Supplemental Online Content

Dieleman JL, Chen C, Crosby SW, et al. US health care spending by race and ethnicity, 2002-2016.

eSupplement. Supplementary Methods and Results

This supplemental material has been provided by the authors to give readers additional information about their work.

Supplement: US Healthcare Spending by Race and Ethnicity, 2002–2016

Authors: Joseph L. Dieleman, PhD; Institute for Health Metrics and Evaluation

Carina Chen, MA; Institute for Health Metrics and Evaluation

Sawyer W. Crosby, BA; Institute for Health Metrics and Evaluation

Angela Liu, MPH; Johns Hopkins University

Darrah McCracken, PhD; Institute for Health Metrics and Evaluation

Ian A. Pollock, MLS; Institute for Health Metrics and Evaluation

Maitreyi Sahu, MPH; Institute for Health Metrics and Evaluation

Golsum Tsakalos, MS; Institute for Health Metrics and Evaluation

Laura Dwyer-Lindgren, PhD; Institute for Health Metrics and Evaluation

Annie Haakenstad, PhD; Institute for Health Metrics and Evaluation

Ali H. Mokdad, PhD; Institute for Health Metrics and Evaluation

Gregory A. Roth, MD, MPH; Institute for Health Metrics and Evaluation

Kirstin Woody Scott, MD, MPhil, PhD; University of Michigan

Christopher J.L. Murray, MD, DPhil; Institute for Health Metrics and Evaluation

Corresponding author: Joseph L. Dieleman, PhD

dieleman@uw.edu

Institute for Health Metrics and Evaluation

University of Washington

Seattle, WA 98195

Revision Date: May 21, 2021

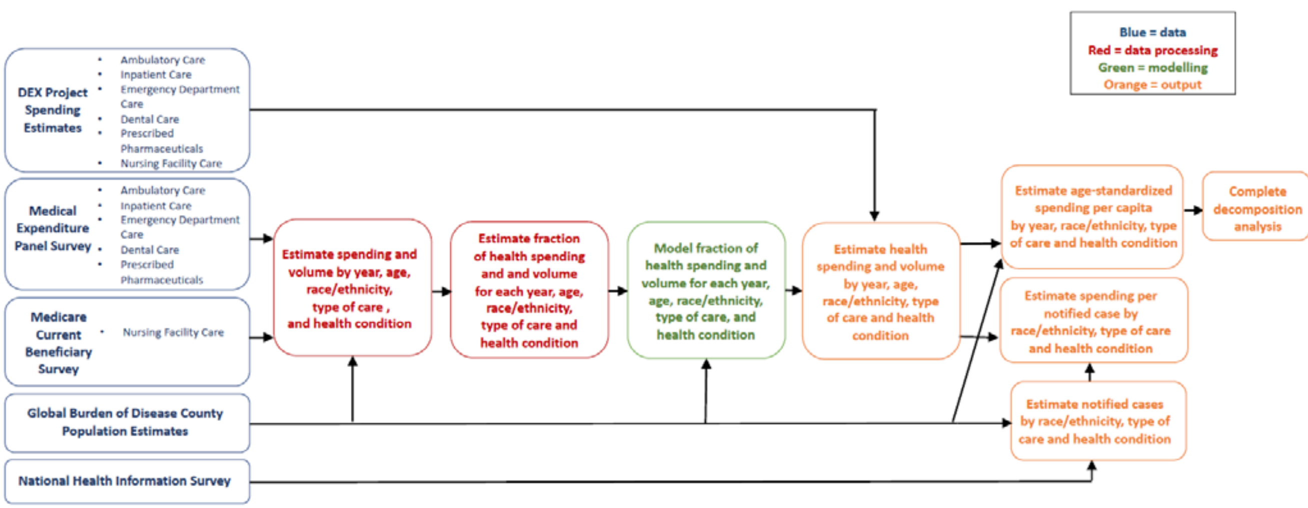
Table of Contents

S1. Conceptual framework

This work contributes to existing research on US healthcare expenditures by comprehensively estimating spending on six race/ethnicity groups with a focus on health spending and utilization disparities for specific health conditions and types of care from 2002 to 2016. This analysis draws heavily from existing research and methods which produced estimates of spending disaggregated by payer, health condition, age and sex, and type of care for 1996–2016.[1] Much of the data processing conducted for this research is identical to that of the above research paper. Namely, the underlying sources of expenditure and utilization data (the Medical Expenditure Panel Survey [2002-2016] and Medicare Current Beneficiary Survey [2002-2012]), the ways in which diagnosis codes were assigned to health conditions, the various adjustments made to address limitations of the survey data, and the process of scaling estimates to the spending totals presented in National Health Expenditure Accounts are identical. Ultimately, the final output and some of the intermediate files from the previous research were inputs to this research. Some of these processes are briefly addressed herein; however, for additional details about these steps, one should refer to the paper cited above, in particular its supplementary appendix. The current appendix focuses on the portions of our methods which are distinct from that previous research, including: (1) the classification of race/ethnicity, (2) aggregation and population adjustments, (3) our modeling framework, (4) age-standardization, (5) the incorporation of notified case data from the National Health Interview Survey, and (6) the decomposition of spending differences into changes in utilization versus changes in price/intensity.

A model of the core steps in the US DEX race and ethnicity process is illustrated in eFigure 1.

eFigure 1: US DEX race and ethnicity process diagram



S2. Sources of data

S2.1 Overview

The Medical Expenditure Panel Survey (MEPS) was a primary data source used to estimate the distribution of annual health spending across age, sex, race/ethnicity category and health condition groups. MEPS is produced by the US Agency for Healthcare Research and Quality (AHRQ), and provides data on the frequency of health services, health status and conditions, payments, and methods of payment for health services. MEPS draws from an annual survey sample of between 21,000 and 37,000 non-institutionalized civilians. Survey weights included in the data were used throughout this study to make MEPS estimates nationally representative. For each health system encounter, MEPS reports information on both payments and causes of health system encounter based on the International Classification of Disease 9th and 10th revisions (ICD9 and ICD10). All data reporting for 2015 or prior are in ICD9, while 2016 data are in ICD10. MEPS is already disaggregated into types of goods and services, which generally correspond closely to the types of services noted in the National Health Expenditure Accounts (NHEA). See previous research cited above to adjustments made to the NHEA data to ensure alignment.

The Medicare Current Beneficiary Survey (MCBS) was used to supplement information from Medicare and Medicaid claims for nursing facilities. The MCBS is a nationally representative sample of those on Medicare, including spending in nursing homes. The MCBS includes not only nursing care spending covered by Medicare, but also supplemental insurance and out-of-pocket spending. These spending estimates were stratified by age, year, sex, race/ethnicity category, and cause.

Notified cases data were obtained from the National Health Interview Survey (NHIS) for 2016 by race/ethnicity group. Notified cases are individuals who reported being told by a healthcare provider that they have a health condition. We used these data to assess how healthcare spending varied after case notification across race/ethnicity groups by reporting spending per notified case. Of the 154 health conditions in the Disease Expenditure Project 2019, seven conditions were selected for the analysis of spending per notified case: (1) cardiovascular diseases, (2) cerebrovascular diseases, (3) diabetes, (4) low-back and neck pain, (5) asthma, (6) chronic obstructive pulmonary disease (COPD), and (7) hypertension. These seven conditions were selected primarily for two reasons: (1) the notified case estimates were available through the NHIS, and (2) the conditions were prevalent enough that disaggregating by race and disease did not result in prohibitively small sample sizes. In addition, cancers were excluded, because the large prevalence of skin cancers among White individuals combined with low spending per case generated results that were not helpful in illuminating broader patterns.

Finally, we extracted population estimates for 2002–2016 from the Global Burden of Diseases, Injuries, and Risk Factors Study 2019 (GBD 2019). Age-specific population data on race/ethnicity groups were extracted from the US Census Bureau population estimates. Additional data from the National Center for Health Statistics (NCHS) bridged-race population series was used to split the 0–4 years age group in the Census estimates.[2–5]

For all survey data, sample weights provided by the agency that conducted the survey were used for all analyses in order to account for complex survey design.

S2.2 Limitations of nursing facility care data

Data from NNHS, CMS-SNF, and Medicaid were used to estimate spending and volume for the nursing care type of service. All three data sources have limitations. NNHS is nationally representative, but it is sparse and only covers three years between 1996 and 2016. CMS-SNF is more comprehensive for short-term nursing home visits but not nationally representative, as it only tracks patients at skilled-nursing facilities (SNFs) who are Medicare-eligible. Medicaid tracks Medicaid beneficiaries, but it is still not nationally representative of all nursing home spending and volume.

In the Disease Expenditure project, these three data sources are combined to apply the time trends found in CMS-SNF and Medicaid to the sparse yet nationally representative estimates of NNHS. Short-term and long-term stays are known to have different disease profiles, and they are also known to have changed differently over the past 15 years. Consequently, nursing care spending was estimated separately for short-term and long-term stays. The results were then aggregated to estimate all health spending in nursing homes from 1996 to 2016. However, the volume of nursing facility care utilization was not estimated in the Disease Expenditure project; for this reason, we were not able to estimate race/ethnicity-specific volume for nursing facility care.

S2.3 Data Completeness

eTable 1A: Number of health system encounters (visits, admissions or prescriptions) included in our dataset, by survey and year

data source	year	number of observations	number of weighted observations
MEPS_AM	2002	195,692	1,654,353,537
MEPS_AM	2003	167,209	1,669,822,830
MEPS_AM	2004	169,251	1,697,815,742
MEPS_AM	2005	163,674	1,697,808,399
MEPS_AM	2006	165,230	1,661,520,961
MEPS_AM	2007	144,560	1,636,950,211
MEPS_AM	2008	143,745	1,659,427,636
MEPS_AM	2009	163,561	1,687,244,821
MEPS_AM	2010	136,160	1,679,052,935
MEPS_AM	2011	142,271	1,718,751,134
MEPS_AM	2012	152,591	1,719,110,971
MEPS_AM	2013	174,679	1,898,774,003
MEPS_AM	2014	177,104	2,062,581,042
MEPS_AM	2015	182,928	2,043,327,505
MEPS_AM	2016	175,788	1,965,319,663

eTable 1B: Number of health system encounters (visits, admissions or prescriptions) included in our dataset, by survey and year

data source	year	number of observations	number of weighted observations
MEPS_ER	2002	6,419	47,805,076
MEPS_ER	2003	5,625	46,874,664
MEPS_ER	2004	5,436	46,419,540
MEPS_ER	2005	5,184	45,351,396
MEPS_ER	2006	5,195	45,350,195
MEPS_ER	2007	4,566	44,181,042
MEPS_ER	2008	5,026	47,386,552
MEPS_ER	2009	5,656	47,315,456
MEPS_ER	2010	4,449	43,071,755
MEPS_ER	2011	5,155	47,826,619
MEPS_ER	2012	5,669	49,783,254
MEPS_ER	2013	6,832	59,296,415
MEPS_ER	2014	6,547	61,278,219
MEPS_ER	2015	6,545	61,255,585
MEPS_ER	2016	6,243	58,251,490

eTable 1C: Number of health system encounters (visits, admissions or prescriptions) included in our dataset, by survey and year

data source	year	number of observations	number of weighted observations
MEPS_IP	2002	3,775	29,396,748
MEPS_IP	2003	3,265	29,271,550
MEPS_IP	2004	3,304	30,194,246
MEPS_IP	2005	3,211	29,483,380
MEPS_IP	2006	3,198	29,085,372
MEPS_IP	2007	2,929	30,283,472
MEPS_IP	2008	2,687	29,035,848
MEPS_IP	2009	3,191	28,820,630
MEPS_IP	2010	2,728	29,714,899
MEPS_IP	2011	2,698	27,923,333
MEPS_IP	2012	2,937	29,280,993
MEPS_IP	2013	2,732	29,828,924
MEPS_IP	2014	2,813	28,362,423
MEPS_IP	2015	2,815	29,063,135
MEPS_IP	2016	2,690	28,498,964

eTable 1D: Number of health system encounters (visits, admissions or prescriptions) included in our dataset, by survey and year

data source	year	number of observations	number of weighted observations
MEPS_RX	2002	332,322	2,692,969,879
MEPS_RX	2003	298,293	2,801,512,035
MEPS_RX	2004	310,430	2,936,055,909
MEPS_RX	2005	312,191	3,001,414,767
MEPS_RX	2006	336,109	3,097,579,693
MEPS_RX	2007	293,986	3,070,805,431
MEPS_RX	2008	287,606	3,149,914,963
MEPS_RX	2009	326,575	3,185,711,528
MEPS_RX	2010	295,028	3,273,394,914
MEPS_RX	2011	308,248	3,304,283,602
MEPS_RX	2012	318,671	3,214,685,627
MEPS_RX	2013	321,552	3,309,785,256
MEPS_RX	2014	313,123	3,421,182,889
MEPS_RX	2015	324,957	3,410,460,379
MEPS_RX	2016	315,212	3,254,235,252

eTable2A: Response rates for surveys, by data source and year

Data source	Year	Overall response rates
MEPS	2016	46.0%
MEPS	2015	47.7%
MEPS	2014	48.5%
MEPS	2013	52.8%
MEPS	2012	56.3%
MEPS	2011	54.9%
MEPS	2010	53.5%
MEPS	2009	57.2%
MEPS	2008	59.3%
MEPS	2007	56.9%
MEPS	2006	58.3%
MEPS	2005	61.3%
MEPS	2004	63.1%
MEPS	2003	64.5%
MEPS	2002	64.7%

eTable2B: Response rates for surveys, by data source and year

Data source	Year	Overall response rates
MCBS	2012	62.30%
MCBS	2011	62.30%
MCBS	2010	62.40%
MCBS	2009	60.50%
MCBS	2008	66.70%
MCBS	2007	69.80%
MCBS	2006	70.60%
MCBS	2005	71.00%
MCBS	2004	70.90%
MCBS	2003	69.50%
MCBS	2002	70.40%

eTable2C: Response rates for surveys, by data source and year

Data source	Year	Household module response rates	Sample child module response rates	Sample adult module response rates
NHIS	2016	67.90%	61.90%	54.30%
NHIS	2002	89.60%	81.30%	74.30%

eTable 3: Proportion of observations in each race/ethnicity category and observations that are not assigned to a race/ethnicity, by survey

Race/ethnicity	MCBS nursing facility care	MEPS Ambulatory Care	MEPS Emergency Department Care	MEPS Inpatient Care	MEPS Prescribed Pharmaceuticals
American Indian, Alaska Native (non-Hispanic)	0.47%	0.55%	0.91%	0.66%	0.73%
Asian, Native Hawaiian, Pacific Islander (non-Hispanic)	0.94%	4.35%	2.60%	3.18%	3.60%
Black (non-Hispanic)	10.58%	15.06%	23.65%	21.18%	18.98%
Hispanic	3.90%	18.23%	24.52%	20.04%	16.65%
Multiple races (non-Hispanic)	0.62%	2.17%	2.81%	1.93%	1.84%
White (non-Hispanic)	83.32%	59.64%	45.50%	53.02%	58.20%
observations not assigned to a race/ethnicity category	0.17%	0.00%	0.00%	0.00%	0.00%

S3. Demographic categorization

To provide a comprehensive yet granular set of estimates, the type of care, the age, sex, self-identified race and ethnicity of the patient, the diagnosis, and the health care spending and encounters (visits, admission, or prescriptions) were identified and extracted from the Medical Expenditure Panel Survey (MEPS) and the Medicare Current Beneficiary Survey (MCBS). In MEPS data that were used, there are a total of 7,278,266 observations in the dataset from year 2002 to 2016. For year 2002 to 2011, the race and ethnicity status were assigned to each observation based on the demographic variables RACEX and HISPANX. The variable RACEX has six categories: “1 WHITE - NO OTHER RACE REPORTED,” “2 BLACK - NO OTHER RACE REPORTED,” “3 AMER INDIAN/ALASKA NATIVE - NO OTHER RACE,” “4 ASIAN - NO OTHER RACE REPORTED,” “5 NATIVE HAWAIIAN/PACIFIC ISLANDER - NO OTHR,” and “6 MULTIPLE RACES REPORTED” and the HISPANX variable has two categories: “1 HISPANIC” and “2 NOT HISPANIC.” If a surveyed respondent is self-identified as “1 HISPANIC” in the HISPANX variable, then no matter which RACEX category is selected, then this respondent will be identified as a Hispanic individual. Then, for the non-Hispanic surveyed respondents with the “1 WHITE - NO OTHER RACE REPORTED,” “2 BLACK - NO OTHER RACE REPORTED,” “3 AMER INDIAN/ALASKA NATIVE - NO OTHER RACE,” and “6 MULTIPLE RACES REPORTED” categories in RACEX will be identified as White (non-Hispanic), Black (non-Hispanic), American Indian or Alaska Native (non-Hispanic) and multiple race (non-Hispanic). The non-Hispanic surveyed respondents with “4 ASIAN – NO OTHER RACE REPORTED,” “5 NATIVE HAWAIIAN/PACIFIC ISLANDER - NO OTHR” categories in RACEX will be identified as Asian, Native Hawaiian, or Pacific Islander (non-Hispanic).For year 2012 to 2013, race and ethnicity status are determined using the demographic variables RACEV1X and HISPANX. Since the RACEV1X and HISPANX variables store the same categories as the RACEX and HISPANX in year 2002 to 2011, we used the same method to assign a race/ethnicity value to each observation in the data.

For year 2014 to 2016, the race and ethnicity status was assigned to each observation based on the demographic variables RACEV1X and HISPANX. The variable RACEV1X has five categories: “1 WHITE - NO OTHER RACE REPORTED,” “2 BLACK - NO OTHER RACE REPORTED,” “3 AMER INDIAN/ALASKA NATIVE - NO OTHER RACE,” “4 ASIAN/NATV HAWAIIAN/PACFC ISL - NO OTH” and “6 MULTIPLE RACES REPORTED,” and the HISPANX variable has two categories: “1 HISPANIC” and “2 NOT HISPANIC.” If a surveyed respondent is self-identified as “1 HISPANIC” in the HISPANX variable, then no matter which RACEV1X category is selected, then this respondent will be identified as a Hispanic individual. for the non-Hispanic surveyed respondents with the “1 WHITE - NO OTHER RACE REPORTED,” “2 BLACK - NO OTHER RACE REPORTED,” “3 AMER INDIAN/ALASKA NATIVE - NO OTHER RACE,” “4 ASIAN/NATV HAWAIIAN/PACFC ISL - NO OTH,” and “6 MULTIPLE RACES REPORTED” categories in RACEV1X will be identified as White (non-Hispanic), Black (non-Hispanic), American Indian or Alaska Native (non-Hispanic), Asian, Native Hawaiian, or Pacific Islander (non-Hispanic), and multiple race (non-Hispanic).

In MCBS data, there are a total of 13,638 observations in the dataset from year 2002 to 2012. Each observation corresponds to a person in the survey. The race and ethnicity are also reported separately in the MCBS; any beneficiary who reported their ethnicity as Hispanic in the ethnicity variable “D_ETHNIC,” regardless of what race they reported in the race variable “D_RACE,” was classified as Hispanic. For the non-Hispanic individuals, the race and ethnicity status were assigned based on the race variable “D_RACE,” where there are seven categories: “-8 Don’t know,” “1 American Indian,” “2 Asian or Pacific Islander,” “3 Black or African American,” “4 White,” “5 More than One,” “91 Other.” However, to maintain the consistency throughout the study, the race and ethnicity status American Indian or Alaska Native (non-Hispanic), Asian, Native Hawaiian, or Pacific Islander (non-Hispanic), Black (non-Hispanic), White (non-Hispanic), or multiple race (non-Hispanic) were assigned to the non-Hispanic surveyed respondent if they selected “1 American Indian,” “2 Asian or Pacific Islander,” “3 Black or African American,” “4 White,” or “5 More than One,” respectively. The individuals with “-8 Don’t know” and “91 Other” in variable “D_RACE” were not included in this study. These exclusions were associated with 0.16% of the total MCBS survey sample and 0.26% of total spending in MCBS data from 2002 to 2012.

Thus, the categories can be accurately defined as seven mutually exclusive and collectively exhaustive race/ethnicity groups – American Indian or Alaska Native (non-Hispanic); Asian (non-Hispanic), Native Hawaiian, or Pacific Islander (non-Hispanic); Black (non-Hispanic); Hispanic; multiple race (non-Hispanic); or White (non-Hispanic), with Asian (non-Hispanic) and Native Hawaiian or Pacific Islander (non-Hispanic) combined for this research.

