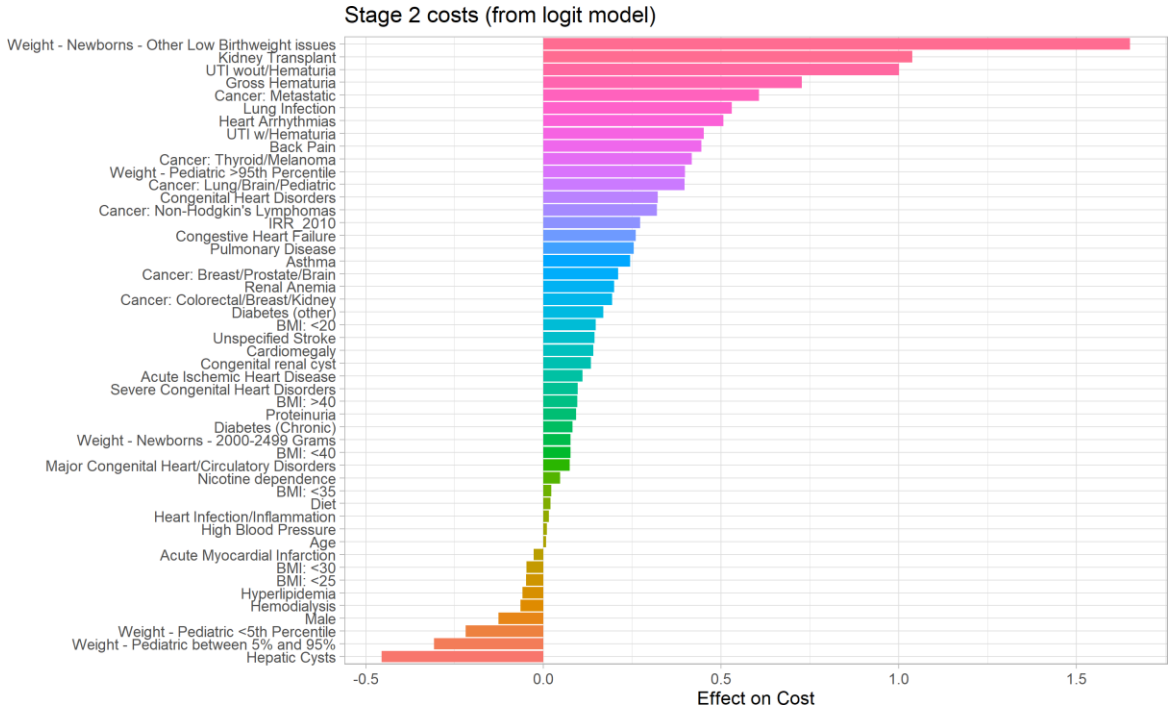


Figure D.21
MEDICARE LOGISTIC REGRESSION COST MODEL FOR STAGE 3

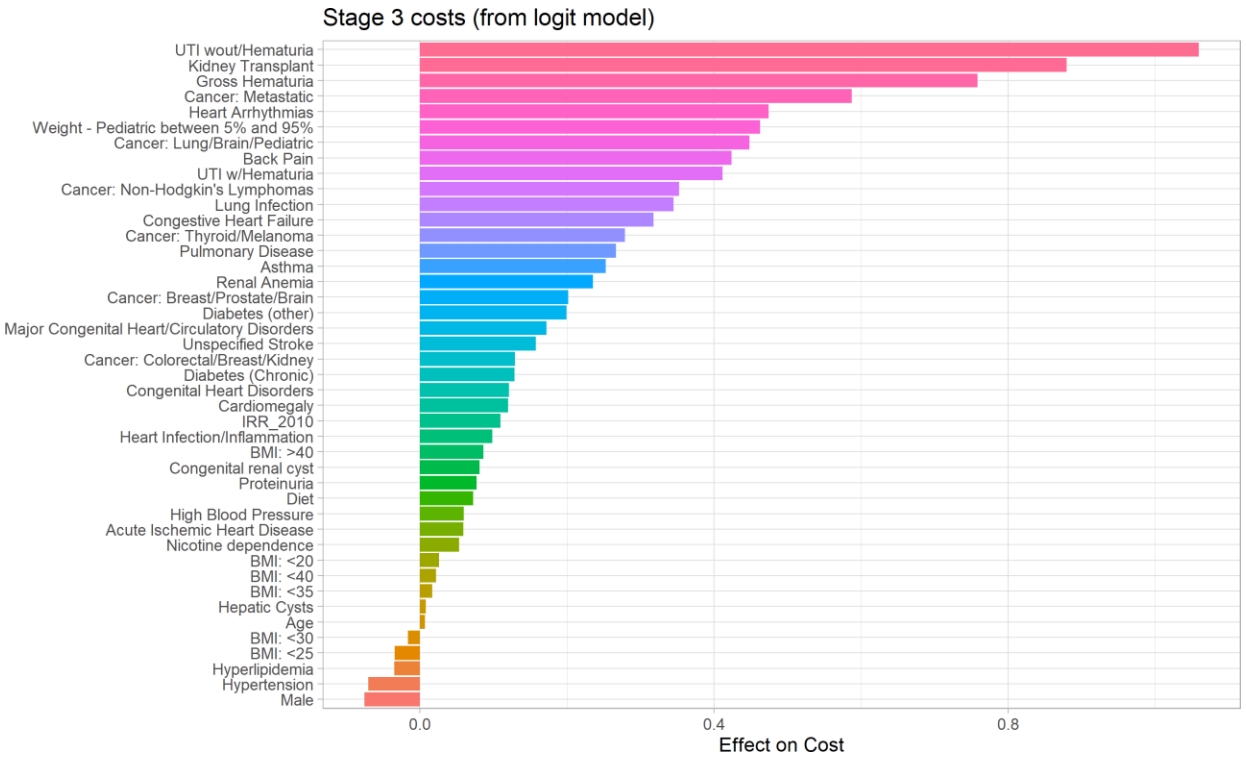


Figure D.22
MEDICARE LOGISTIC REGRESSION COST MODEL FOR STAGE 4

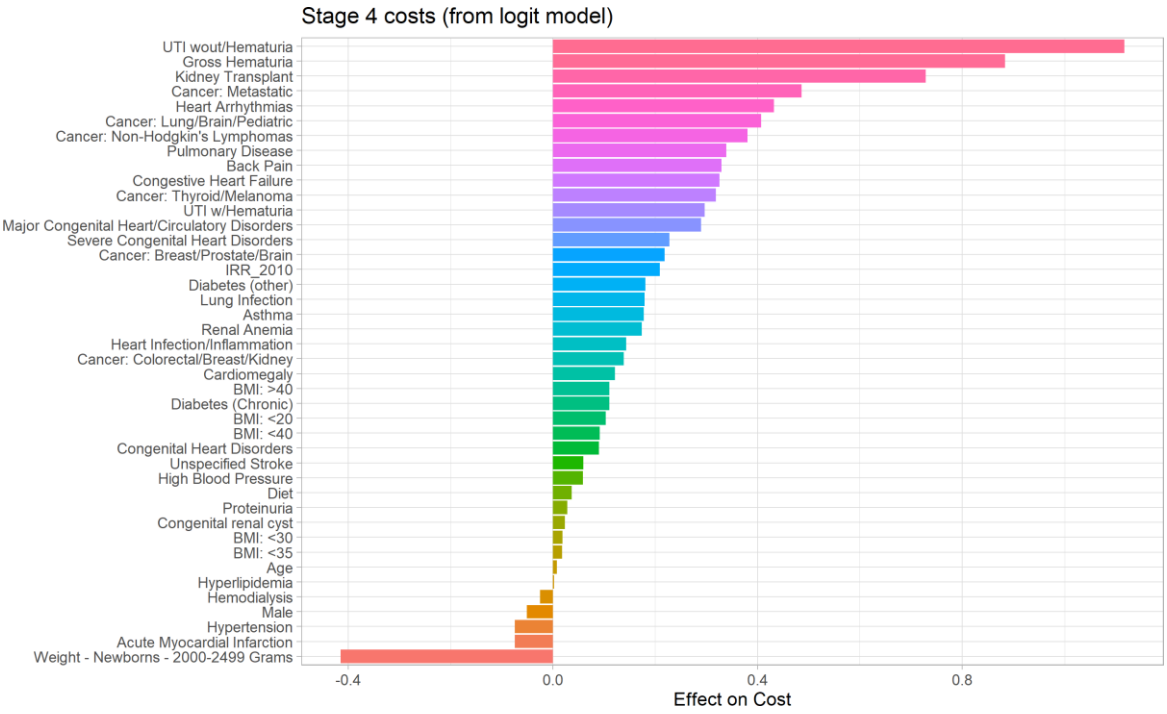


Figure D.23
MEDICARE LOGISTIC REGRESSION COST MODEL FOR STAGE 5

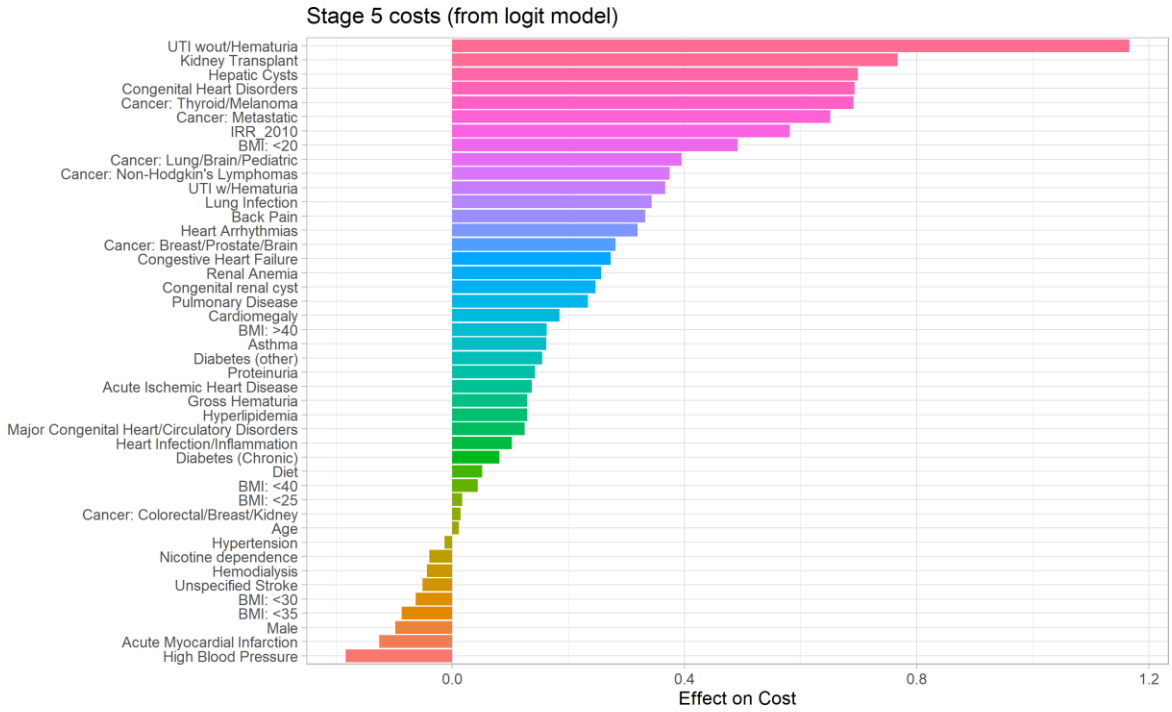
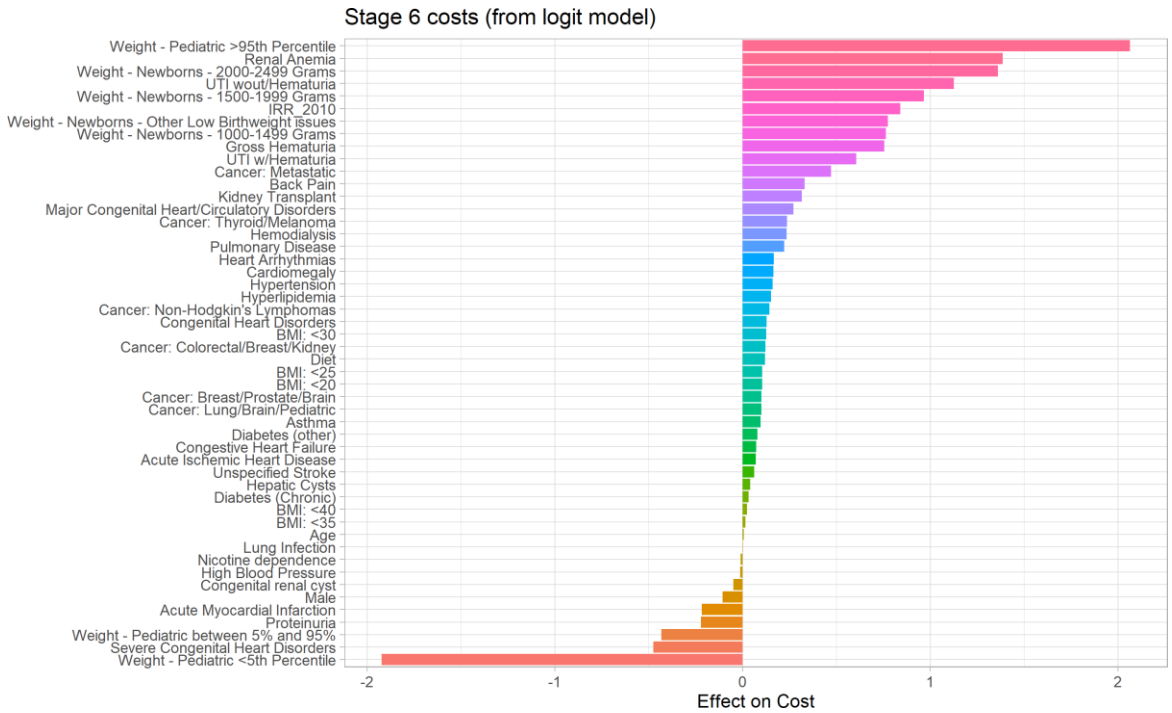