In addition, age-specific population data for race/ethnicity groups were extracted from the US Census Bureau population estimates. Additional data from the National Center for Health Statistics bridged-race population series was used to split the 0–4 age group in the census estimates. For detailed information about assignment and disaggregation of the remaining demographic variables, (like age and sex), as well as epidemiological classifications and types of services, refer to section 3 of the supplemental appendix for the *US Health Care Spending by Payer and Health Condition, 1996–2016* paper.[1]

S4. Generating uncertainty

In order to generate uncertainty, the formatted survey data (MCBS and MEPS) were bootstrapped 1,000 times, producing 1,000 individual samples on which to run further aggregation and analyses. The complex survey design for MCBS and MEPS was taken into consideration for bootstrapping by using the user-written Stata package *bsweights*.[6,7] *bsweights* resamples survey data accounting for stratification and clustering in the original survey design. This process ensured that the bootstrapped draws resembled the original sampling scheme by resampling the entire primary sampling units within the original survey strata. In other words, the patient weights were incorporated into all further estimation in this research project. All subsequent statistical analyses were performed at the bootstrap draw level; that is, all of the following steps were performed 1,000 times for each draw. In order to generate the final figures and numbers, we took the mean of these draws, as well as 95% uncertainty intervals ranging from the 2.5th to the 97.5th percentile of the draws. To evaluate if two estimates were statistically different, we calculated a bootstrap p-value reflective of a two-sided hypothesis test, meaning we counted the number of times that one draw was larger (or smaller) than the other, and then divided by half of the number of draws that were tested (in our case 500). This method does not assume the uncertainty distributions are normal, and can be interpreted the same way a p-value.[8,9]

S5. Adjustments

Many data processing steps were performed, including assigning ICD codes to health conditions, adjusting for comorbidities, converting charges to actual payments, performing several type-of-care-specific adjustments, and scaling estimates to the total amounts presented in National Health Expenditure Accounts. A thorough, detailed, and well documented account of these steps exists already in sections 3 through 6 of the supplementary methods appendix for the *US Health Care Spending by Payer and Health Condition, 1996–2016* paper.[1] For further details about these calculations, please reference the materials mentioned above, as the steps taken for this research are identical to those outlined in the *US Health Care Spending by Payer and Health Condition, 1996–2016* paper.[1]

S6. Aggregation and population adjustment

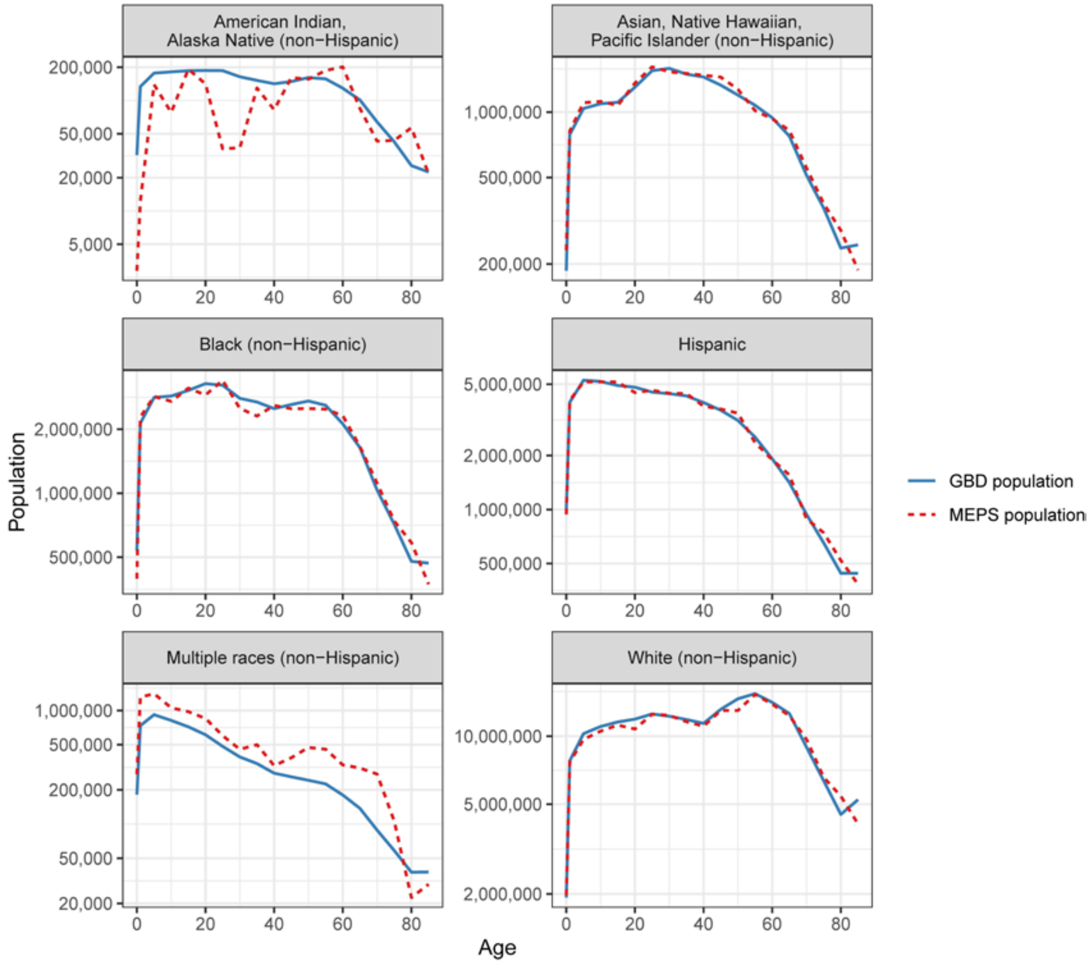
S6.1 Aggregating from the encounter level

In order to aggregate the survey data from individual encounters to total amounts of spending and volume, we took the sum of spending and volume across all surveyed individuals for each race/ethnicity category, as well as each type of care, year, age group, sex, and health condition (for ambulatory, dental, and emergency care, "volume" = visits; for prescribed pharmaceuticals, "volume" = prescriptions; for inpatient and nursing facility care, we included two volume measures for the number of admissions and the number of days spent in a facility). This aggregation was performed at the draw level.

S6.2 Adjusting the race and ethnicity population

A comparison of the racial breakdown in MEPS' surveyed population to other estimates of race/ethnicity population, including the GBD population estimates used in this study, revealed the MEPS population was not representative across race/ethnicity groups. Most importantly, MEPS comparatively overestimated the proportion of the US population that falls within the "multiple races" category, and the American Indian, Alaska Native population age structures were quite different. For this reason, we adjusted the survey data such that it corresponds with GBD population estimates. In order to make this adjustment, we adjusted spending and volume estimated in MEPS by the multiplying by the adjustment scaler $\frac{GBD\ population}{MEPS\ population}$ for each year, age group, sex, and race/ethnicity category. This is equivalent to converting the survey spending data to spending and volume per person (using the survey population), and then multiplying the per-person data with the GBD population estimates to yield total, adjusted spending and volume for each race and ethnicity category, year, age group, and sex.

eFigure 2: Comparing GBD and MEPS populations



S7. Estimating race/ethnicity-specific spending and volume of care

S7.1 Overview

In order to address problems associated with small sample sizes and data irregularity, and to generate a complete set of estimates, the adjusted health condition-, type of care-, age-, sex-, year-, and race/ethnicity-specific spending and volume estimates were regressed on penalized age and year splines, as well as a covariate that measured the fraction of the population in each race/ethnicity group for each age group and year. Models were run independently for each of the 154 health conditions, for aggregate health condition categories (for instance, aggregated categories that capture spending on all cardiovascular diseases, or all conditions), as well as for each race/ethnicity group, six types of care, and three metrics (spending, admissions, and days spent in facility) resulting in 12,276 individual models, or 12,276,000 at the draw level. Each model produced estimates for each age, sex, and year.

S7.2 Model specification

Modeling was performed using the gam ("Generalized Additive Models with Integrated Smoothness Estimation") function within the "mgcv" R package. Age group and year (used to generate the splines) were normalized to range from 0 to 1. All models included a covariate that measured the fraction of the population in each race/ethnicity group (for each age group and year). All models used data weights, which were generated based on the existing DEX estimates of health expenditures to represent the proportion of spending for a given model that occurred in each year, age group, and sex; this encouraged a better model fit for data points which represent higher levels of spending or volume (i.e., for models predicting spending on dementia, weights were higher for older age groups and more recent years since more spending and volume for dementia occurs in older age groups and more recent years).
The remaining model specification depended largely on the data density. Models with the highest data density across both age groups and year were modeled using a two-dimensional spline (tensor product) across age group and year. Models with the second-highest degree of data density were modeled using one-dimensional age and year splines, which were found to be more stable than the tensor product for conditions with many missing values. If a model had high data density across years, but low data density across age groups, only a year spline was used. Similarly, when there were data for many age groups, but limited years, only an age-spline was used. Models with insufficient data to be modeled with either an age or year spline were modeled with only the population covariate. Models which failed due to lack of data were filled with the predictions from a model with a higher level of aggregation (i.e., if a model for one of the specific 154 health conditions failed, the predictions for a more aggregated health condition were used). All tensor product models used 4 knots, which gave sufficient flexibility while allowing models to converge. Models with one-dimensional splines used between 4 and 6 knots, with more knots for models with higher data density. Finally, if the data density was not significantly different between the two sexes, models were run separately for each sex; otherwise, one model was run on the combined data from both sexes. Below, we have provided two tables to outline this process; the first table (eTable 4) depicts how we determined whether a separate model was run for each sex. The second (eTable 5) shows how we determined which types of splines were used to model.

eTable 4: Cutoffs for splitting models by sex

Specification	Condition
Model separately for each sex	One sex does not have a count of 3+ additional age groups or years above the other group (i.e., the data density is similar between the two sexes)
Model on the combined dataset for both sexes	One sex does have a count of 3+ additional age groups or years above the other group (i.e., one group has much higher data density)

eTable 5: Spline inclusion cutoffs

Specification	Age data density	Year data density	Special case
Model using the tensor product of age and year	Raw data exist for every possible** age for the given condition/type of care	Raw data exist for every possible year for the given condition/type of care	Model represented the data aggregated across all conditions for a particular type of care
Model using a one-dimensional spline across age and a one-dimensional spline across year	Raw data exist for at least 40% of the possible age groups for the given condition/type of care	Raw data exist for at least 50% of the possible years for that condition/type of care	
Model using a one-dimensional spline across year only	Raw data exist for fewer than 40% of the possible age groups for the given condition/type of care	Raw data exist for at least 50% of the possible years for that condition/type of care	
Model using a one-dimensional spline across age only	Raw data exist for at least 40% of the possible age groups for the given condition/type of care	Raw data exist for fewer than 50% of the possible years for that condition/type of care	

** For instance, for spending on neonatal conditions, there is only one possible age group, but for diabetes and Alzheimer's, there are 19 and 10 possible age groups, respectively.

S7.3 Raking

In order to ensure that the modeled proportions summed to equal 1 across the six race/ethnicity categories, estimates below 0 or above 1 were forced to equal 0 and 1, and the final estimates were proportionally raked to sum to 1. Raking modeled estimates was determined to be preferable to centered-log-transforming the data prior to modeling, as the latter led to significant underestimates of spending and volume for white people and overestimates of spending on race/ethnicity categories which accounted for a small proportion of the total spending and volume. Raking was done at the draw level.

Across all of the health conditions referenced in this research, 6.7% of the draw-level modeled estimates were forced to 1 or 0. This is expected given that the proportion of spending for some race/ethnicity categories is often close to 1 or 0, and we expect there to be variation across the draws. Furthermore, most of these impossible predictions did not extend far beyond 1 or 0; in the aggregate, across all draws, this adjustment resulted in an average absolute change of only 0.0054 in the predicted draw values, which represent the proportion of spending or volume for each race/ethnicity. Following this change, raking the estimates introduced an average absolute shift of only 0.01 in the modeled draw proportions.

S7.4 Extending long-term care estimates through 2016

Because the survey data from MCBS did not extend past 2012, we found that our long-term care models performed poorly when expected to extrapolate through 2016. For this reason, we specified our long-term care models to predict from 2002 through 2012 only, and then extended the 2012 predictions through 2016. Visual inspection showed that there was very little variation in the fraction of spending of volume for each race/ethnicity group over time, especially at the aggregate levels.

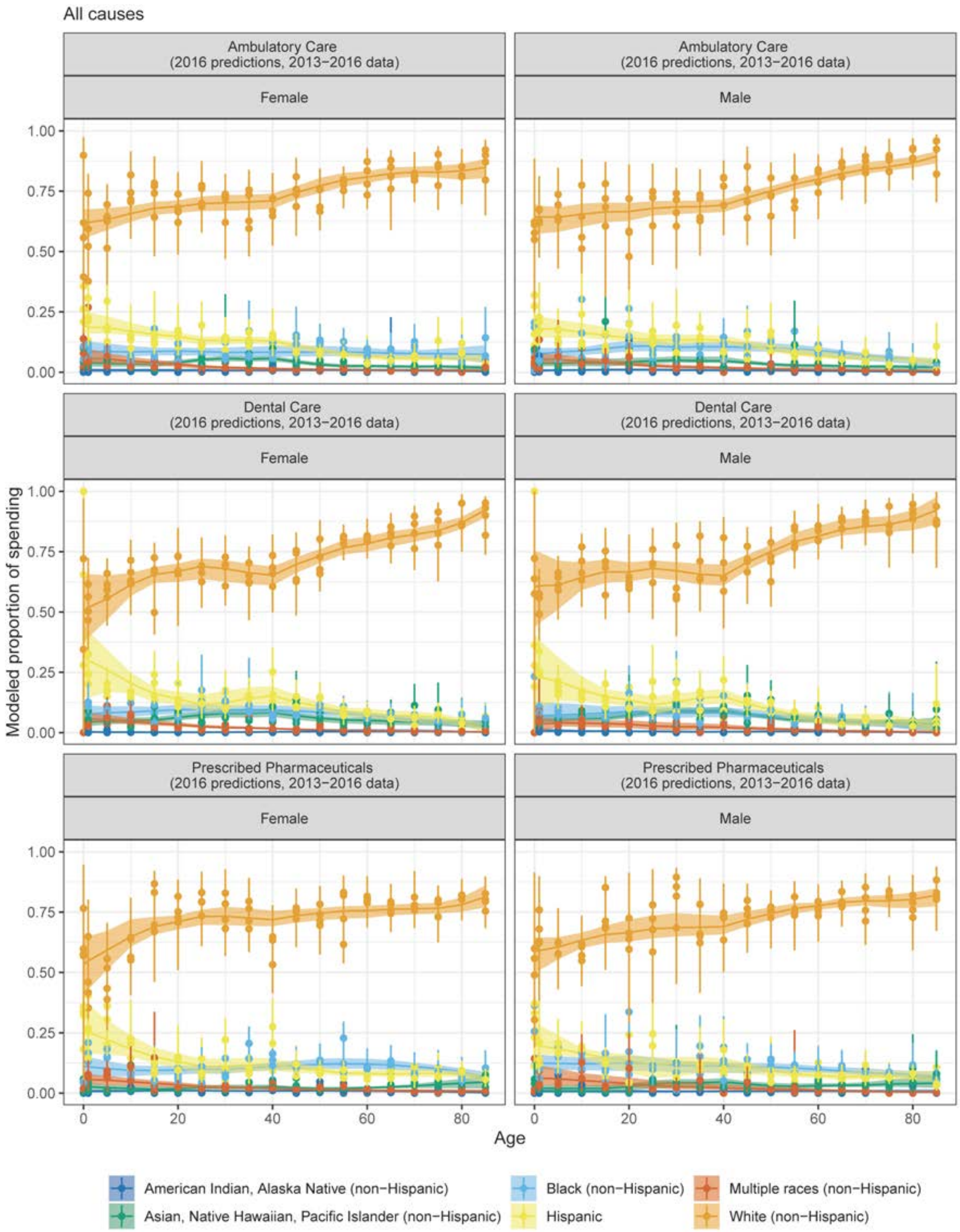
S7.5 Reviewing the estimates

The modeled proportions were vetted extensively using interactive visualizations produced with the R “shiny” package. Specifically, we examined the model fit across age and year for each race/ethnicity, type of care, health condition, and sex. In order to make our results transparent, we have provided figures that depict the age-trend predictions for the models estimating the proportion of spending in each race/ethnicity category. The estimates presented below exhaustively include each type of care and condition referenced in the paper and the appendix (that is, they include the 14 conditions from Figure 4 (in the paper) and eFigure 5, and the “All causes” aggregate of all conditions used in the other figures, along with all six types of care, where applicable). For brevity, we did not include figures depicting the time-trend of each model (which also influenced the overall model fit), and we only included the estimates for the latest year of raw data (that is, 2012 for long-term care, and 2016 for all other types of care); however, to illustrate the influence of neighboring years, each plot includes the raw data points for the latest three years of data. Colored lines represent the mean of the draw-model predictions, and shaded areas represent 95% uncertainty intervals. Data points represent the mean of the raw data’s bootstrapped draws, and vertical lines illustrate the 95% uncertainty intervals of these draws. eFigure 3 presents the spending models, and eFigure 4 presents the volume models.

S8 Reporting

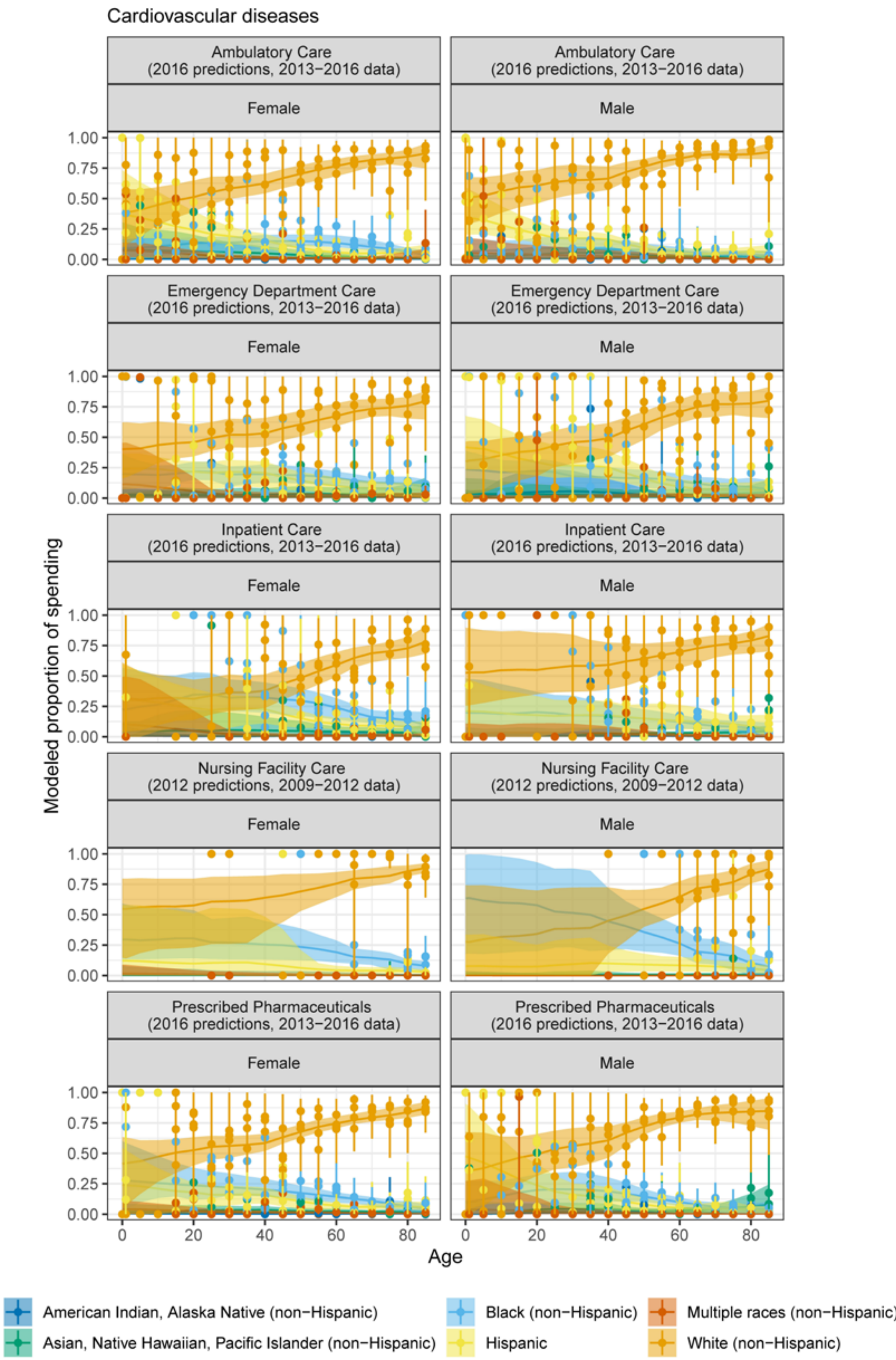
To quantify variation in spending levels among race/ethnicity groups, we used the coefficient of variation, which is a statistical measure that reports the dispersion of an outcome relative to the mean.