Figure D.24
MEDICARE LOGISTIC REGRESSION COST MODEL FOR STAGE 6



D.2.2 GAMMA REGRESSION

The following pages present results of using the Medicare gamma regression cost model for stage 1 (Figure D.25), stage 2 (Figure D.26), stage 3 (Figure D.27), stage 4 (Figure D.28), stage 5 (Figure D.29) and stage 6 (Figure D.30).

Figure D.25

MEDICARE GAMMA REGRESSION COST MODEL FOR STAGE 1

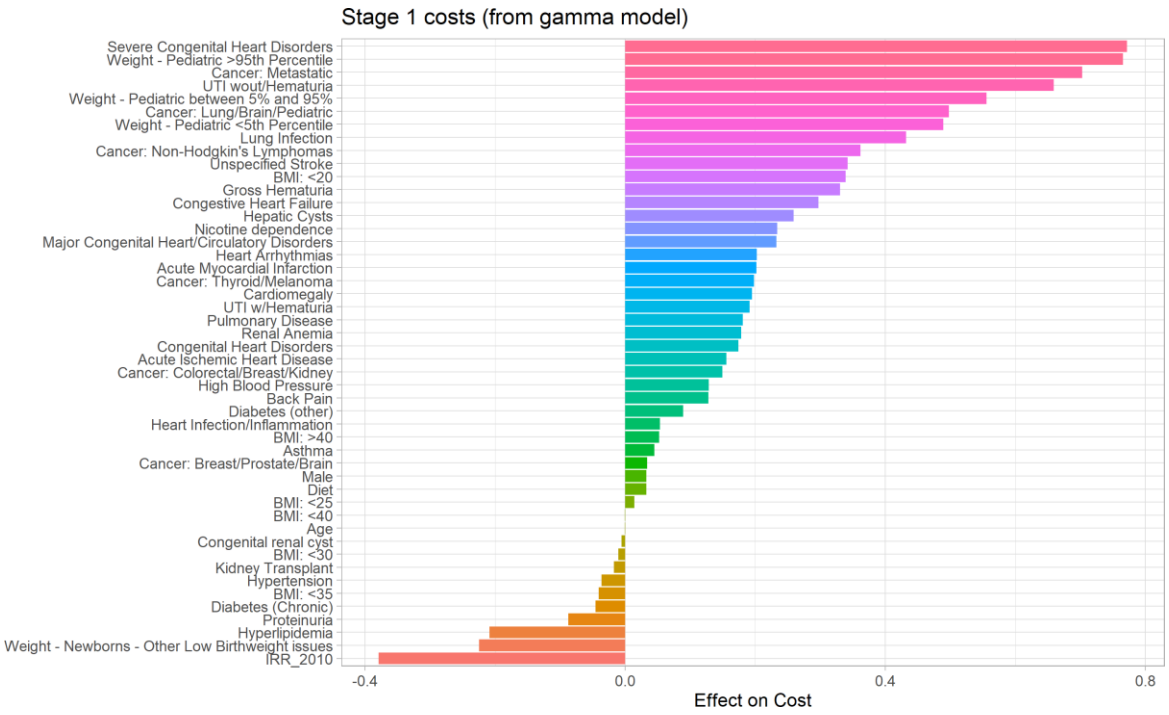


Figure D.26

MEDICARE GAMMA REGRESSION COST MODEL FOR STAGE 2

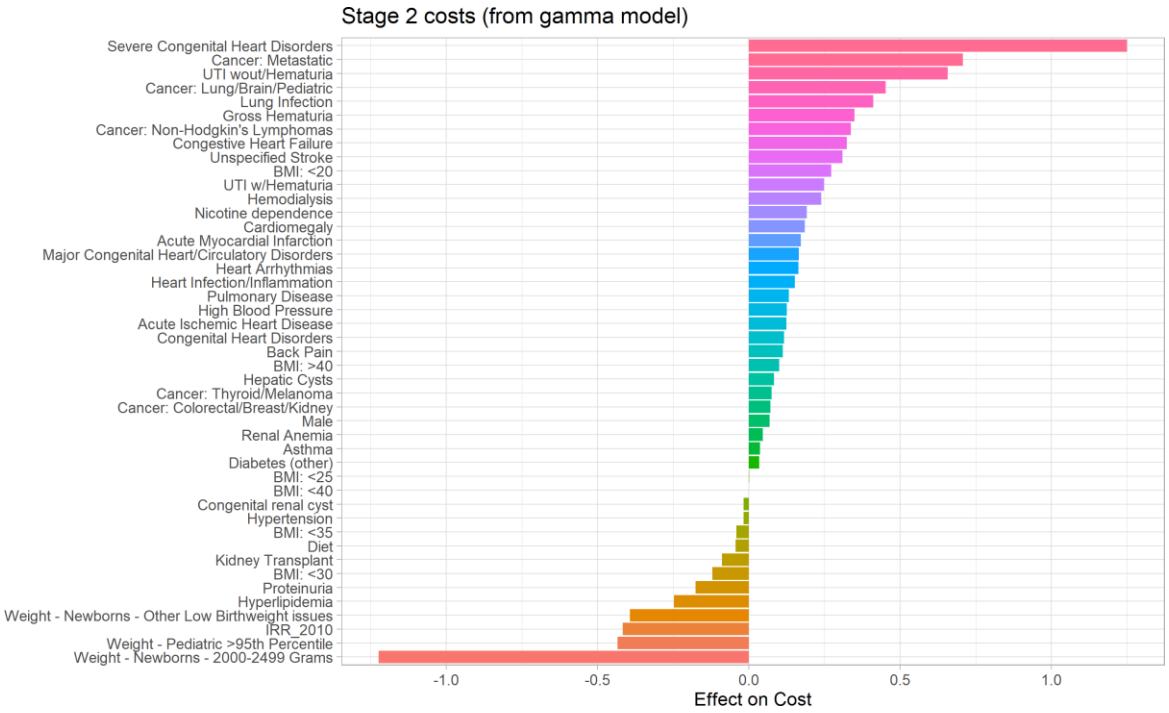


Figure D.27

MEDICARE GAMMA REGRESSION COST MODEL FOR STAGE 3