eFigure 3: Spending models

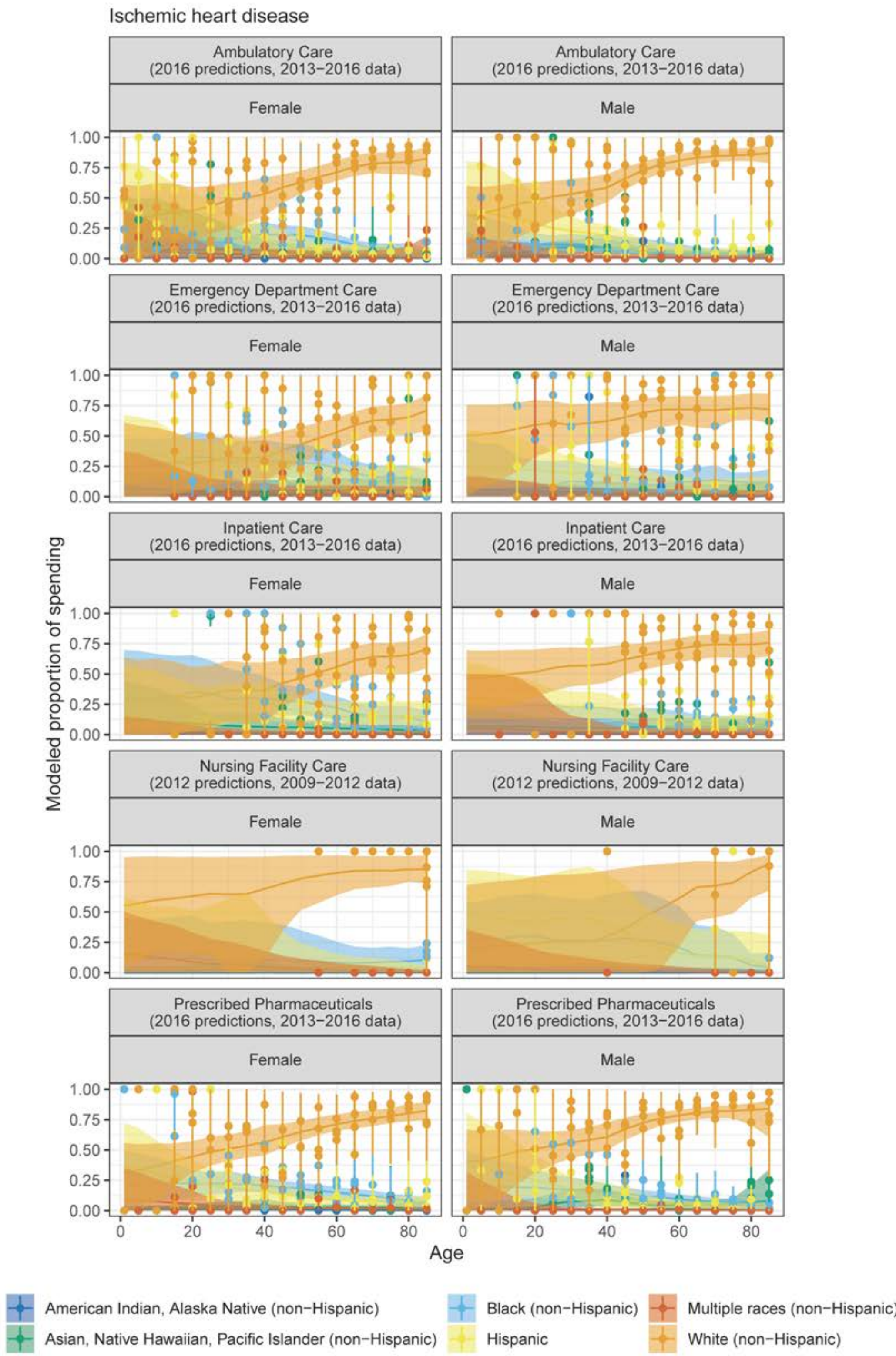


eFigure 3: Spending models

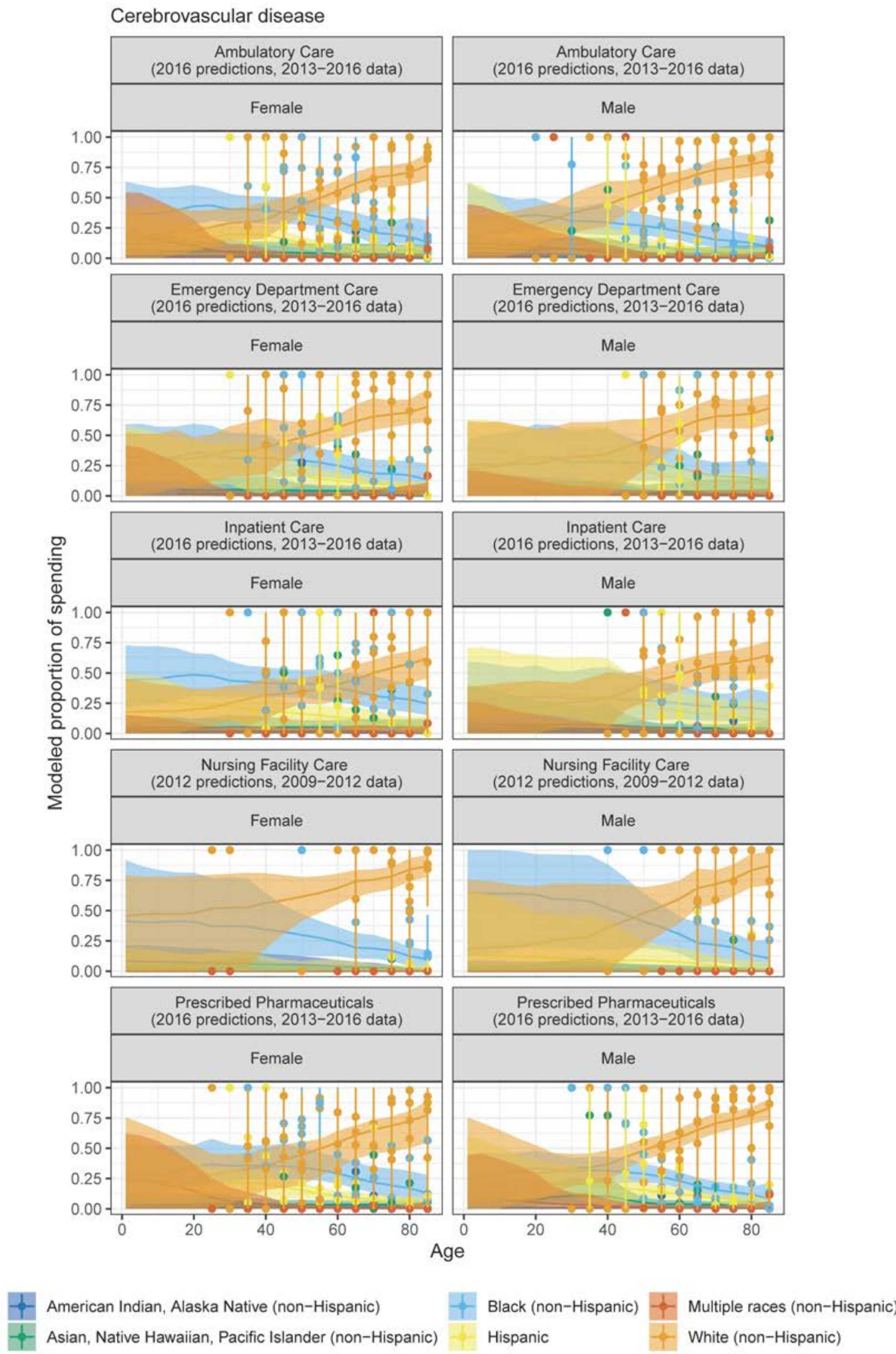
eFigure 3: Spending models



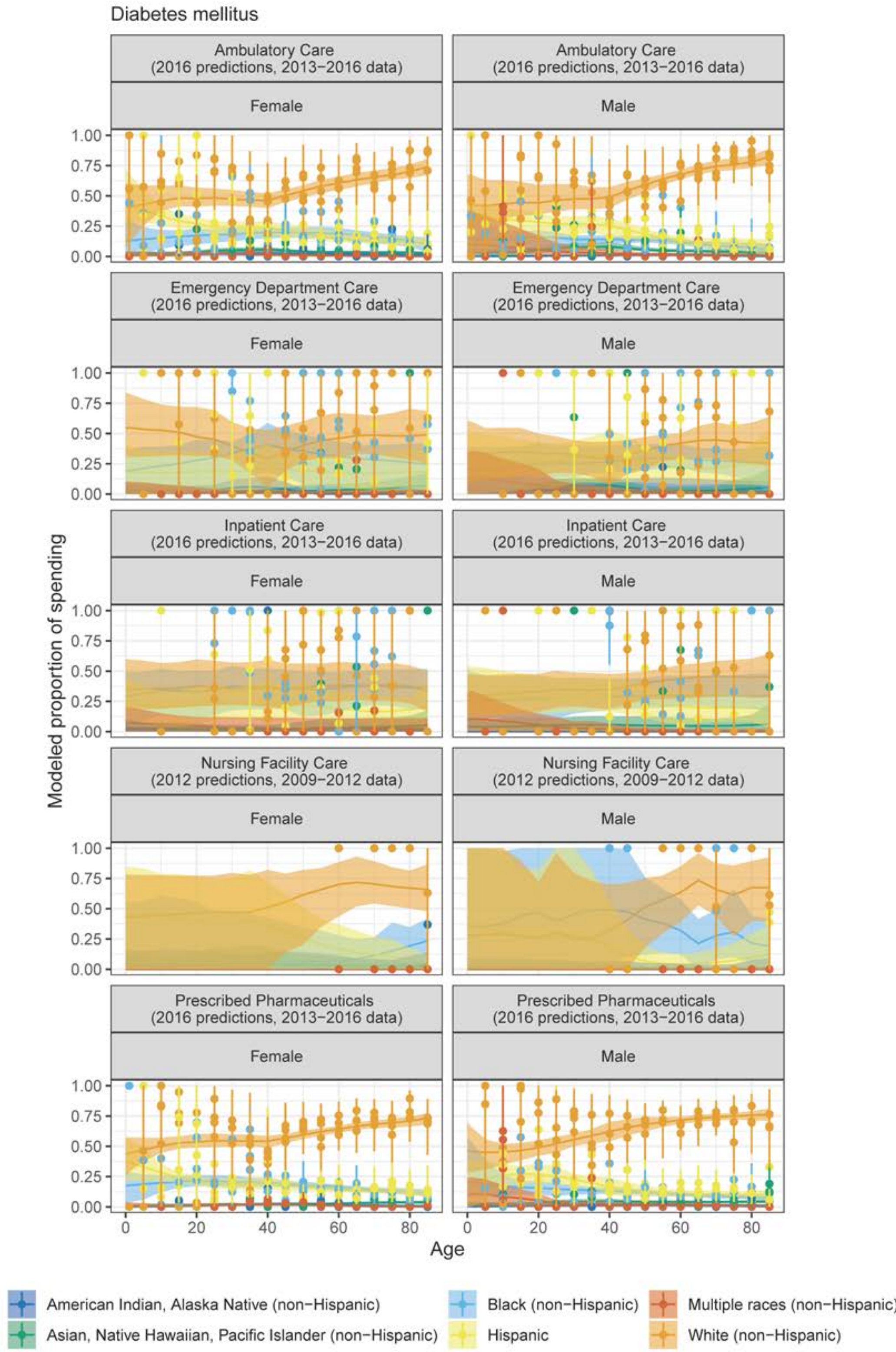
eFigure 3: Spending models



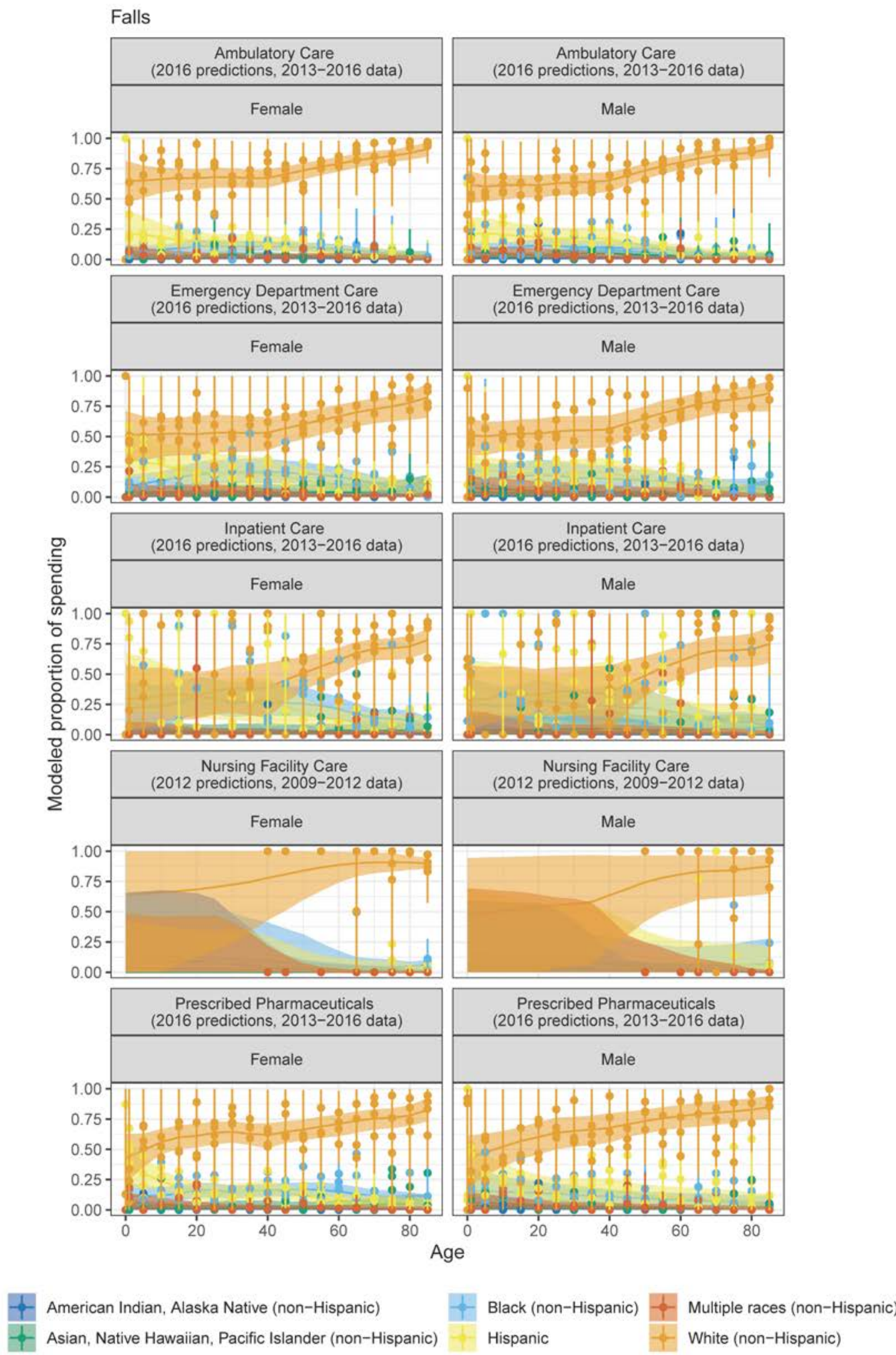
eFigure 3: Spending models



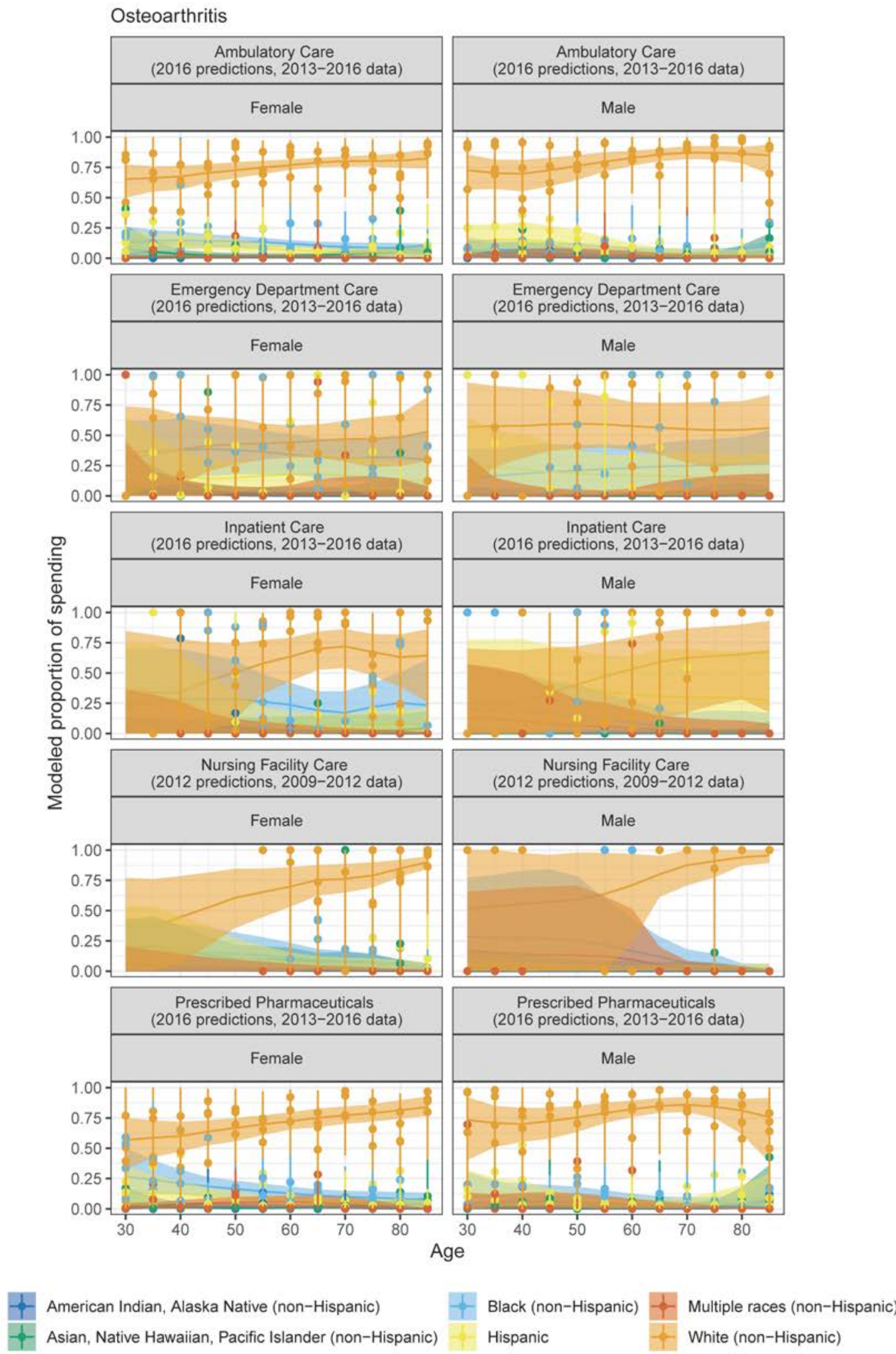
eFigure 3: Spending models



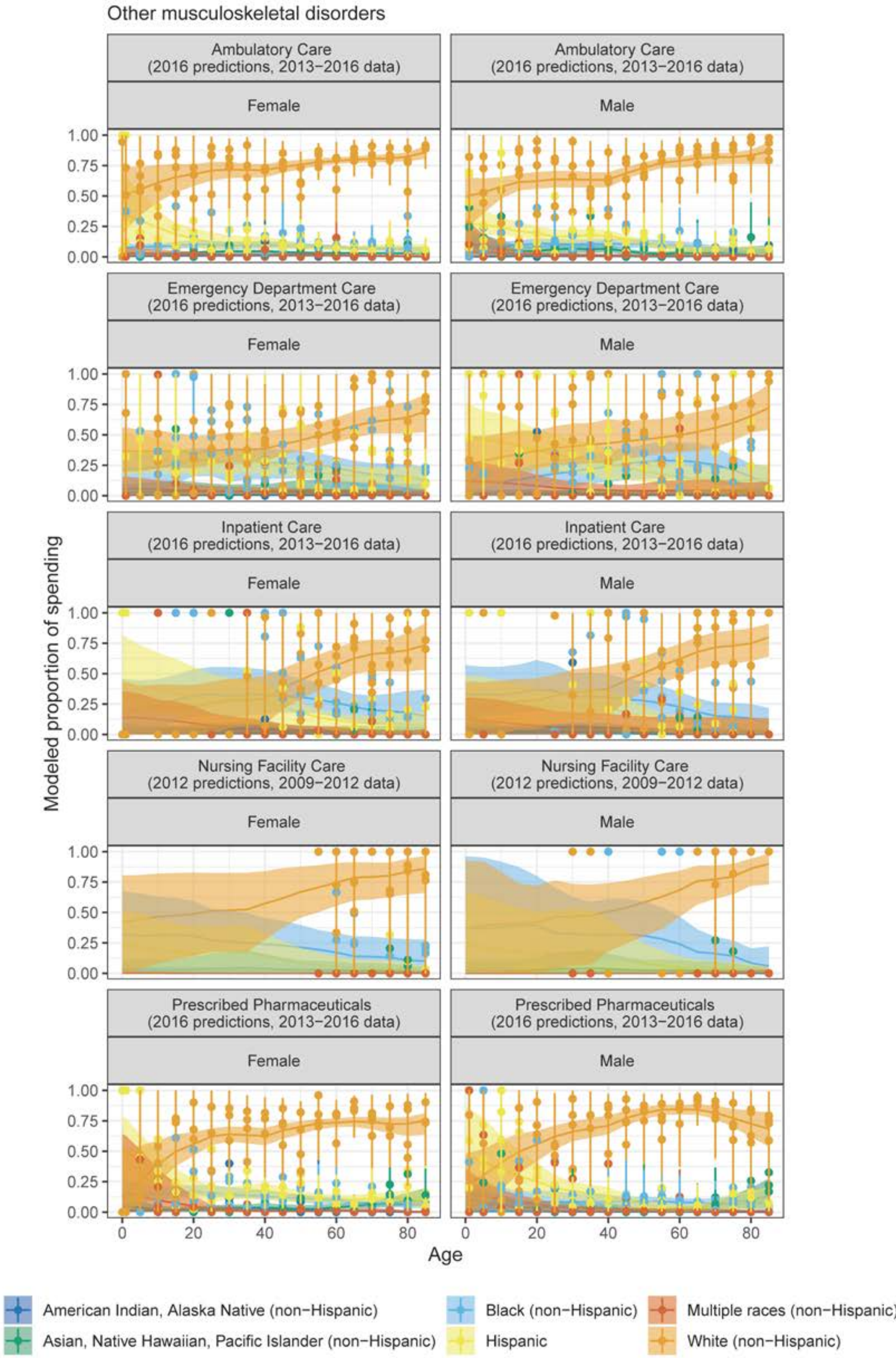
eFigure 3: Spending models



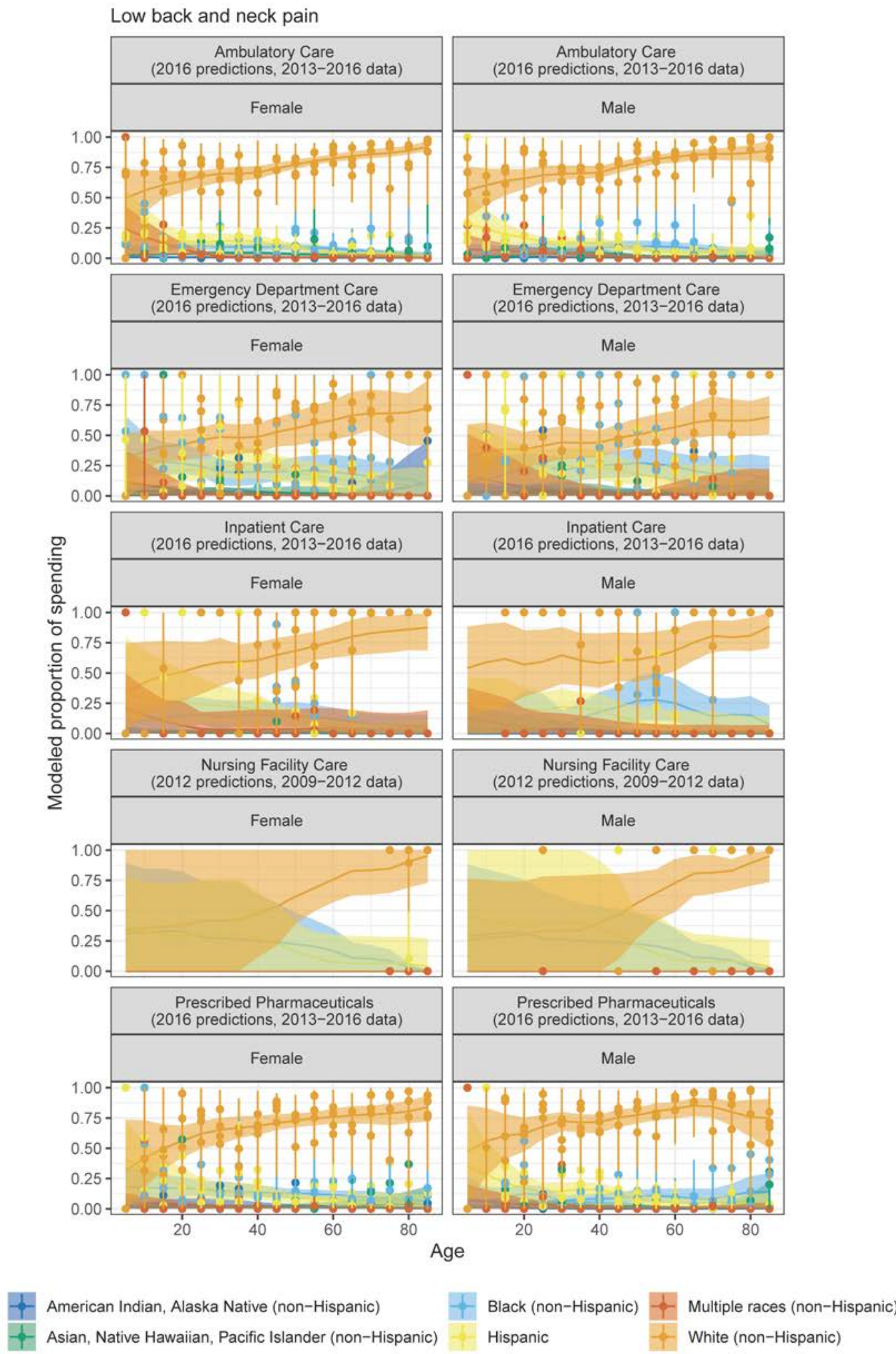
eFigure 3: Spending models



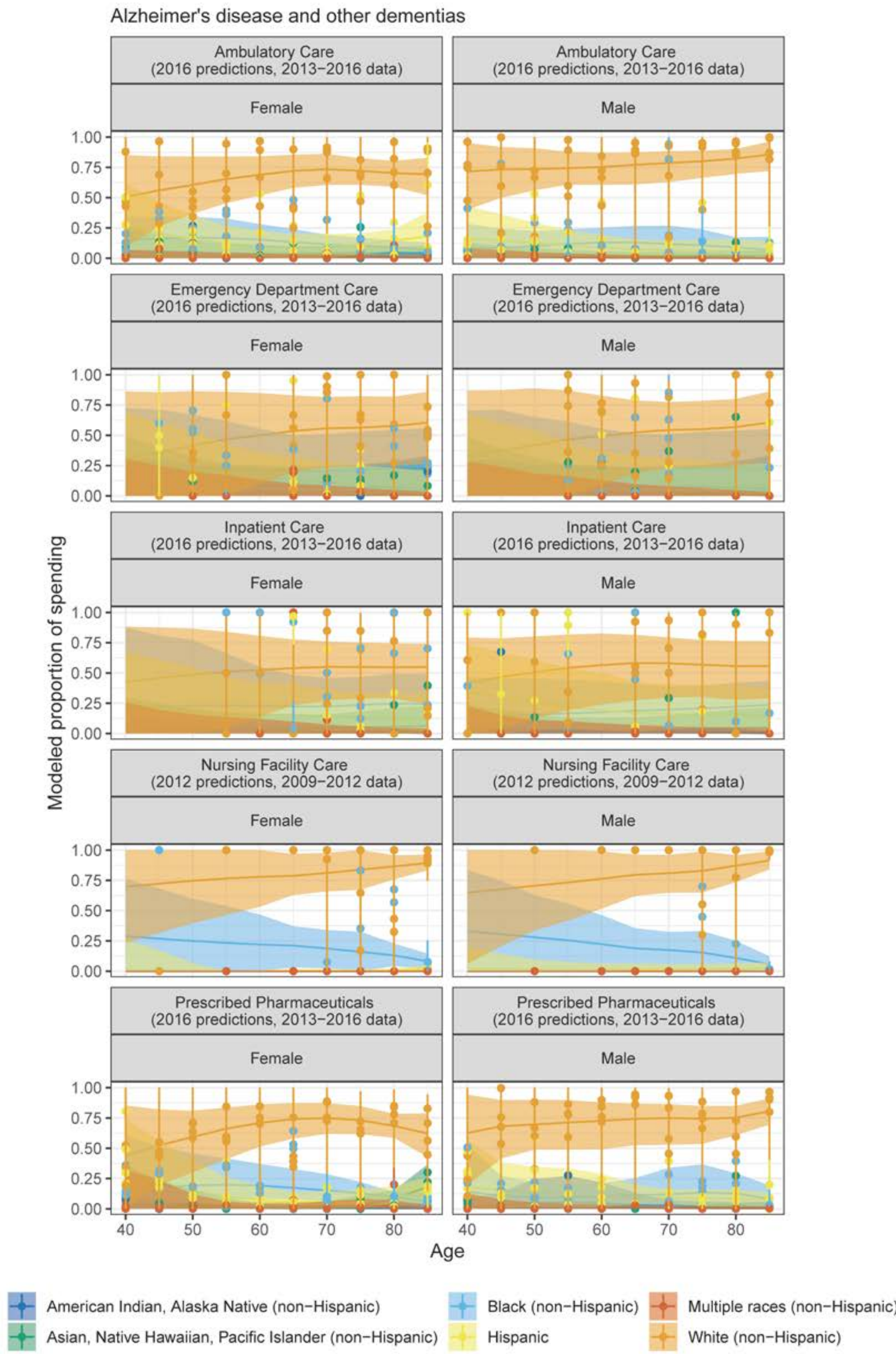
eFigure 3: Spending models



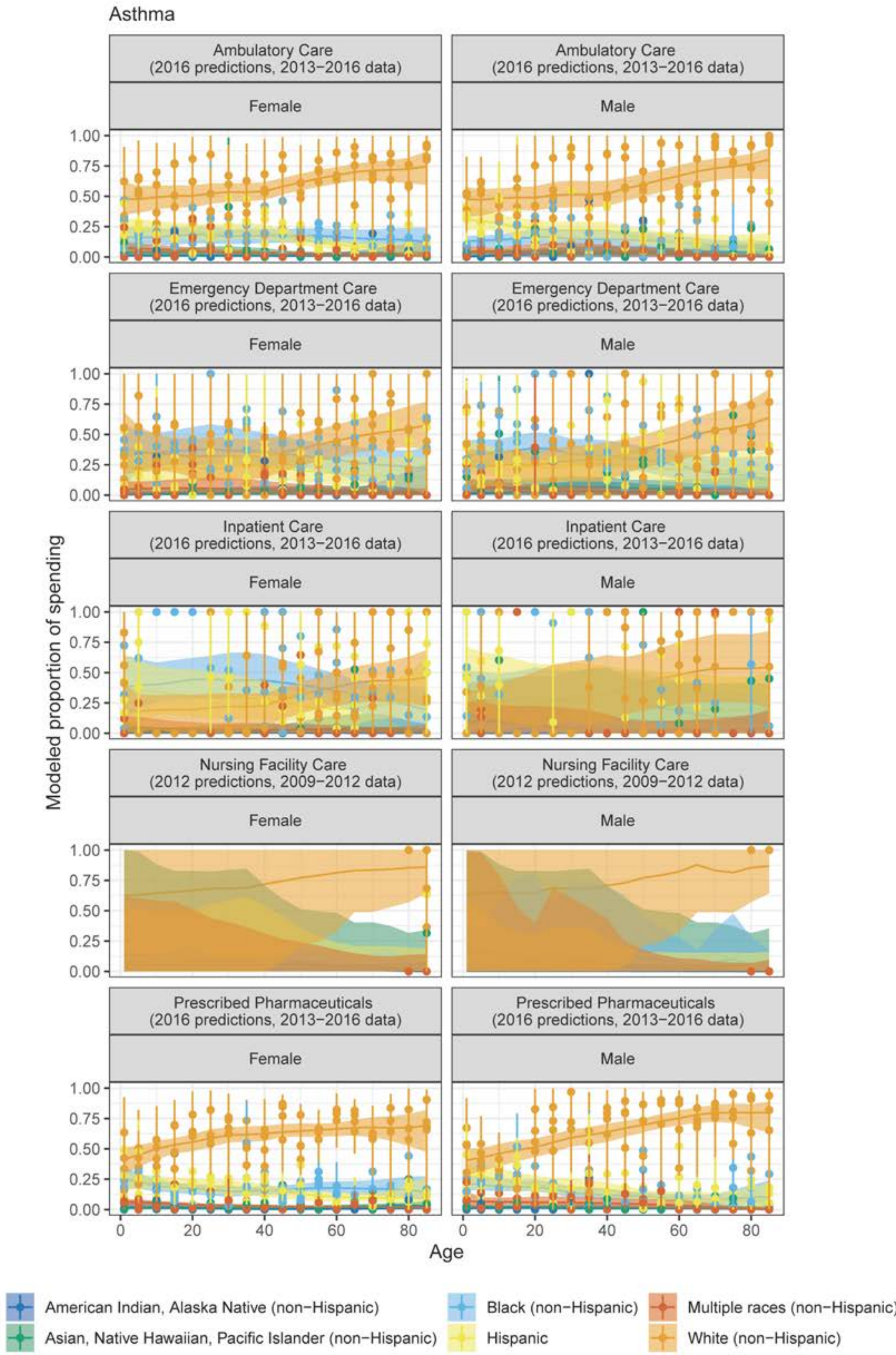
eFigure 3: Spending models



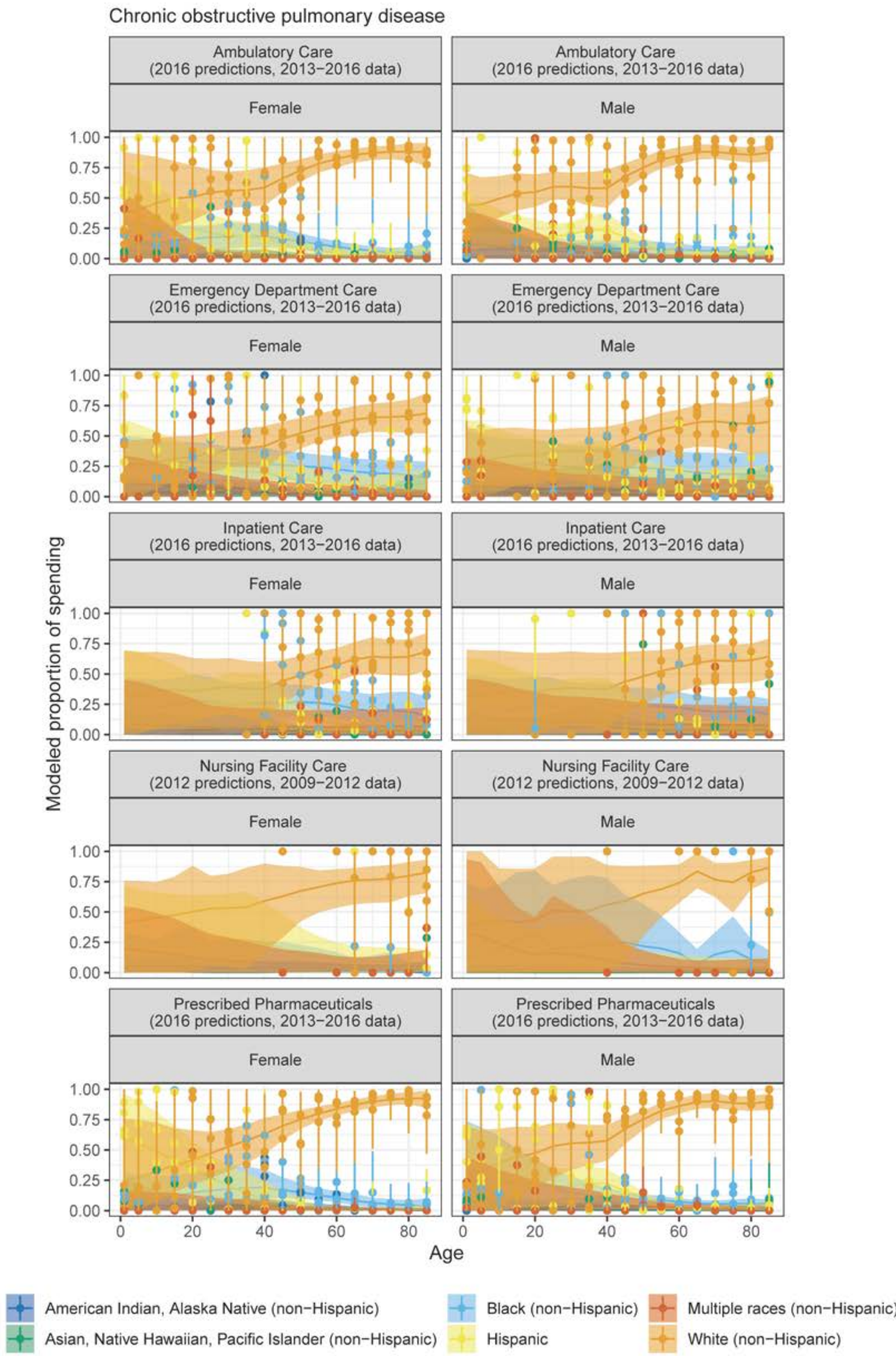
eFigure 3: Spending models



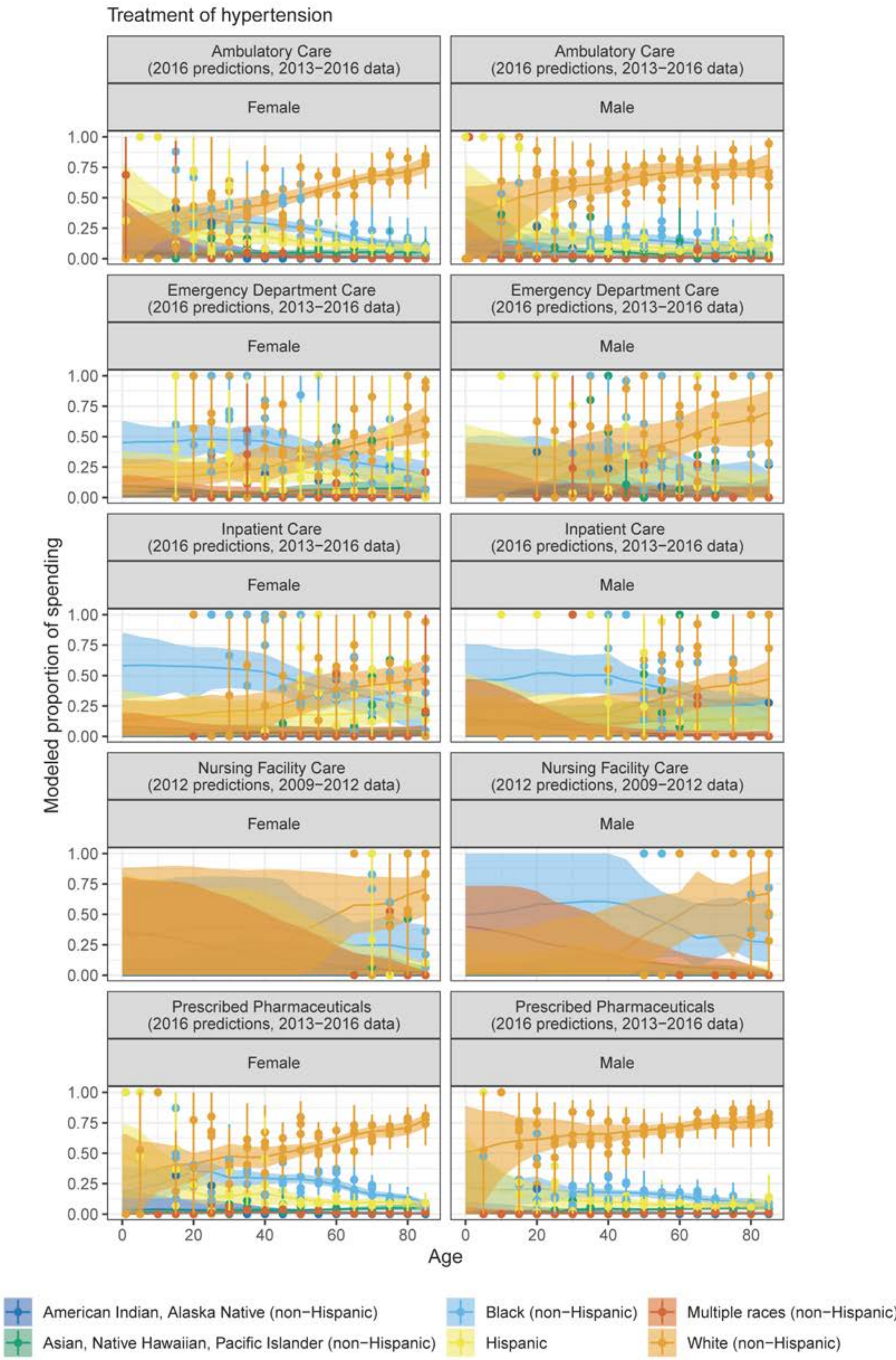
eFigure 3: Spending models



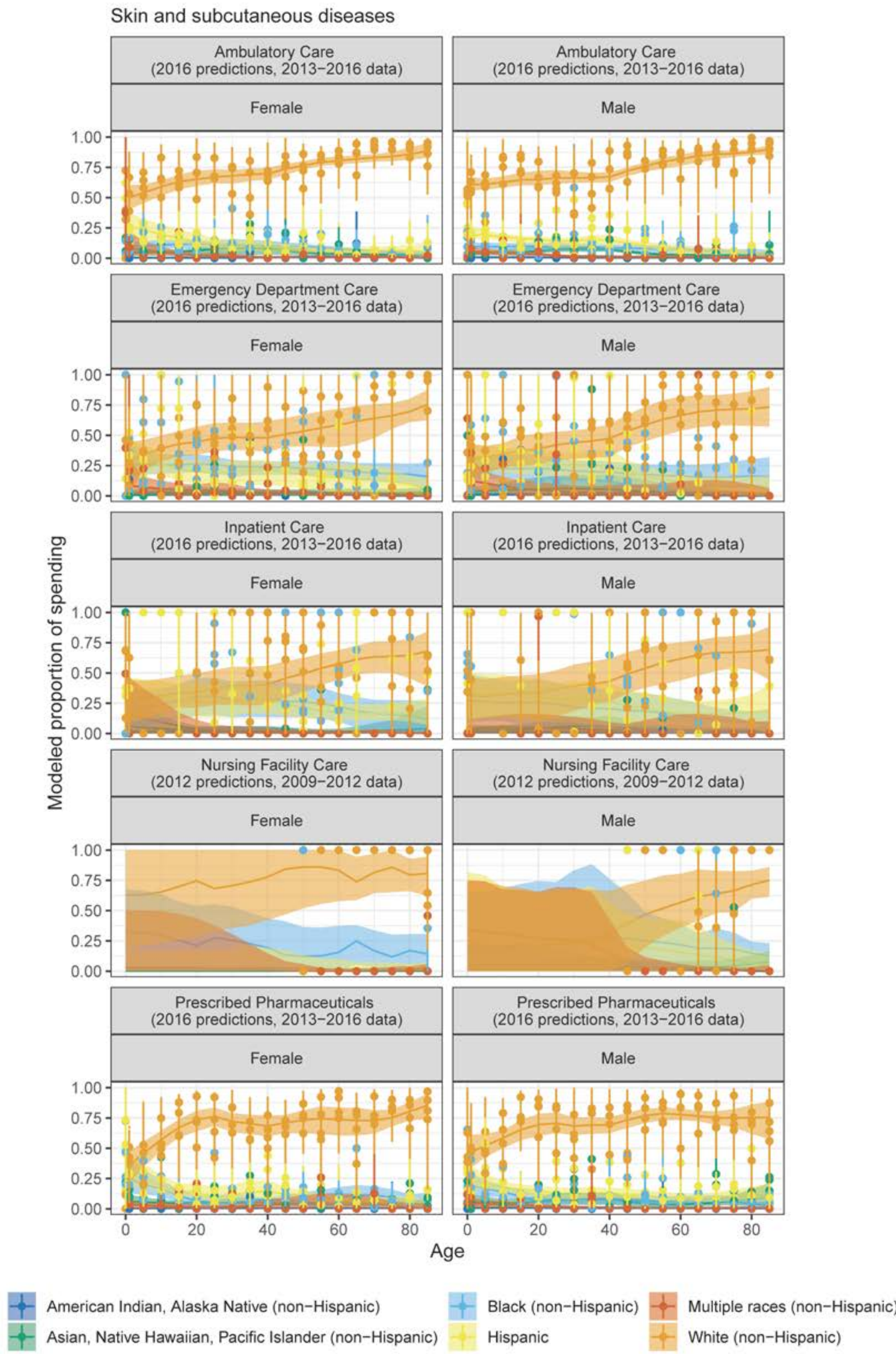
eFigure 3: Spending models



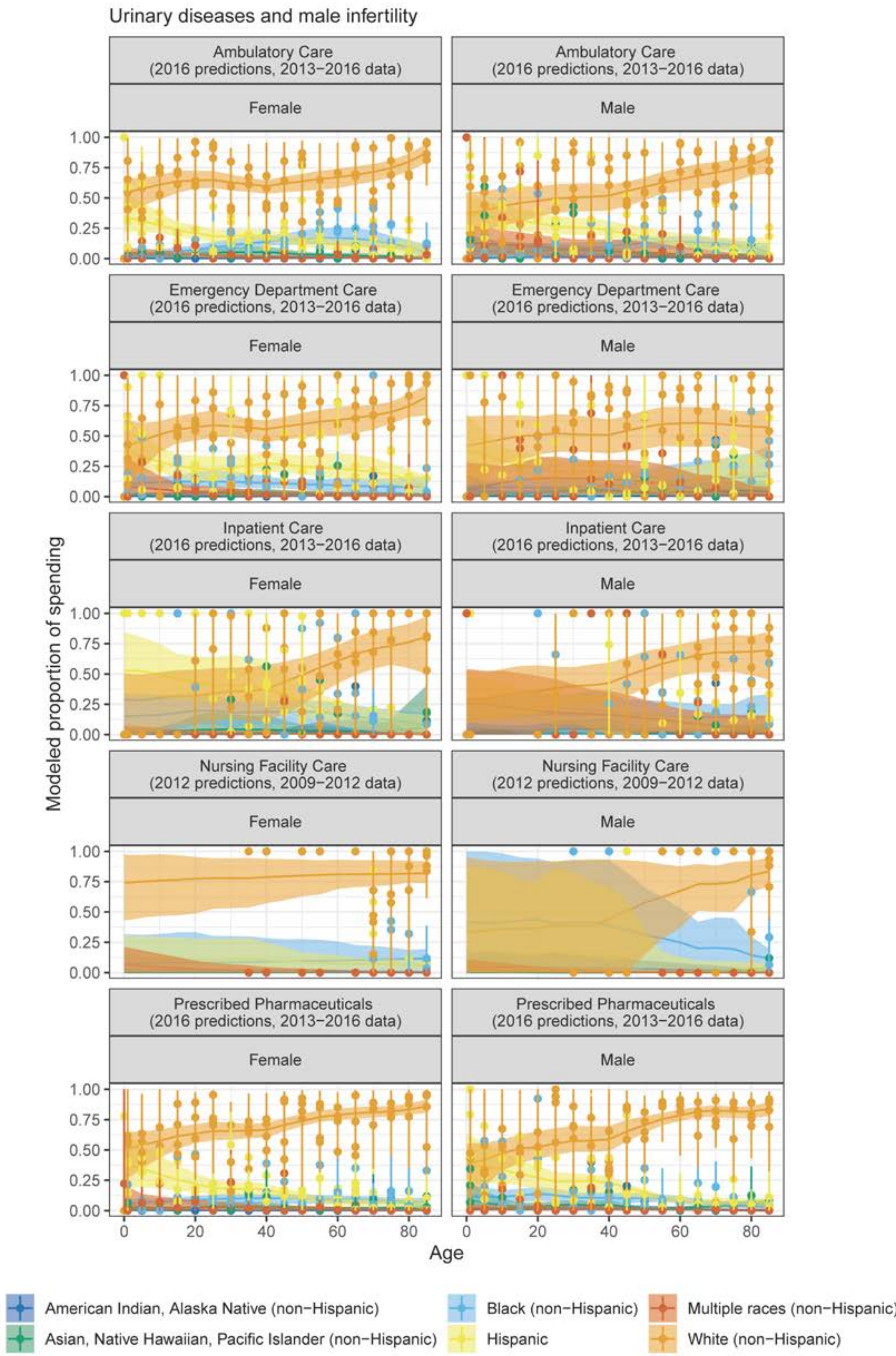
eFigure 3: Spending models



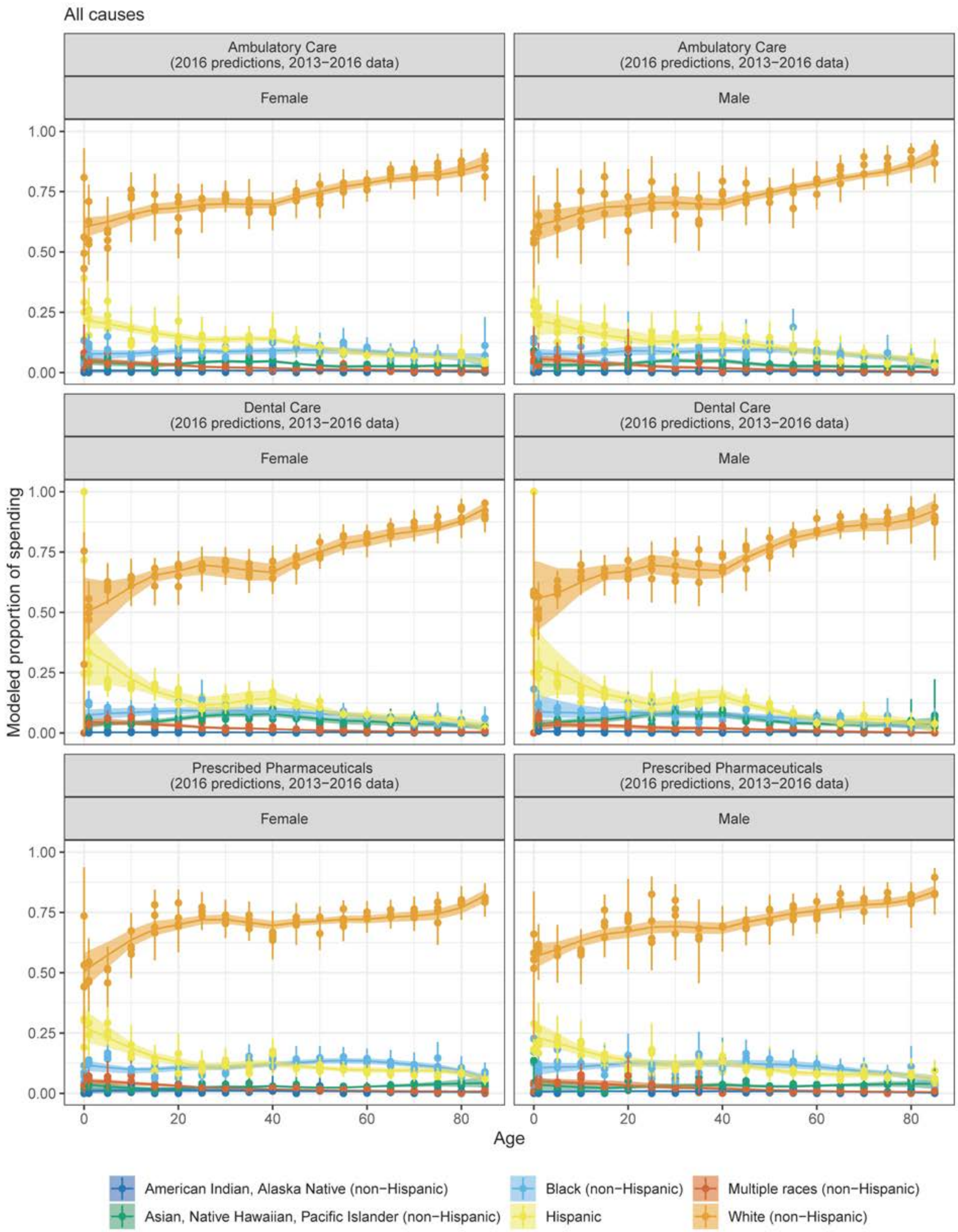
eFigure 3: Spending Models



eFigure 3: Spending Models

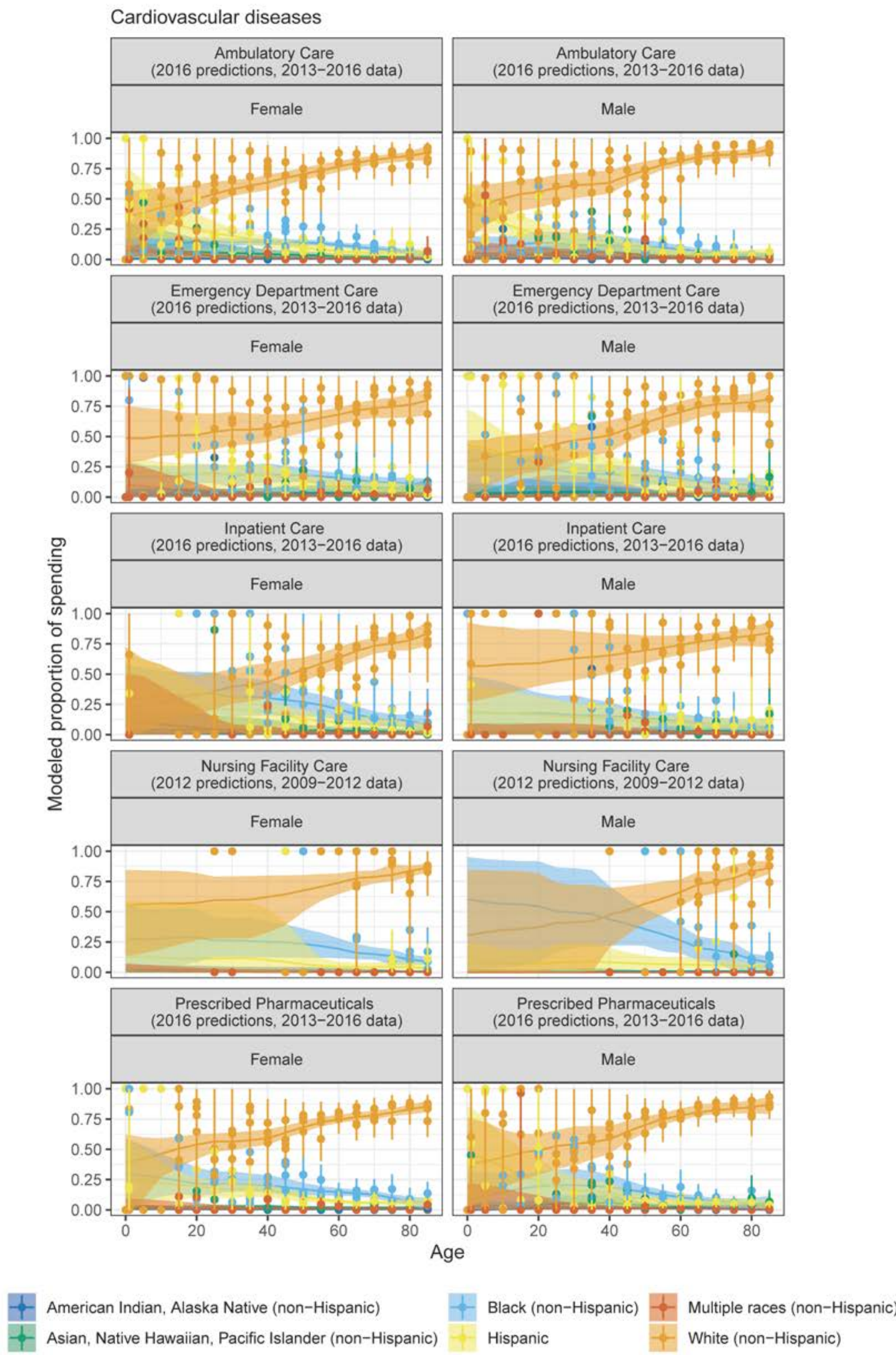


eFigure 4: Volume Models

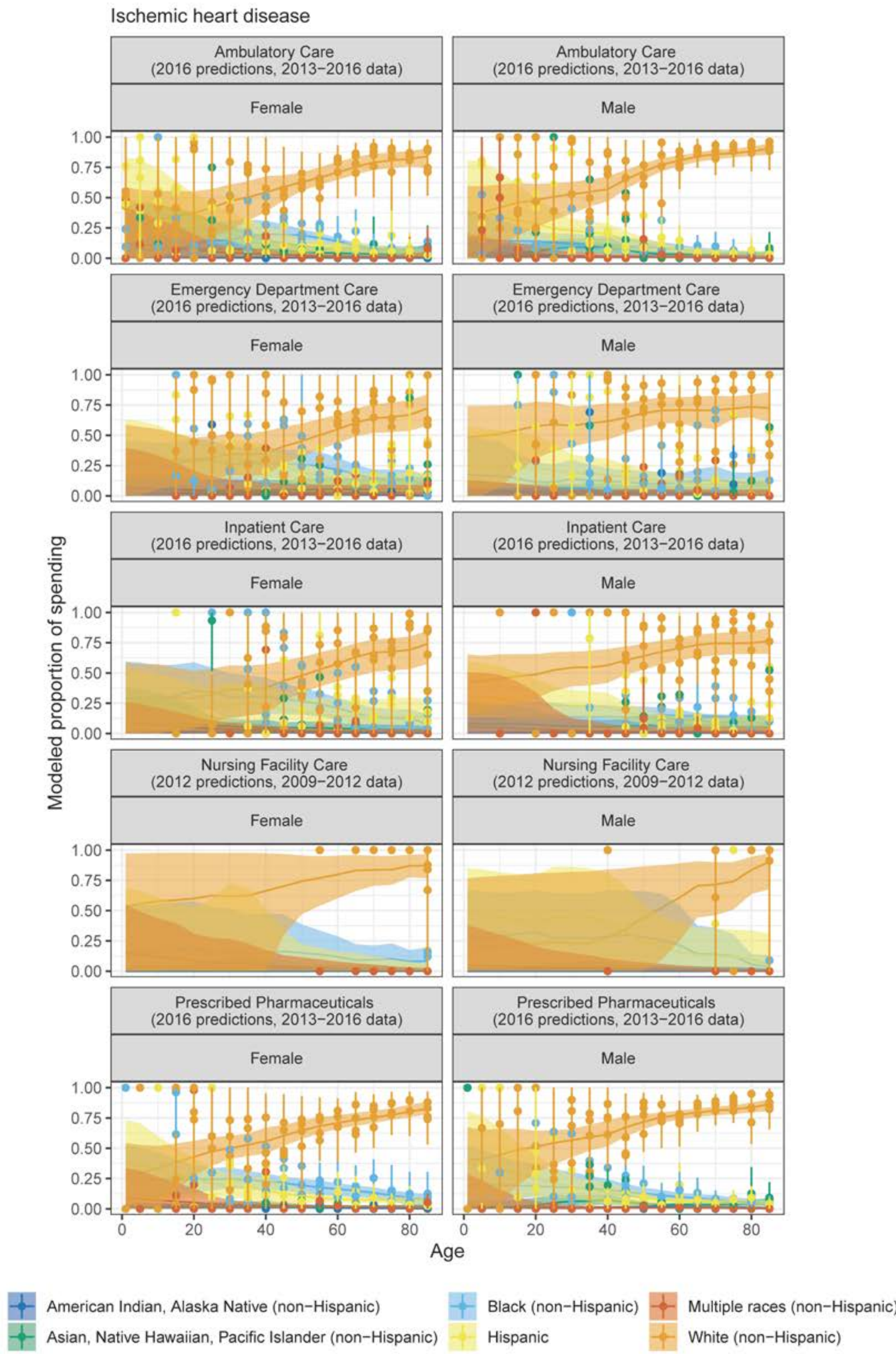


eFigure 4: Volume Models

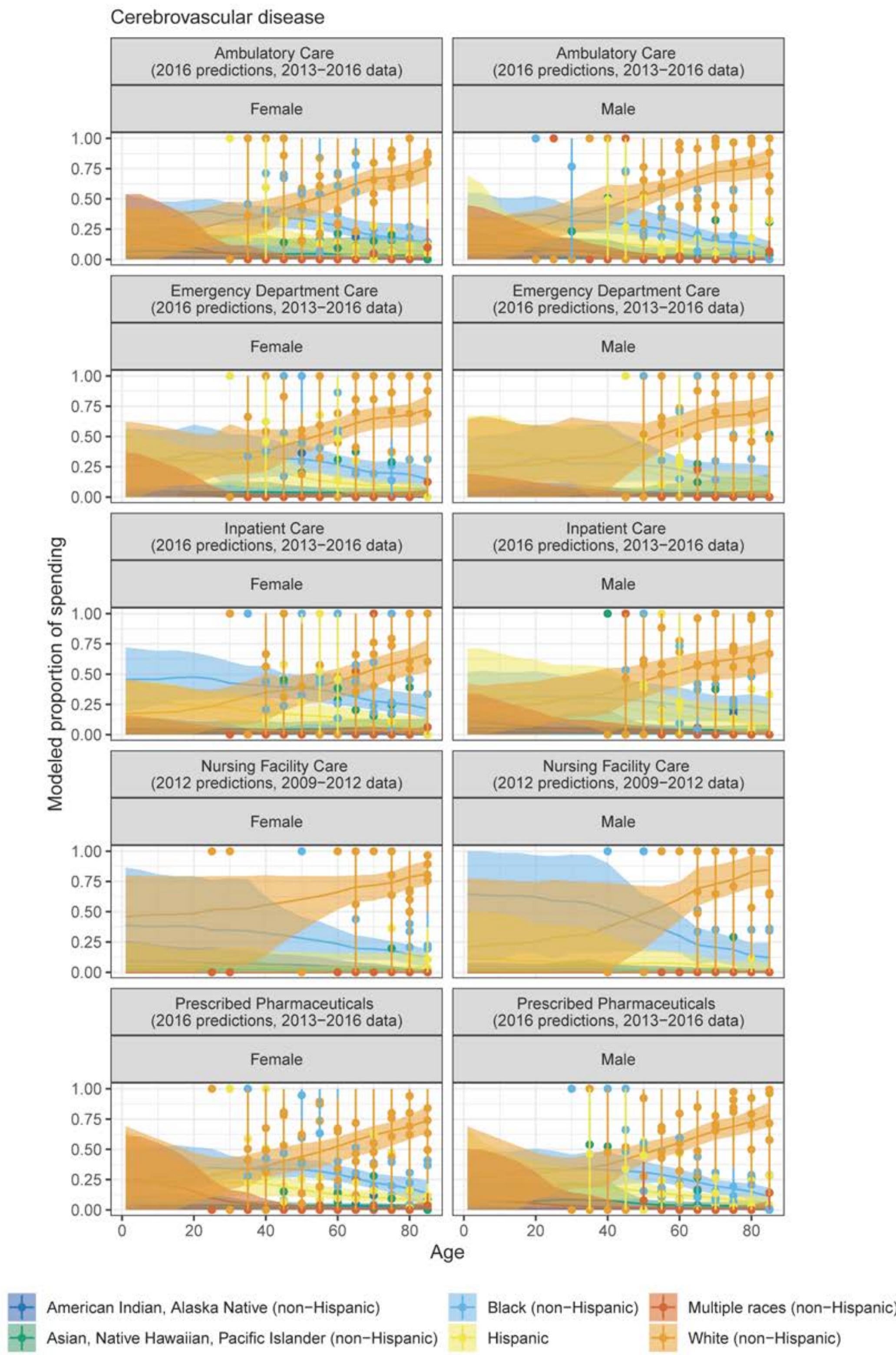
eFigure 4: Volume Models



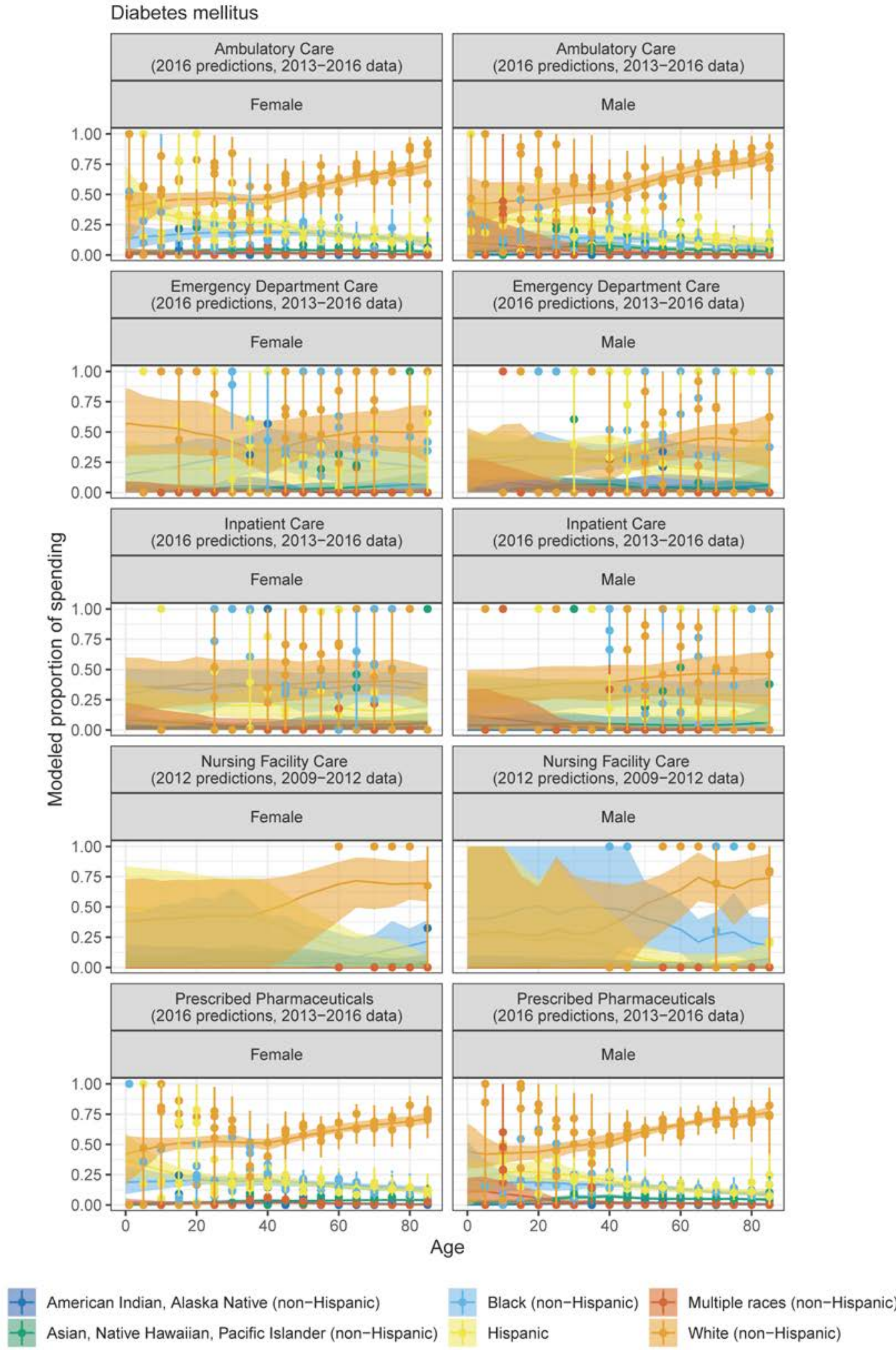
eFigure 4: Volume Models



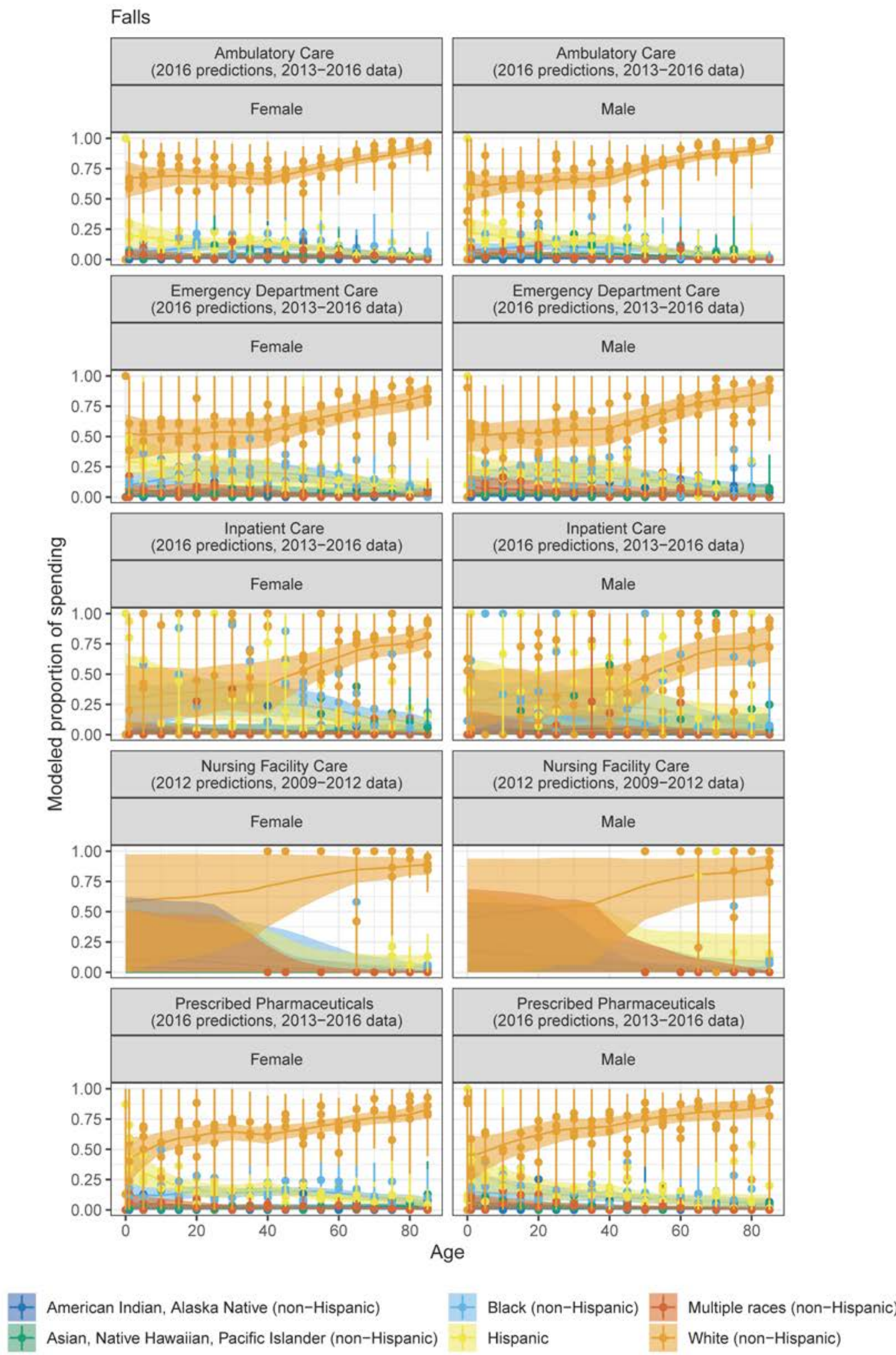
eFigure 4: Volume Models



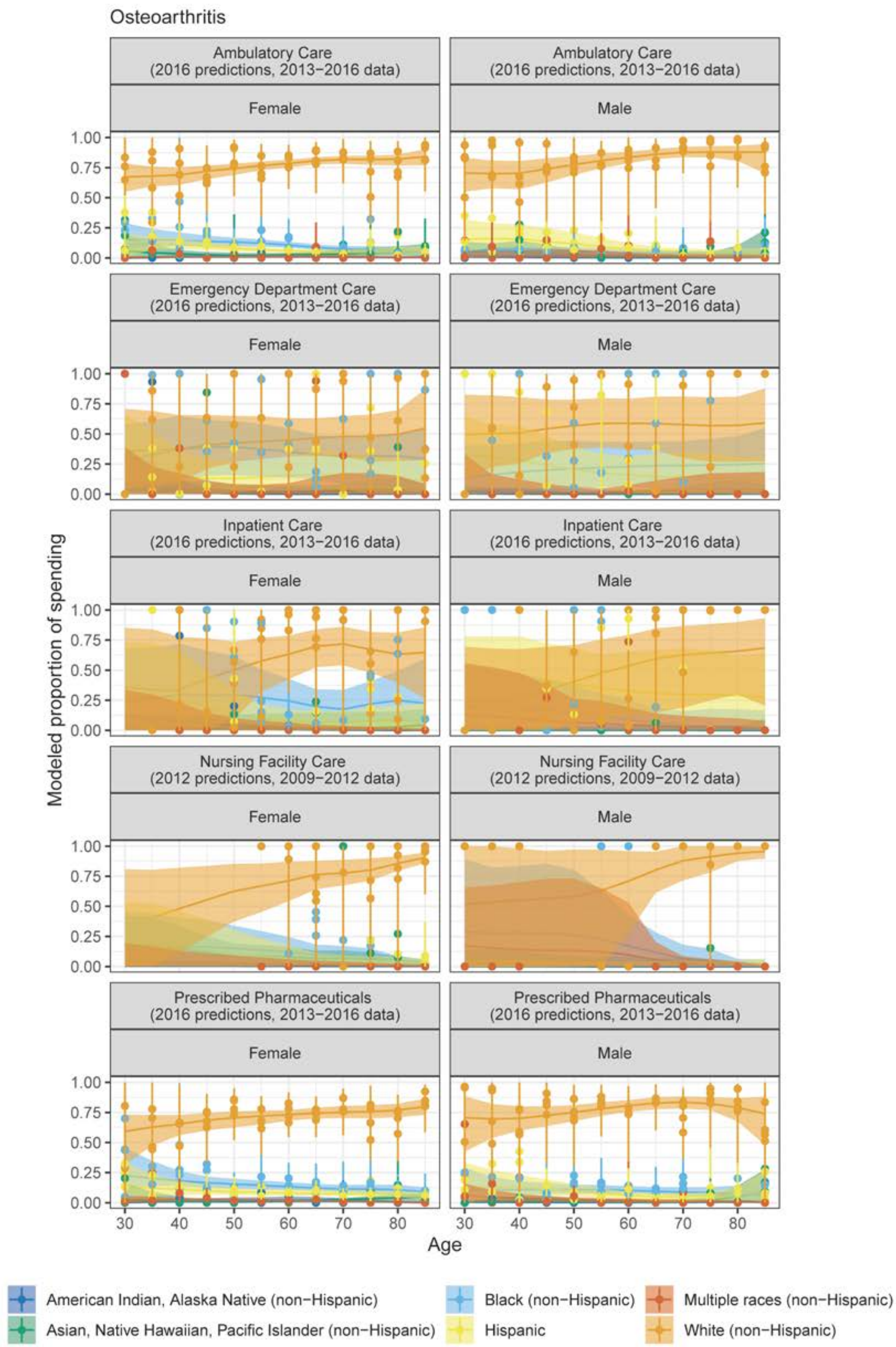
eFigure 4: Volume Models



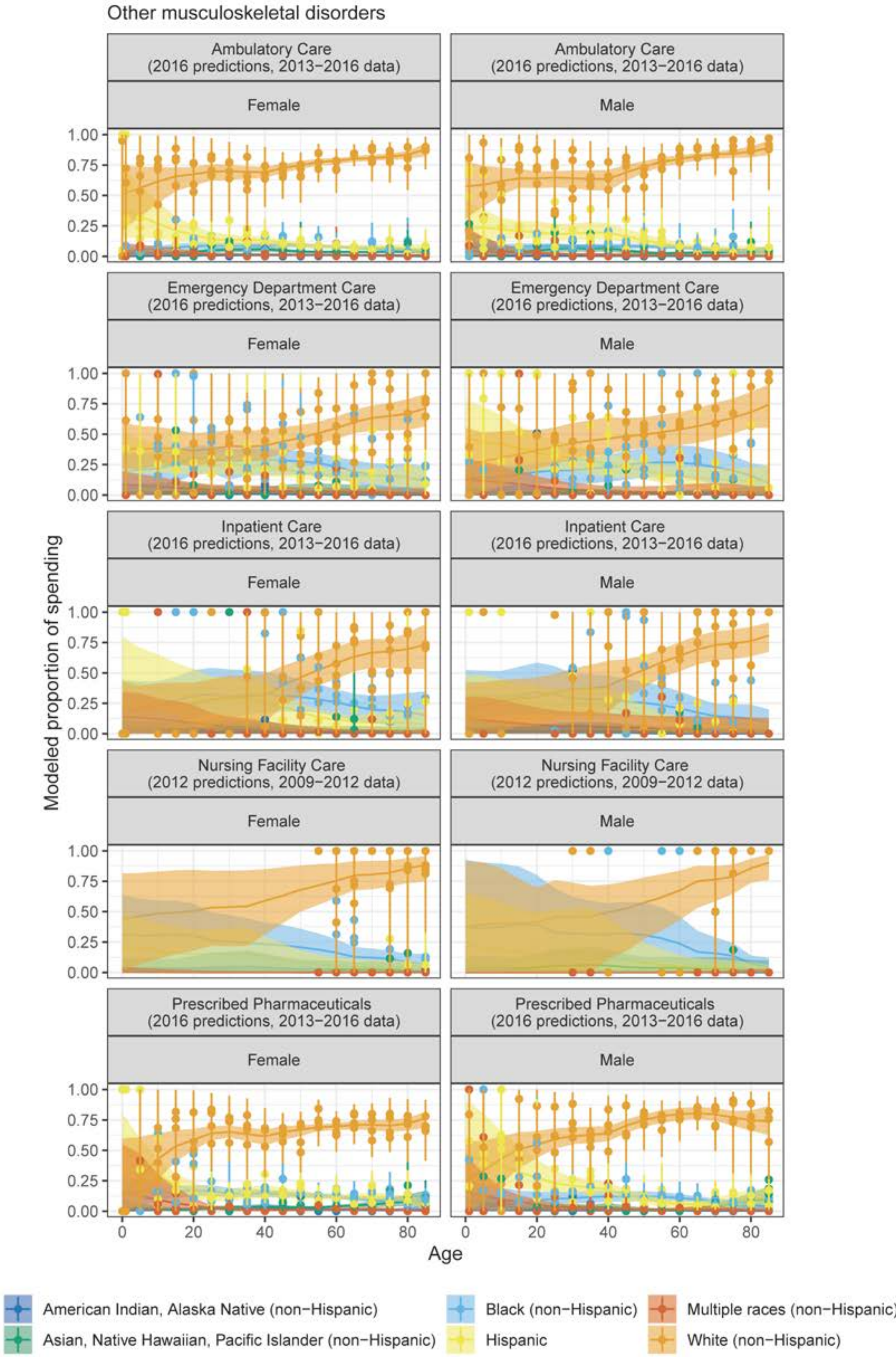
eFigure 4: Volume Models



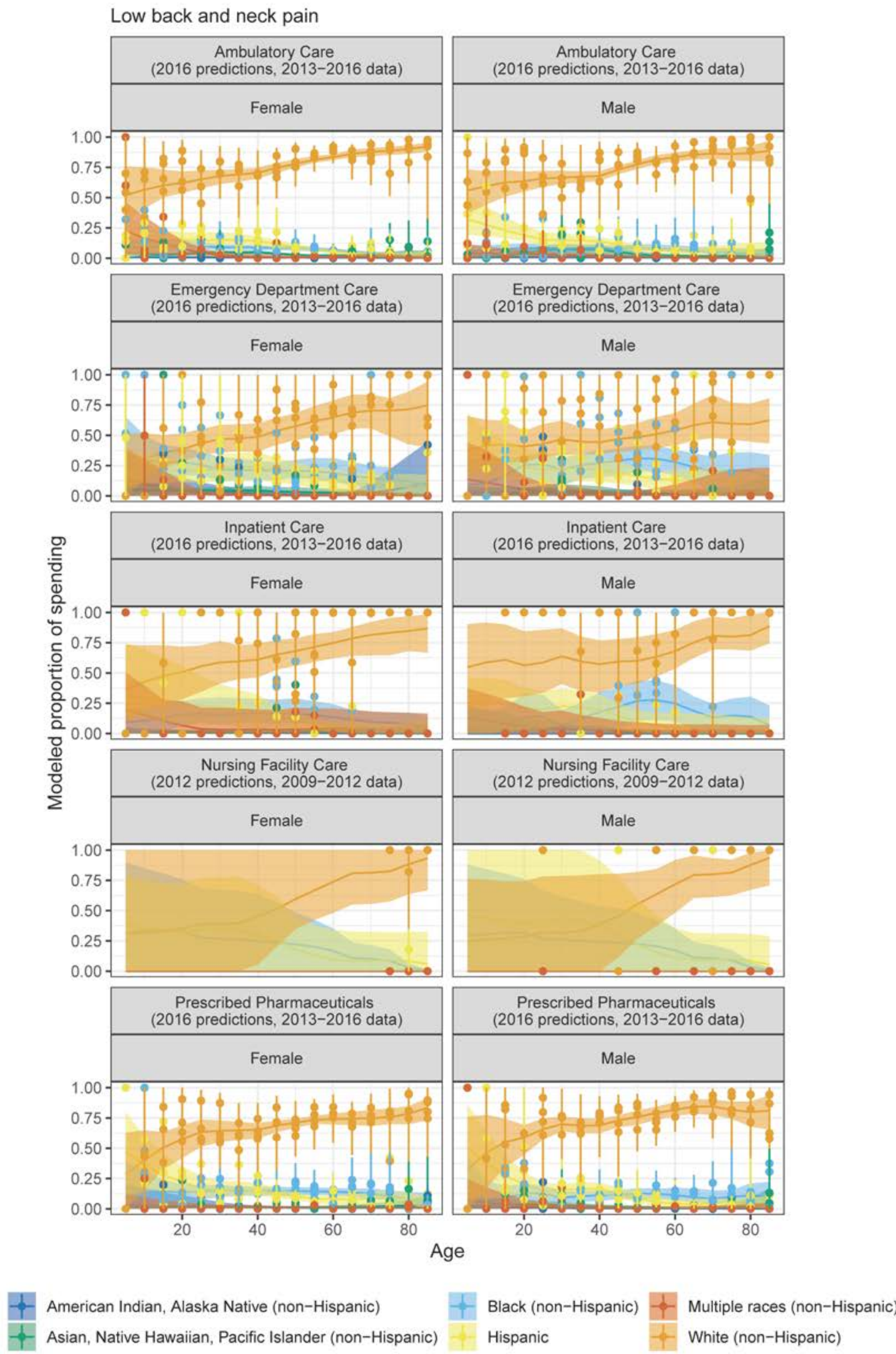
eFigure 4: Volume Models



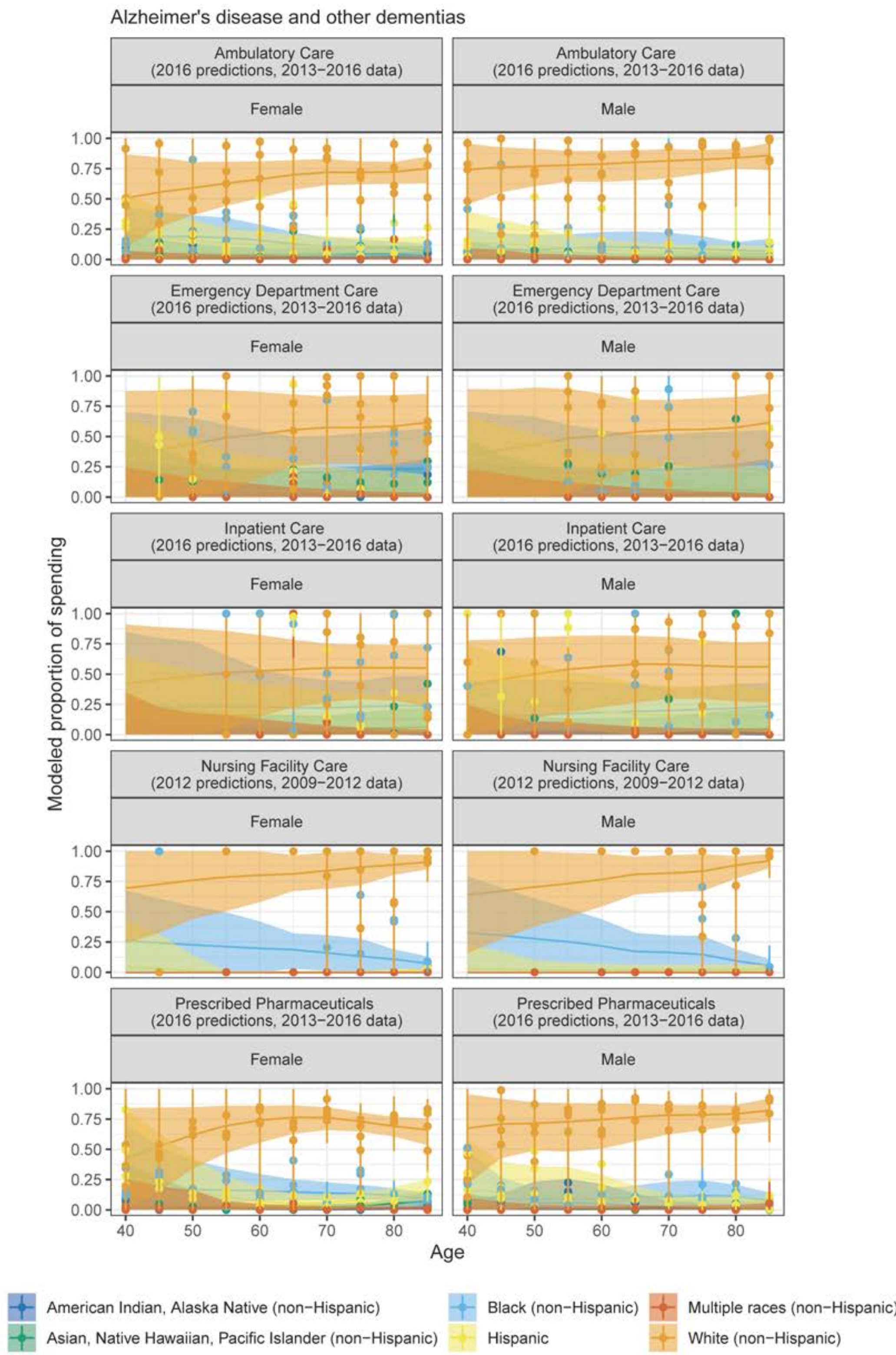
eFigure 4: Volume Models



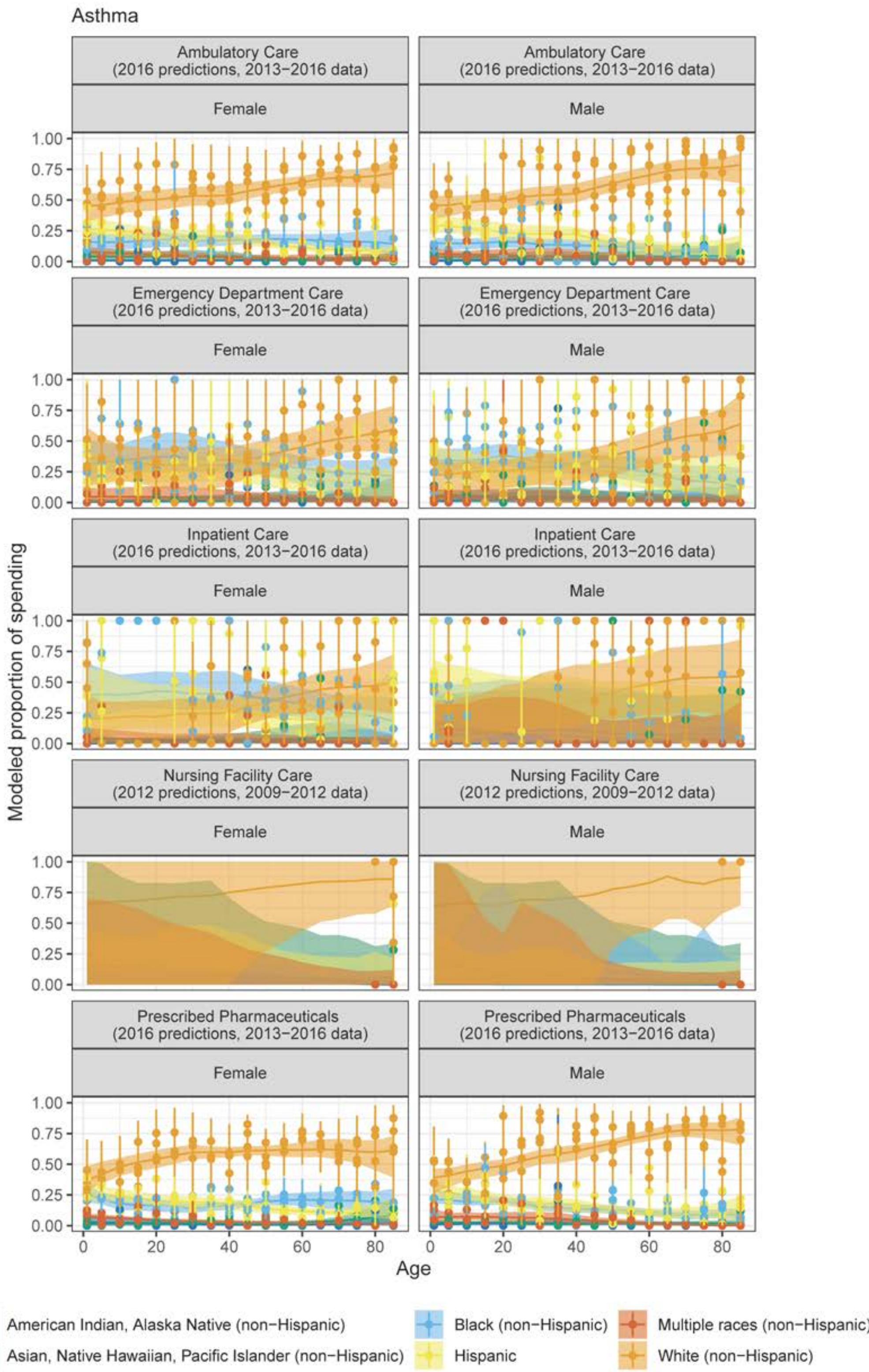
eFigure 4: Volume Models



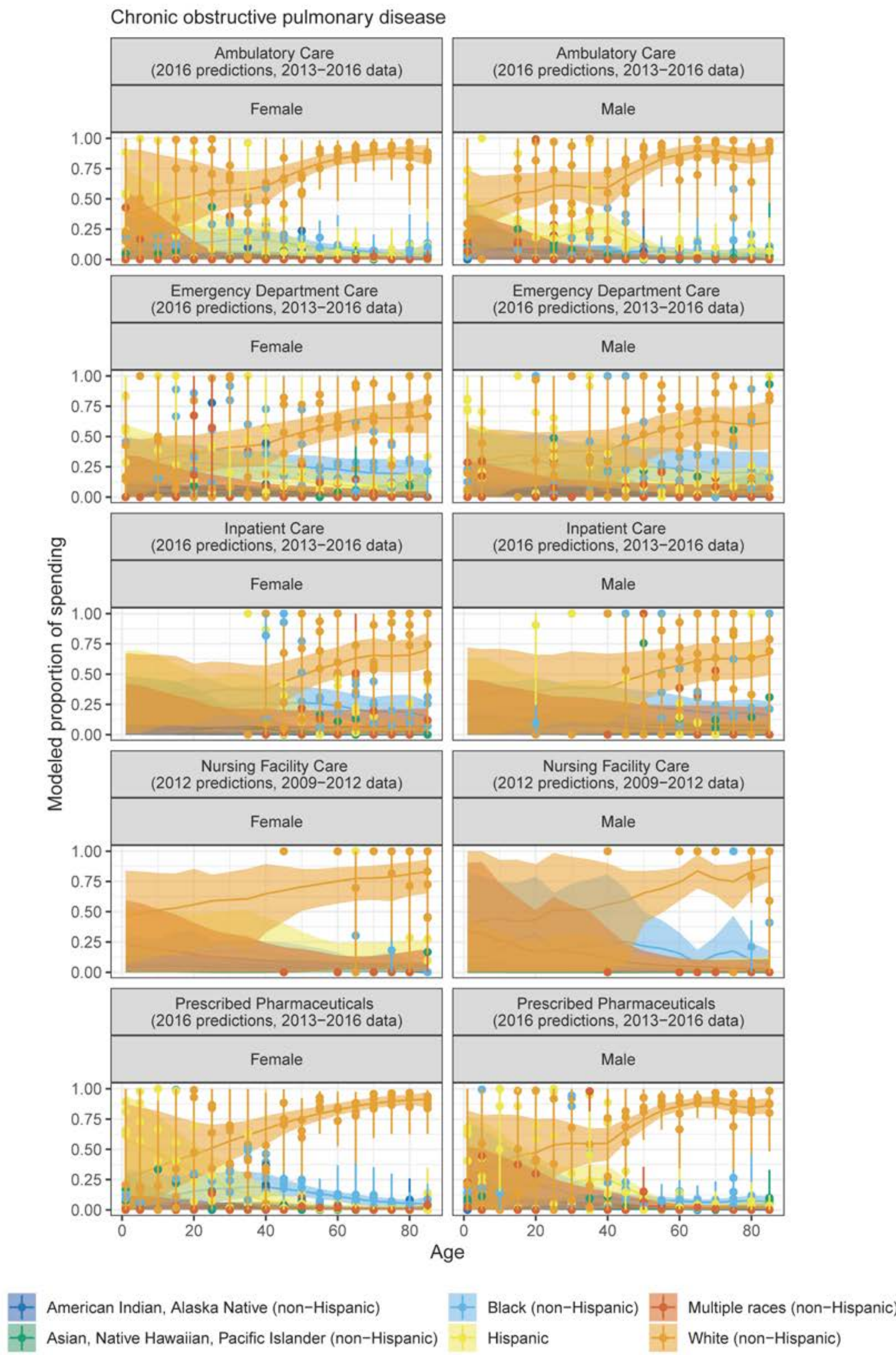
eFigure 4: Volume Models



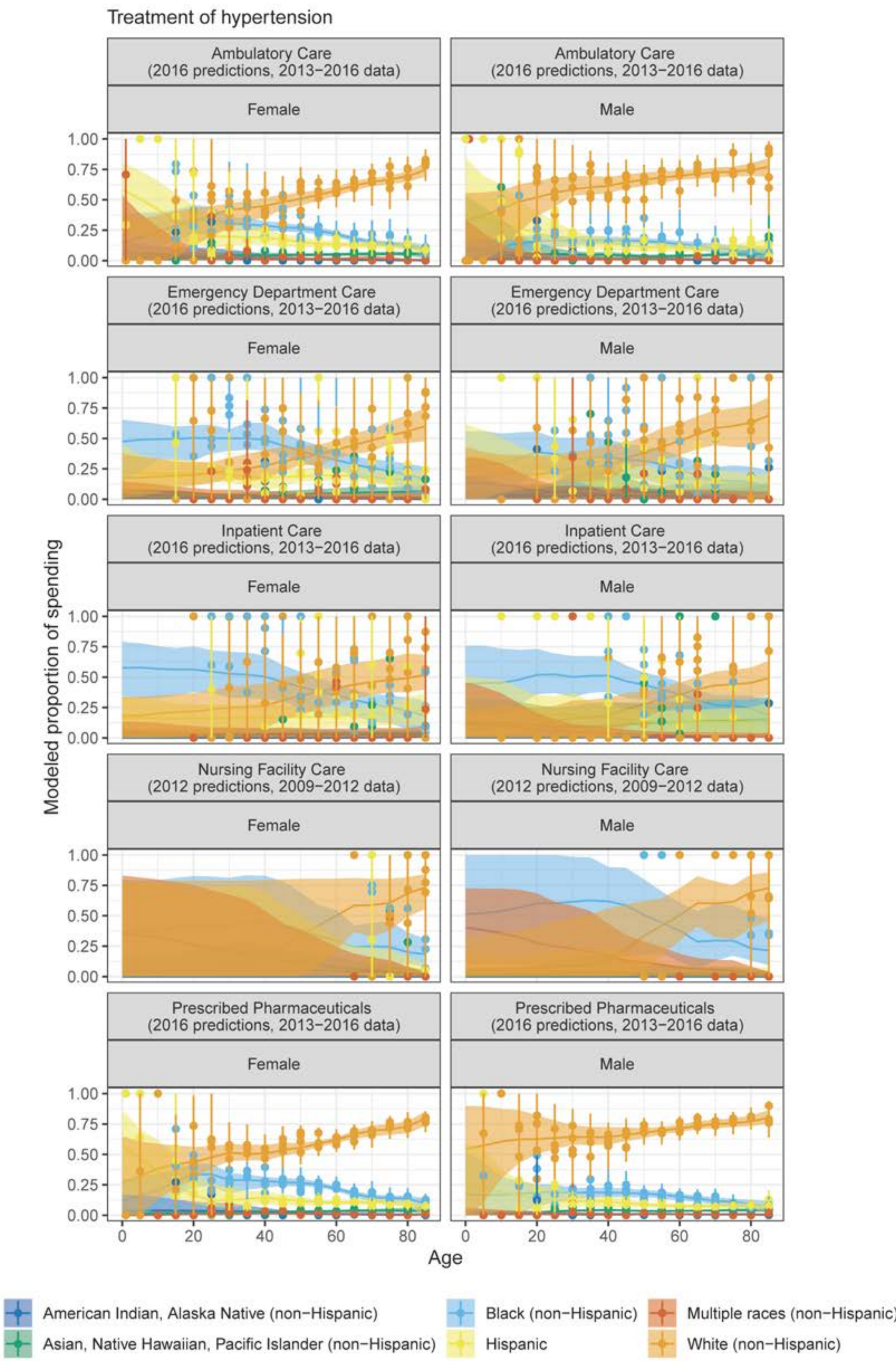
eFigure 4: Volume Models



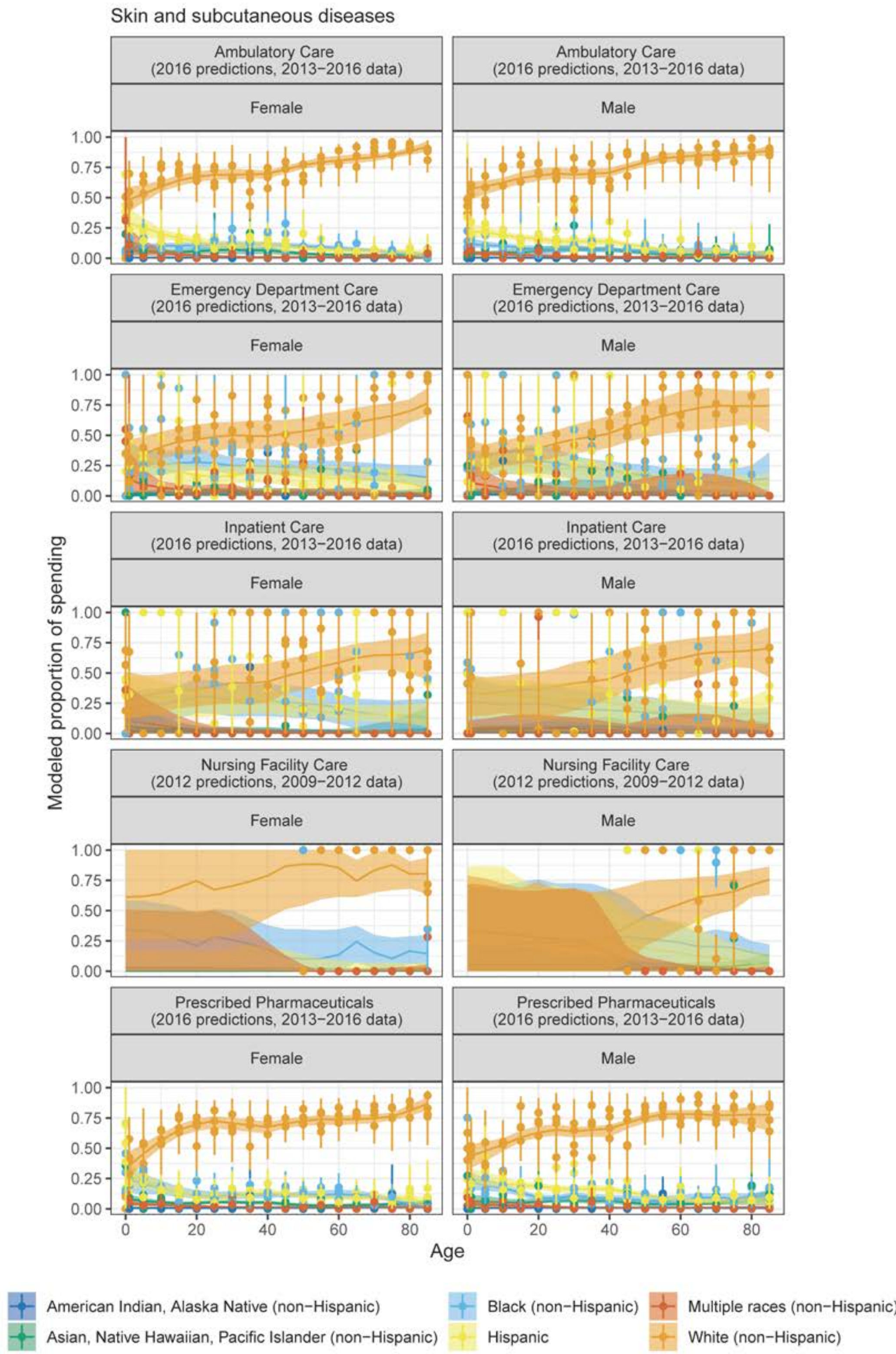
eFigure 4: Volume Models



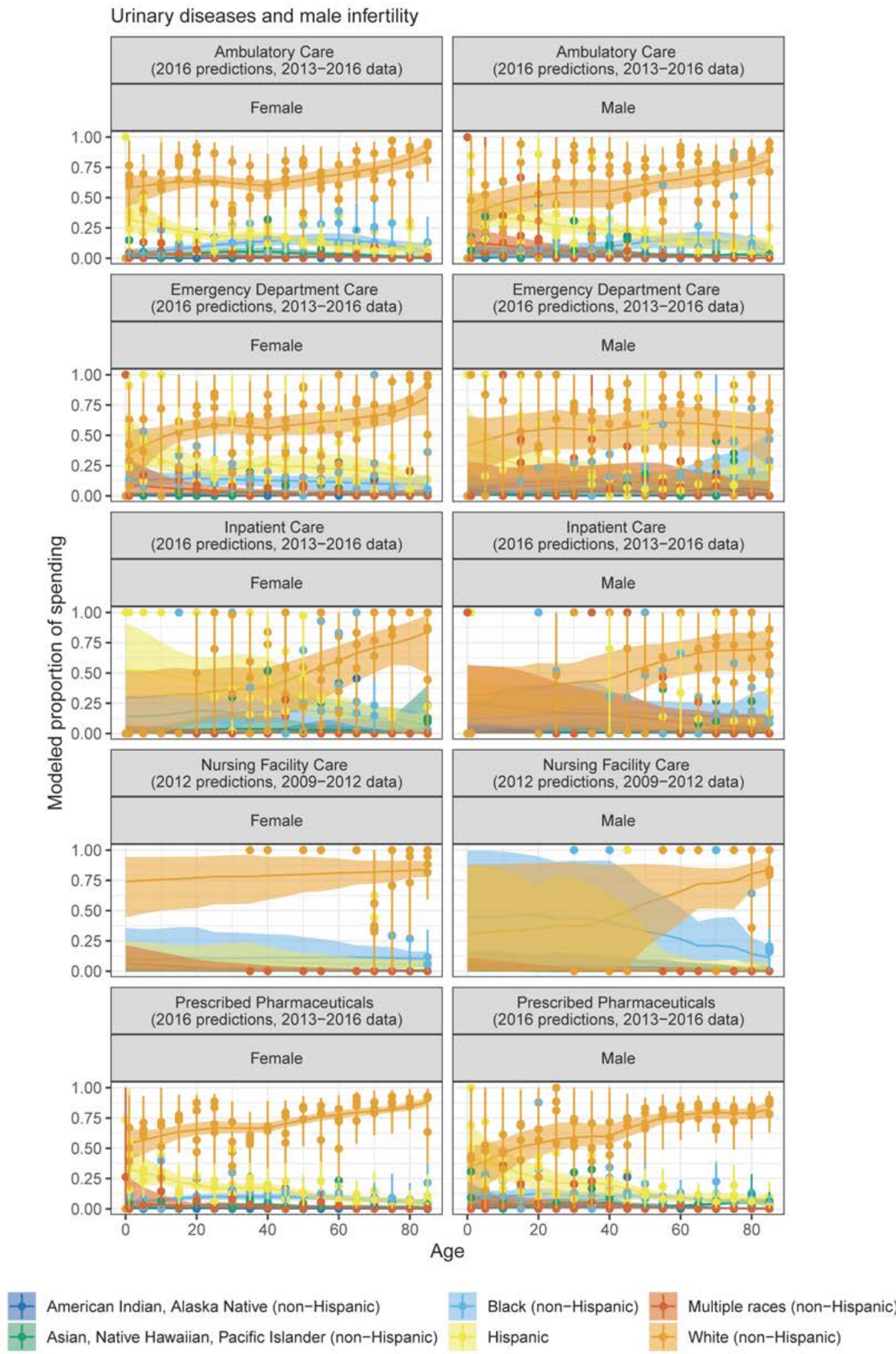
eFigure 4: Volume Models



eFigure 4: Volume Models



eFigure 4: Volume Models



S8. Scaling to the total envelope of spending

In order to convert the modeled proportions of spending for each race/ethnicity to estimates of total spending and volume, we multiplied these proportions by the DEX estimates of total spending or volume estimates for each health condition (or aggregate health condition), year, age, sex, and type of care by race. This is equivalent to proportionally splitting the spending the total spending for each of the above categories into six race/ethnicity groups. This calculation, which was performed at the draw level, is illustrated in the equations below.

$$Spending_{[r]} = DEX\ spending_{[t]} \times\ spending\ modeled\ proportions_{[r]}$$

$$Volume_{[r]} = DEX\ volume_{[t]} \times\ volume\ modeled\ proportions_{[r]}$$