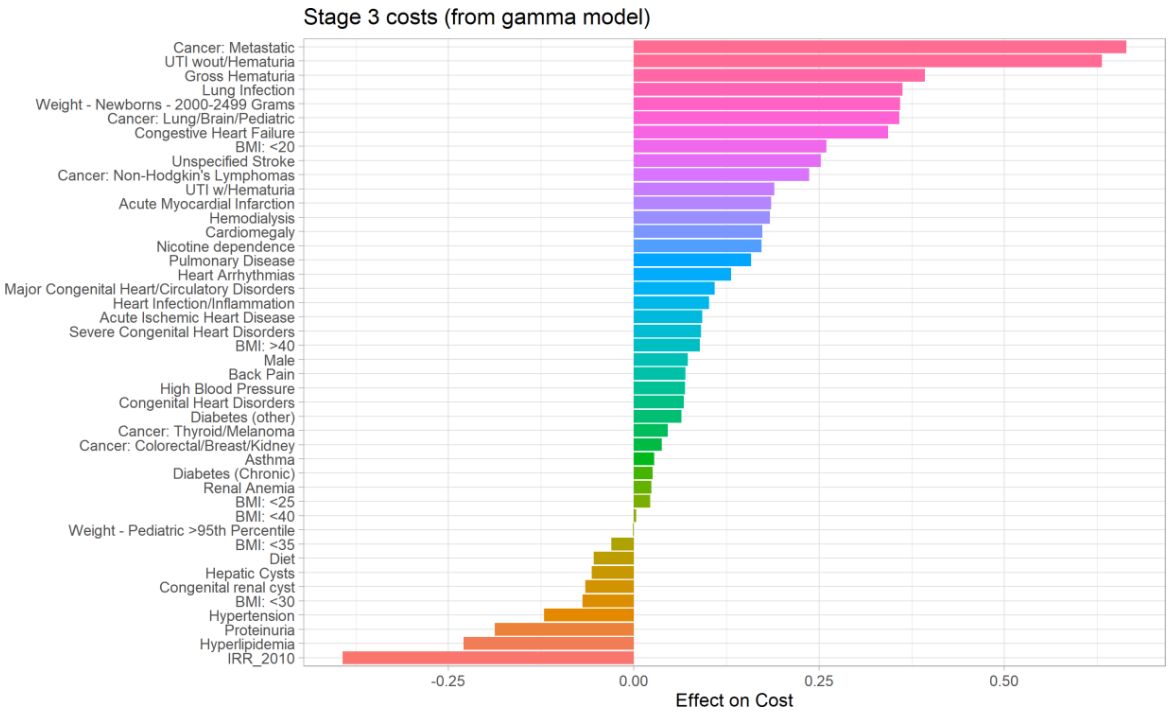


Figure D.28

MEDICARE GAMMA REGRESSION COST MODEL FOR STAGE 4

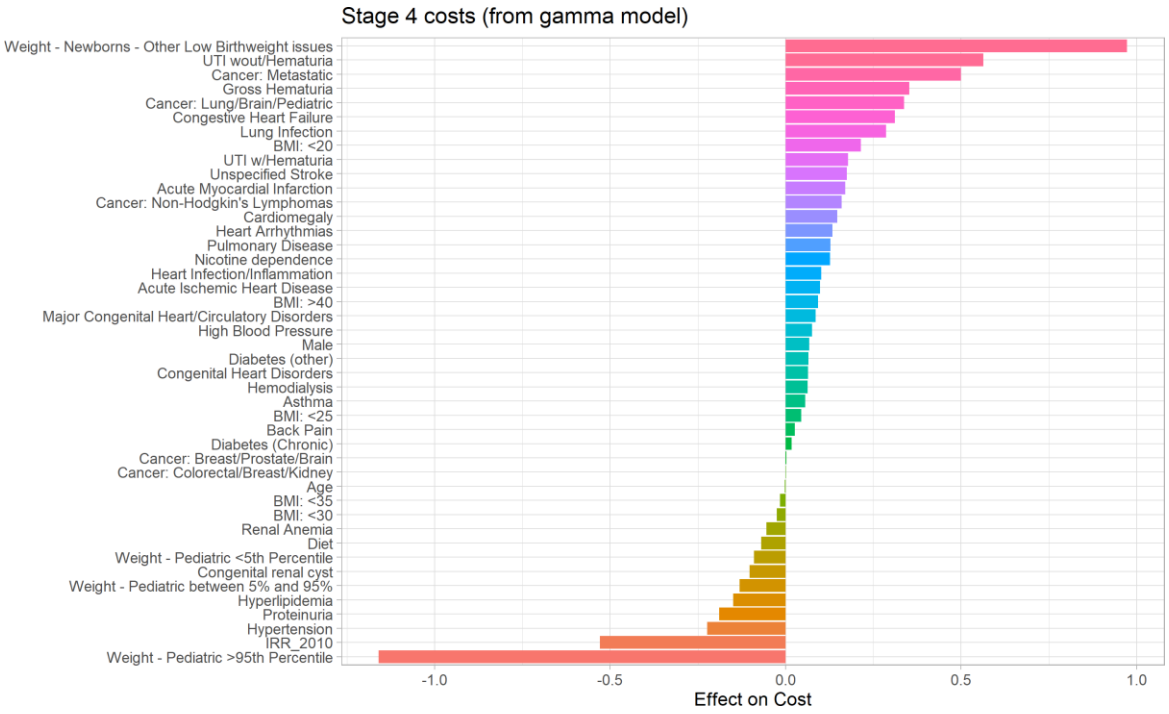


Figure D.29

MEDICARE GAMMA REGRESSION COST MODEL FOR STAGE 5

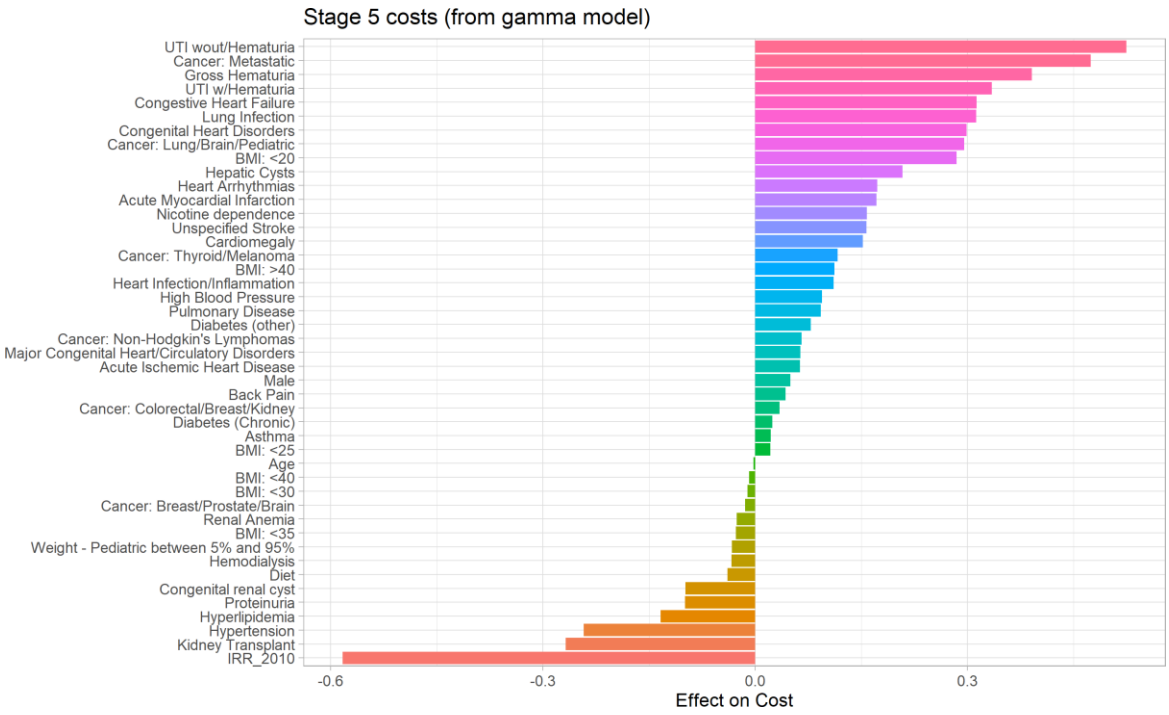
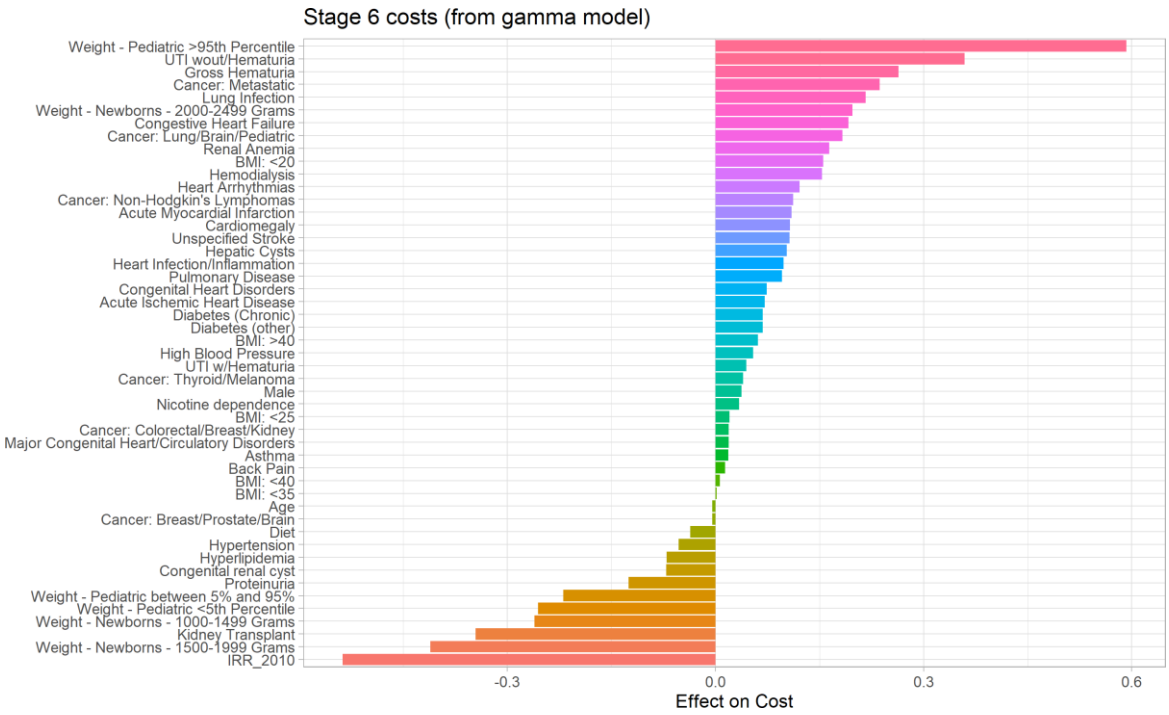