$[r]$: $year\text{-}, age\text{-}, sex\text{-}, type\text{-}, and\ health\ condition\text{-}, race\ and\ ethnicity\text{-}specific$
$[t]$: $year\text{-}, age\text{-}, sex\text{-}, type\text{-}, and\ health\ condition\text{-}\ specific$

This calculation allowed us to generate specific estimates of the race/ethnicity-specific spending or volume for each year, age, sex, type of care, and cause category, along with uncertainty estimates.

S9. Age-standardization

Due to the varying age profiles of each race/ethnicity groups' population, it is important to standardize per person spending and volume estimates to a reference population such that the estimates can be compared when aggregated across age. We used direct age-standardization to standardize all race/ethnicity categories to the 2016 all-race population age profile. In order to age-standardize, the following steps were performed: (1) Using race- and age-specific populations, we generated age-specific estimates of spending and volume per person for each race/ethnicity. (2) Using the all-race 2016 reference population, we generated population weights by calculating the proportion of the total population in each age group. (3) For each race/ethnicity, we took a weighted mean across the age-specific per person spending and volume estimates (generated in step 1) using the population weights generated in step 2. This calculation is illustrated below. We performed age-standardization for each health condition, type of care, and year of spending. In addition, age-standardized estimates were calculated at the draw level to propagate uncertainty.

$$Age\ std.spend\ per\ capita_{[r]} = \sum\nolimits_{[a]} spend\ per\ capita_{[r,a]} * \ 2016\ pop\ weight_{a}$$

$[r]:\ race\text{-}\ and\ ethnicity\text{-}specific$
$[a]:\ age\text{-}specific$

S10. Decomposition

For each type of care and race/ethnicity group, healthcare spending per person is a product of two factors: (i) health care utilization (i.e., utilization rates, which are visits per person, admissions per person, or prescriptions per person) and (ii) price and intensity of care (spending per visit, spending per admission, or spending per prescription). In an effort to better understand how these two factors impacted spending for each race/ethnicity group, we ran a Das Gupta decomposition to isolate their effects. Das Gupta decomposition is common in demographic research to assess the relative effects associated with fundamental drivers that collectively lead to change in an underlying rate or count. Specifically, we decomposed the difference between each race/ethnicity group's spending per person and the all-population mean spending per person to isolate (a) the differences in spending attributable to differences in utilization rates and (b) differences in spending attributable to differences in the price and intensity of care. This calculation is illustrated in the equations below. Because of data limitations, differences in spending attributable to differences in price and intensity were combined and not measured separately.

$$Effect_{[U]} = \frac{U_{[r]} + U_{[m]}}{2}(PI_{[r]} - PI_{[m]})$$

$$Effect_{[PI]} = \frac{PI_{[r]} + PI_{[m]}}{2}(U_{[r]} - U_{[m]})$$

U : utilization
PI : price and intensity
$[r]$: *a given race/ethnicity*
$[m]$: *the all-race population mean*

Decomposition was run for each type of care, and calculations were performed at the draw level to propagate uncertainty. Finally, we normalized the outputs so they could be expressed as a percentage difference from the all-race population's spending.

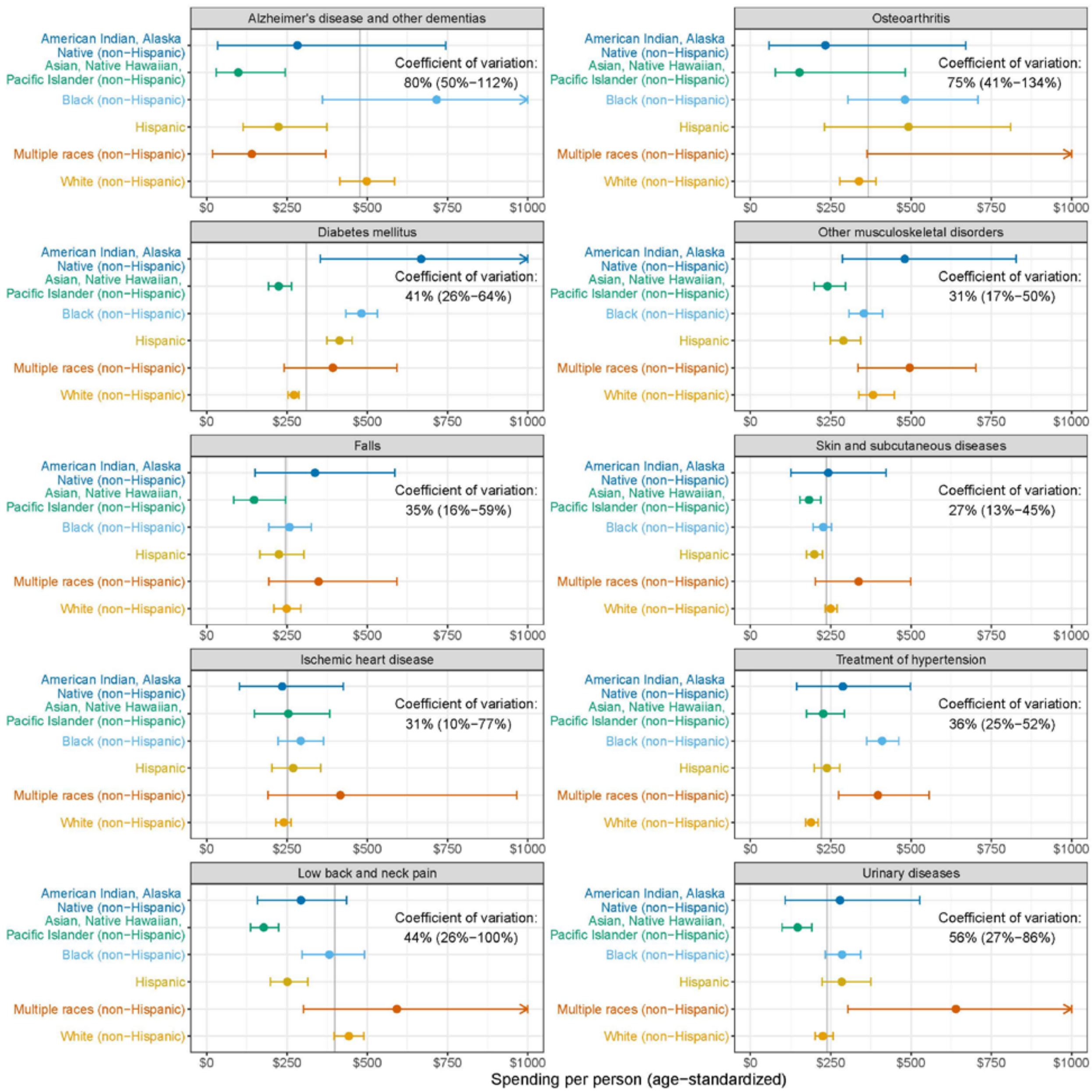
S11. Estimating spending and utilization per notified case

To calculate the number of notified cases for select health conditions, we calculated the weighted proportions of survey respondents who reported they were told by a healthcare provider that they have a health condition for each age, sex, and race/ethnicity group. Then, these proportions were applied to population drawn from the Global Burden of Disease study (GBD) to estimate the total number of notified cases in the United States by age, sex, and race/ethnicity categories in 2016.