Figure D.30

MEDICARE GAMMA REGRESSION COST MODEL FOR STAGE 6



D.2.3 TWEEDIE REGRESSION

The following pages present results of using the Medicare Tweedie regression cost model for stage 1 (Figure D.31), stage 2 (Figure D.32), stage 3 (Figure D.33), stage 4 (Figure D.34), stage 5 (Figure D.35) and stage 6 (Figure D.36)

Figure D.31
MEDICARE TWEEDIE REGRESSION COST MODEL FOR STAGE 1

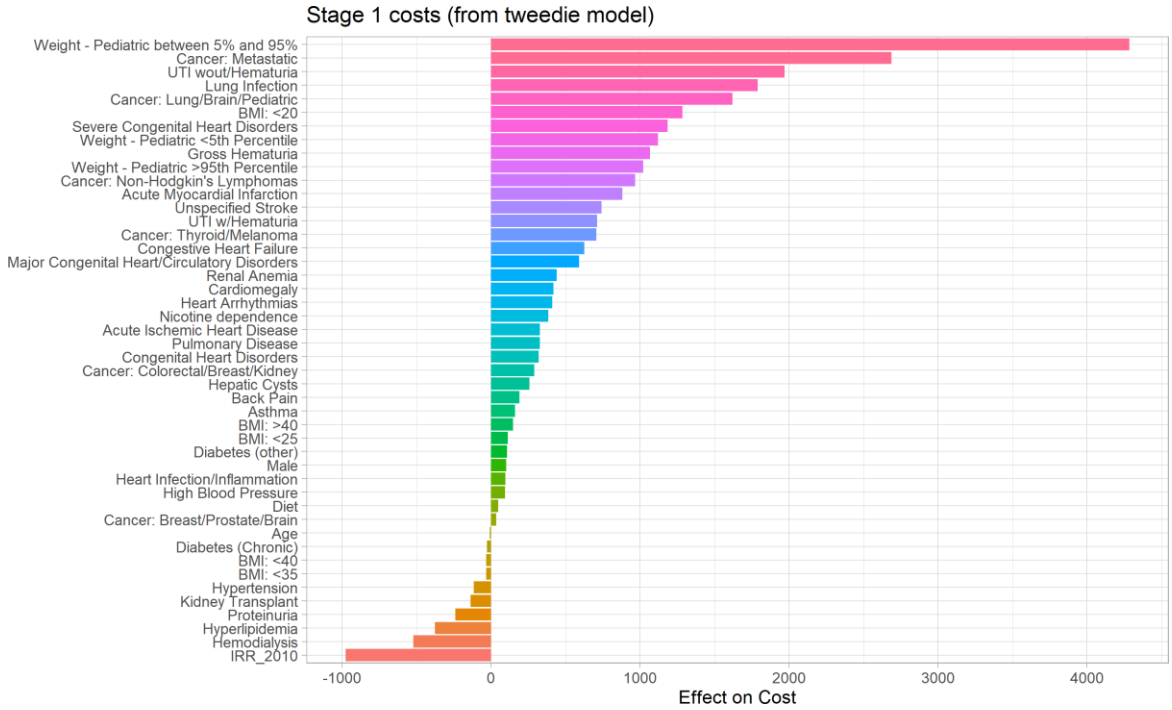


Figure D.32
MEDICARE TWEEDIE REGRESSION COST MODEL FOR STAGE 2

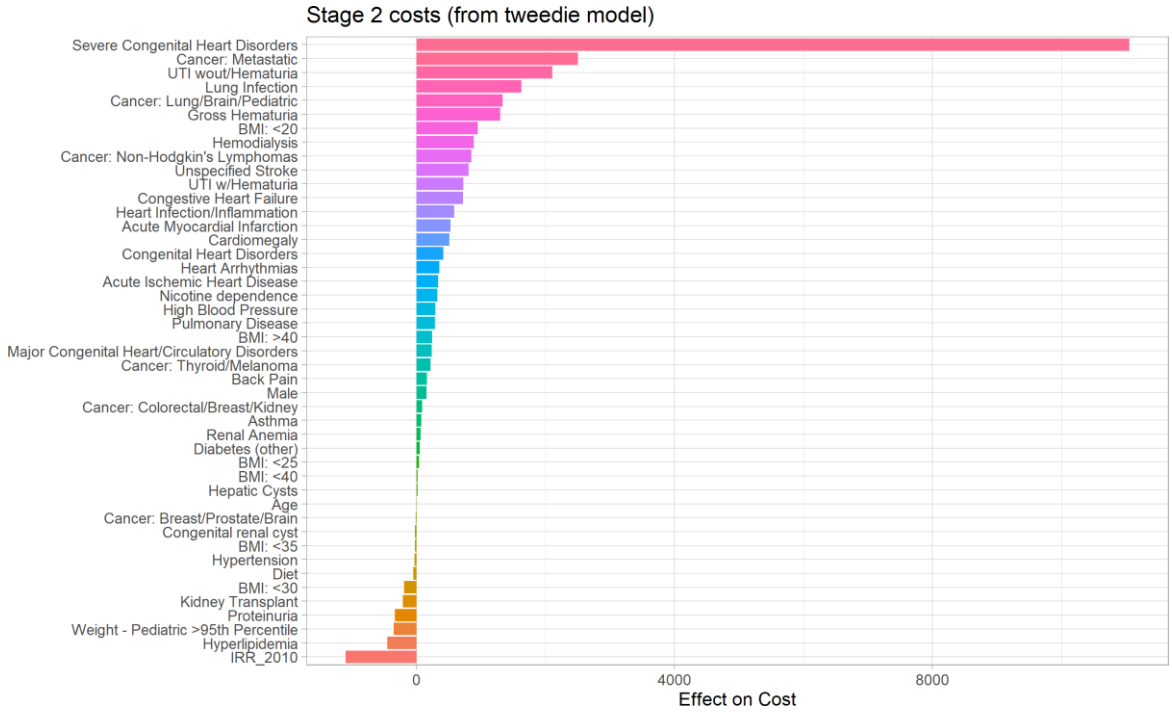


Figure D.33
MEDICARE TWEEDIE REGRESSION COST MODEL FOR STAGE 3

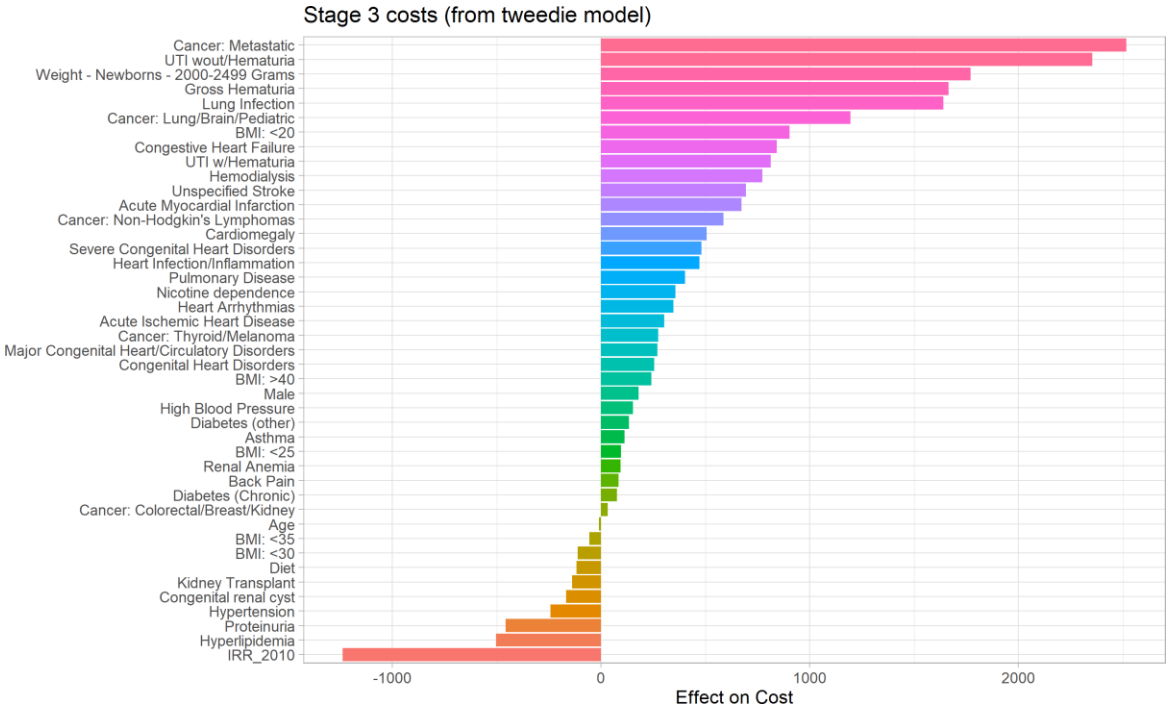


Figure D.34
MEDICARE TWEEDIE REGRESSION COST MODEL FOR STAGE 4

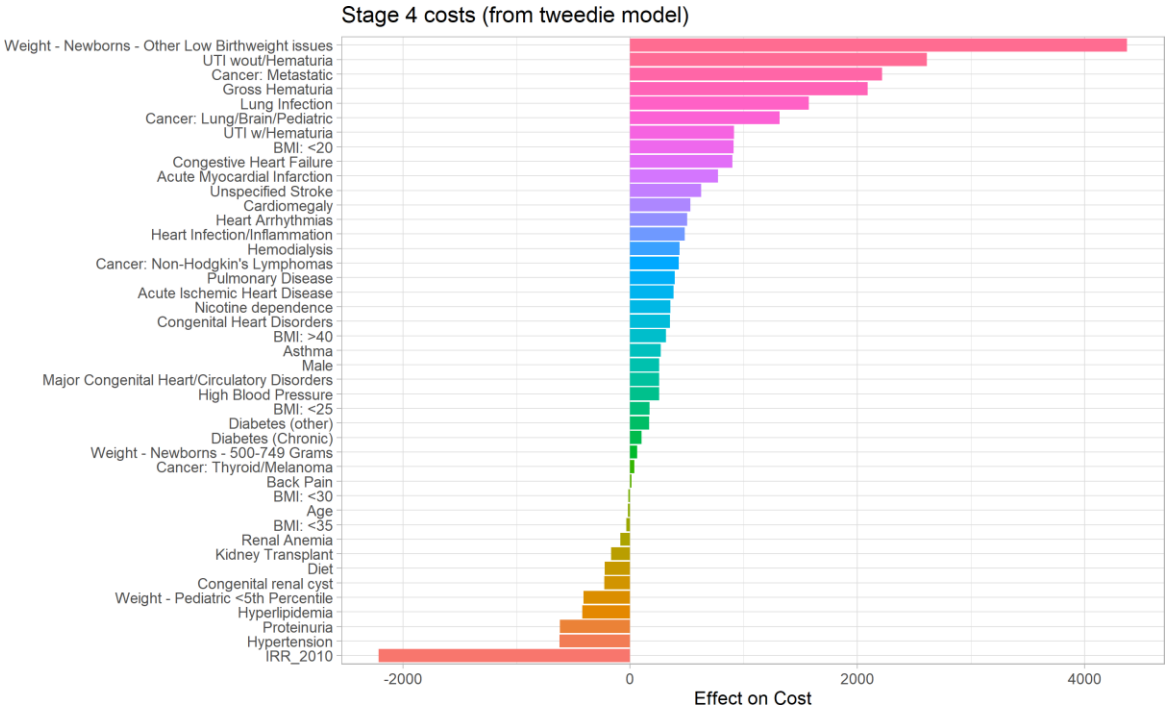


Figure D.35

MEDICARE TWEEDIE REGRESSION COST MODEL FOR STAGE 5

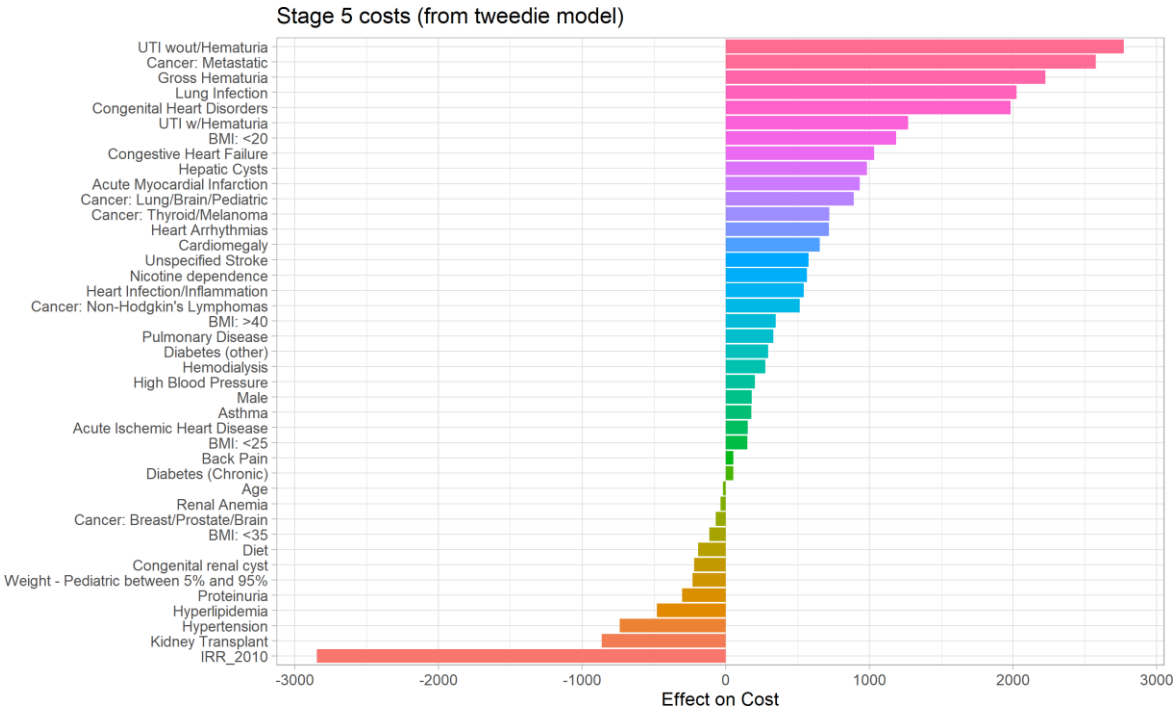
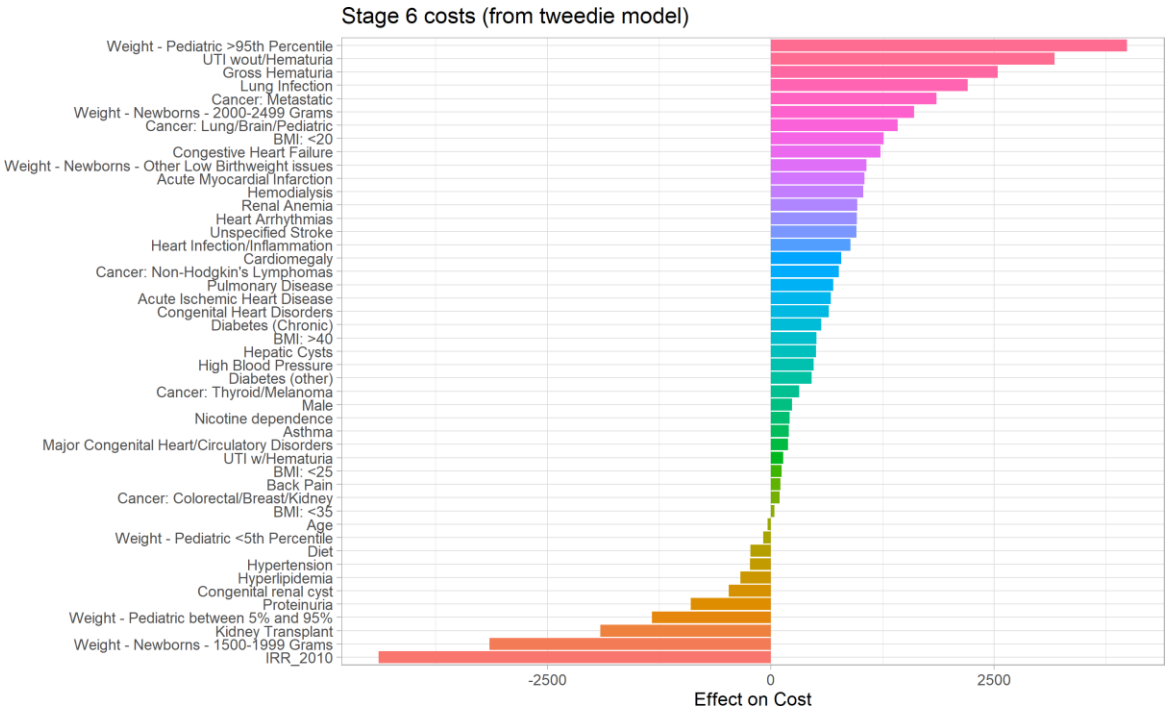


Figure D.36

MEDICARE TWEEDIE REGRESSION COST MODEL FOR STAGE 6



About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org