The following seven health conditions were chosen because they comprise a large portion of disease burden and healthcare spending and there was a sufficient sample size for analysis: cardiovascular disease, cerebrovascular disease, diabetes, low-back and neck pain, asthma, COPD, and treatment of hypertension. The notified cases were combined with the estimates of spending and volume to calculate per notified case spending and volume specific to each race/ethnicity. The spending per notified case estimates were calculated at the draw level to propagate uncertainty.

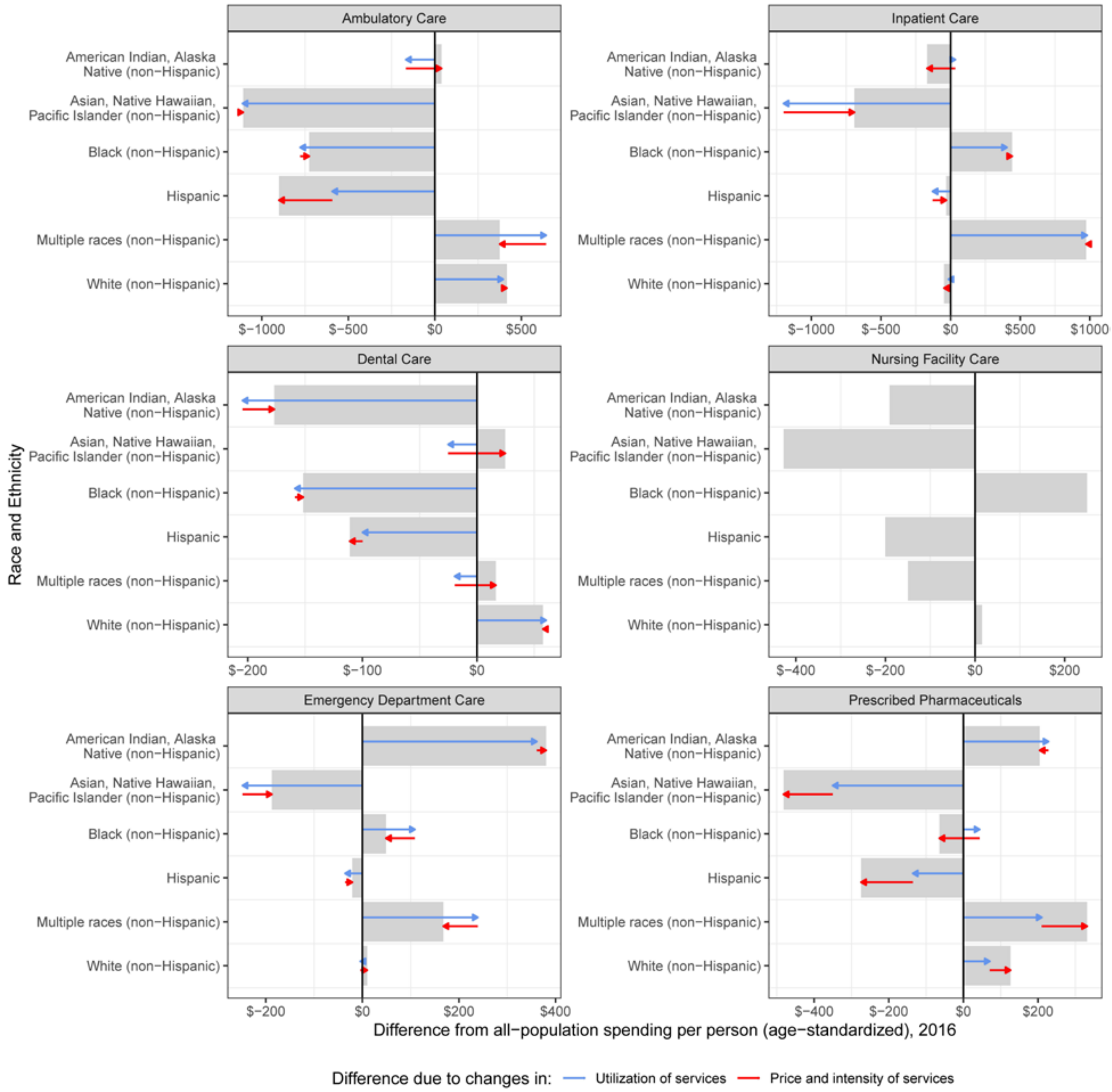
S12. Additional Results

eFigure 5: Estimated age-standardized healthcare spending per person by race/ethnicity for the 10 health conditions with the highest spending in 2016



Error bars indicate 95% uncertainty intervals. Values in parentheses for coefficient of variation represent 95% uncertainty intervals. Spending reported in 2016 US dollars per person.

eFigure 6: Age-standardized spending per person attributable to utilization and price and intensity of services in 2016 (in USD)



This figure is the same as Figure 3 of the main paper, except that: (1) we present the results in raw US dollars rather than in relative percentage terms, and (2) we display the results separated by type of care rather than by race/ethnicity group. Each grey bar represents the relative difference in spending for each race/ethnicity group, relative to all-population spending for each type of care. Types of care are mutually exclusive. Arrows indicate to what extent the difference from all-population spending was due to utilization or prices and intensity of treatment. Decomposition was not performed for nursing facility care due to the lack of utilization data for this type of care.

eTable6A: Statistically significant absolute differences in spending, utilization, and price and intensity of care for seven key diseases, 2016

			Ambulatory care		Emergency care		Inpatient care			Prescribed pharmaceuticals	
Race and ethnicity	Health condtion	Disease specific spending per case	Visits per case	Spending per visit	Visits per case	Spending per visit	Admissions per case	Beddays per admission	Spending per bedday	Prescriptions per case	Spending per prescription
American Indian, Alaska Native (non Hispanic)	Asthma										
	Cardiovascular diseases		0.673								
	Cerebrovascular disease										
	COPD						0.011				
	Diabetes mellitus			$981							
	Hypertension										
	Low back and neck pain	$624	0.763				0.001				
Asian, Native Hawaiian, Pacific Islander (non-Hispanic)	Asthma	$500								1.221	
	Cardiovascular diseases										
	Cerebrovascular disease							5.3			
	COPD										
	Diabetes mellitus	$2,363									$112
	Hypertension									5.964	
	Low back and neck pain	$785					0.002			0.238	
Black (non-Hispanic)	Asthma	$999			0.074		0.012			2.458	$192
	Cardiovascular diseases	$5,946	0.837								
	Cerebrovascular disease	$7,537					0.208			1.253	
	COPD	$6,372	0.824		0.390		0.233				
	Diabetes mellitus		1.545		0.068		0.044				
	Hypertension	$1,172			0.043		0.008	3.7		8.579	
	Low back and neck pain		0.781		0.043			3.4			
Hispanic	Asthma	$870			0.060		0.011				
	Cardiovascular diseases		0.822						$6,439	2.420	
	Cerebrovascular disease							5.6			
	COPD				0.329			4.2		4.003	
	Diabetes mellitus				0.053					11.007	
	Hypertension			$712	0.032		0.005			6.080	
	Low back and neck pain	$628	0.696				0.003	3.2		0.323	
Multiple races (non-Hispanic)	Asthma										
	Cardiovascular diseases	$3,174	0.700							1.828	
	Cerebrovascular disease										
	COPD		0.644								
	Diabetes mellitus										
	Hypertension	$1,314								6.090	
	Low back and neck pain									0.345	
White (non-Hispanic)	Asthma	$689	0.277		0.017		0.003	3.6		1.903	$233
	Cardiovascular diseases		1.334		0.071				$4,441	3.441	$95
	Cerebrovascular disease	$4,361					0.101	5.4	$3,322	0.757	
	COPD	$3,457			0.105		0.068	4.3			$207
	Diabetes mellitus		2.008		0.024		0.015			12.740	$149
	Hypertension	$858	0.706		0.013		0.002	4.0			
	Low back and neck pain	$1,221	1.359		0.015			3.6		0.552	

This figure is the same as Figure 4 of the main paper, except that: (1) we present the results in absolute amounts rather than in relative percentage terms, and (2) we include columns for spending per utilization. Only values with a bootstrap p-value < 0.05 were included in this table. Because of relatively large uncertainty intervals associated with small samples, many of the values were suppressed.

eTable6B: Statistically significant relative differences in spending, utilization, and price and intensity of care for seven key diseases, 2016

			Ambulatory care		Emergency care		Inpatient care			Prescribed pharmaceuticals	
Race and ethnicity	Health condtion	Disease specific spending per case	Visits per case	Spending per visit	Visits per case	Spending per visit	Admissions per case	Beddays per admission	Spending per bedday	Prescriptions per case	Spending per prescription
American Indian, Alaska Native (non-Hispanic)	Asthma										
	Cardiovascular diseases		-44%								
	Cerebrovascular disease										
	COPD						-88%				
	Diabetes mellitus			86%							
	Hypertension										
	Low back and neck pain	-43%	-35%				-85%				
Asian, Native Hawaiian, Pacific Islander (non-Hispanic)	Asthma	-34%								-39%	
	Cardiovascular diseases										
	Cerebrovascular disease							-2%			
	COPD										
	Diabetes mellitus	-29%									-22%
	Hypertension									-21%	
	Low back and neck pain	-28%					-62%			-53%	
Black (non-Hispanic)	Asthma	32%			136%		117%			23%	-12%
	Cardiovascular diseases	18%	-31%								
	Cerebrovascular disease	53%					65%			46%	
	COPD	62%	-28%		165%		148%				
	Diabetes mellitus		-17%		84%		99%				
	Hypertension	28%			106%		126%	-3%		14%	
	Low back and neck pain		-34%		111%			-3%			
Hispanic	Asthma	15%			92%		91%				
	Cardiovascular diseases		-32%						33%	-26%	
	Cerebrovascular disease							4%			
	COPD				123%			-2%		-35%	
	Diabetes mellitus				43%					-9%	
	Hypertension			-15%	56%		43%			-19%	
	Low back and neck pain	-43%	-41%				-46%	-9%		-36%	
Multiple races (non-Hispanic)	Asthma										
	Cardiovascular diseases	-37%	-42%							-44%	
	Cerebrovascular disease										
	COPD		-43%								
	Diabetes mellitus										
	Hypertension	43%								-19%	
	Low back and neck pain									-32%	
White (non-Hispanic)	Asthma	-9%	-6%		-47%		-44%	3%		-5%	6%
	Cardiovascular diseases		10%		-6%				-8%	5%	-3%
	Cerebrovascular disease	-12%					-19%	-2%	-13%	-12%	
	COPD	-12%			-29%		-28%	1%			2%
	Diabetes mellitus		8%		-34%		-32%			5%	5%
	Hypertension	-6%	-4%		-38%		-42%	3%			
	Low back and neck pain	11%	15%		-23%			2%		9%	

This figure is the same as Figure 4 of the main paper, except that we include columns for spending per utilization. Only values with a bootstrap p-value < 0.05 were included in this table. Because of relatively large uncertainty intervals associated with small samples, many of the values were suppressed.

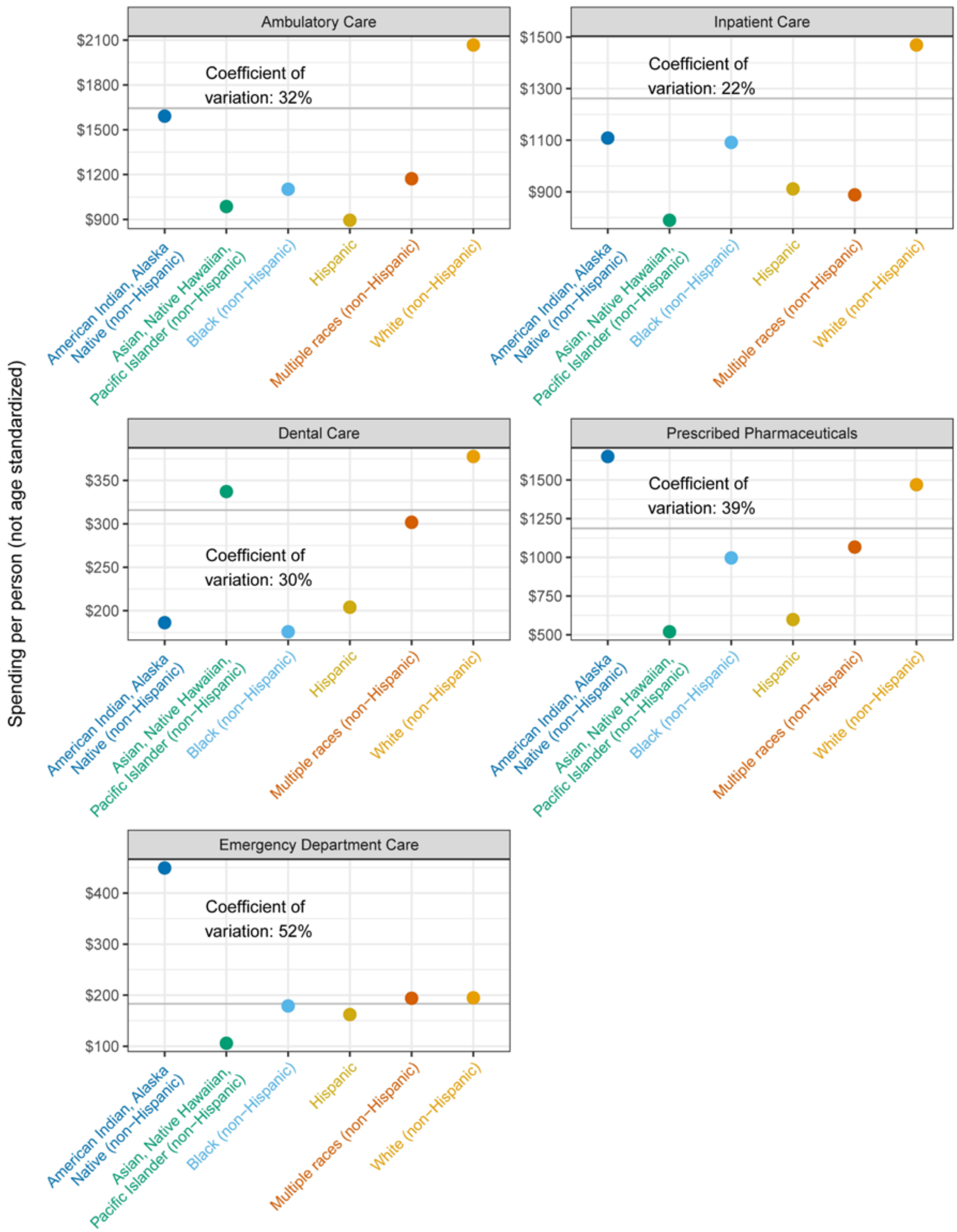
S13. Robustness checks

In order to evaluate how sensitive our results are to our modeling and age-standardization, we performed several checks. First, we compared our estimates to those observed in MEPS. In most cases, our modeling should serve primarily to strengthen and clarify patterns that likely exist in the original survey data, although the age-standardization may lead to important differences. For these comparisons, we created alternate versions of the spending per person illustrated in Table 1 and Figure 2 (from the paper) using raw, unadjusted MEPS data. Due to the lack of data past 2012 in the MCBS, we were not able to include nursing facility care in these figures.

eTable 7: Spending per person based on raw MEPS data (2016)

	Spending per person (across all ages, not age-standardized, and excluding long term care)
American Indian, Alaska Native (non-Hispanic)	$ 4,987
Asian, Native Hawaiian, Pacific Islander (non-Hispanic)	$ 2,738
Black (non-Hispanic)	$ 3,545
Hispanic	$ 2,770
Multiple races (non-Hispanic)	$ 3,623
White (non-Hispanic)	$ 5,579
Coefficient of Variation	30%

eFigure 7: Spending per person and type of care based on raw MEPS data (2016)



We expect our final results to differ somewhat from the raw MEPS data for several reasons. First and foremost, our final estimates include spending on nursing facility care, while these comparisons do not because MEPS does not include that population. Second, our estimates are scaled to the total amount of healthcare spending in inpatient care, ambulatory care, dental care, emergency department care, nursing facility care, and prescribed pharmaceuticals (separately for each) in the US to reflect total spending. Unadjusted MEPS estimates are known to under-count low-frequency, high spending patients. Third, our estimates have undergone several adjustments to address various limitations of the survey data. Fourth, our estimates have been modeled to address problems with small sample sizes and to generate a complete set of estimates. Finally, our estimates are age-standardized, while the raw MEPS data is not. This will mean that for race/ethnicity groups with relatively younger populations, the MEPS raw estimates will be lower. These differences will be especially severe for types of care that have much of the spending on older adults, such as inpatient care. Despite these differences, the raw data and the modeled, age-standardized estimates, and most importantly the key patterns described in the text are similar, although key differences, especially related to Black individuals and inpatient care do exist.

As an additional check, we compared our non-age-standardized final results against the spending per person summarized from MEPS by the Agency for Healthcare Research and Quality (AHRQ) on their website.[10] In order to compare against the race/ethnicity variables presented by AHRQ, we aggregated American Indian, Alaska Native, and Multiple Races into one group. This comparison is presented below.

eTable 8: IHME estimates compared with AHRQ estimates

	IHME estimates of spending per person (Not age standardized. Including nursing facility care)	AHRQ spending per person (Not age standardized. Not including nursing facility care)	IHME estimates of spending per person as a percent of spending on white people	AHRQ spending per person as a percent of spending on white people
American Indian, Alaska Native OR Multiple Races (non-Hispanic)	$ 6,553	$ 4,303	73%	71%
Asian, Native Hawaiian, Pacific Islander (non-Hispanic)	$ 4,444	$ 2,976	50%	49%
Black (non-Hispanic)	$ 6,546	$ 4,009	73%	67%
Hispanic	$ 4,682	$ 3,093	52%	51%
White (non-Hispanic)	$ 8,941	$ 6,020	100%	100%

Despite the fact that AHRQs estimates include spending on home care, over-the-counter medicine, and medical devices (both durables and non-durables) and do not include spending on nursing facility care (all of which make the AHRQ numbers distinct from our estimates), eTable 8 shows that our findings are largely in line with AHRQs presentation of MEPS data, especially when spending on each race/ethnicity is considered as a proportion of the spending on white people.

S14. GATHER Compliance

This study complies with the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) recommendations.[11] We have documented the steps involved in our analytical procedures and detailed the data sources used. See Table 3 for the GATHER checklist. The GATHER recommendations can be found here: http://gather-statement.org/

eTable 9: GATHER Compliance Checklist

#	GATHER checklist item	Description of compliance	Reference
Objectives and funding			
1	Define the indicators, populations, and time periods for which estimates were made.	Narrative provided in paper and methods appendix describing indicators, definitions, and populations	Main text (Methods—Overview, Geographical units and time periods) and methods appendix
2	List the funding sources for the work.	Funding sources listed in paper	Main text (Summary)
Data Inputs			
For all data inputs from multiple sources that are synthesized as part of the study:			
3	Describe how the data were identified and how the data were accessed.	Narrative provided in paper and methods appendix describing data-seeking methods	Main text (Methods) and methods appendix
4	Specify the inclusion and exclusion criteria. Identify all ad-hoc exclusions.	Narrative provided in paper and methods appendix describing inclusion and exclusion criteria	Main text (Methods) and methods appendix
5	Provide information on all included data sources and their main characteristics. For each data source used, report reference information or contact name/institution, population represented, data collection method, year(s) of data collection, sex and age range, diagnostic criteria or measurement method, and sample size, as relevant.	Metadata for data sources by component, geography, cause, risk, or impairment is available through an interactive, online data record	Link to the GHDx to be provided upon publication.
6	Identify and describe any categories of input data that have potentially important biases (e.g., based on characteristics listed in item 5).	Summary of known biases included in paper narrative	Main text (Limitations)
For data inputs that contribute to the analysis but were not synthesized as part of the study:			

7	Describe and give sources for any other data inputs.	Will be included in GHDx link	Link to the GHDx to be provided upon publication.
For all data inputs:			
8	Provide all data inputs in a file format from which data can be efficiently extracted (e.g., a spreadsheet as opposed to a PDF), including all relevant meta-data listed in item 5. For any data inputs that cannot be shared due to ethical or legal reasons, such as third-party ownership, provide a contact name or the name of the institution that retains the right to the data.	Downloads of input data available through online tools, including data visualization tools; input data not available in tools will be made available upon request	Online data visualization tools and the Global Health Data Exchange, http://ghdx.healthdata.org
Data analysis			
9	Provide a conceptual overview of the data analysis method. A diagram may be helpful.	Flow diagrams of the overall methodological processes, as well as cause-specific modeling processes, have been provided	Main text (Methods) and methods appendix
10	Provide a detailed description of all steps of the analysis, including mathematical formulae. This description should cover, as relevant, data cleaning, data pre-processing, data adjustments and weighting of data sources, and mathematical or statistical model(s).	Flow diagrams and corresponding methodological write-ups have been provided	Main text (Methods) and methods appendix
11	Describe how candidate models were evaluated and how the final model(s) were selected.	Details on evaluation of model performance have been provided	Methods appendix
12	Provide the results of an evaluation of model performance, if done, as well as the results of any relevant sensitivity analysis.	Details on evaluation of model performance have been provided	Methods appendix
13	Describe methods for calculating uncertainty of the estimates. State which sources of uncertainty were, and were not, accounted for in the uncertainty analysis.	Details on uncertainty calculations have been provided	Methods appendix
14	State how analytic or statistical source code used to generate estimates can be accessed.	Access statement provided	Code is provided in an online repository, link to the GHDx to be provided upon publication.
Results and Discussion			

15	Provide published estimates in a file format from which data can be efficiently extracted.	Results are available through online data visualization tools and the Global Health Data Exchange	Online data tools (data visualization tools and the Global Health Data Exchange, link to the GHDx to be provided upon publication.)
16	Report a quantitative measure of the uncertainty of the estimates (e.g. uncertainty intervals).	Uncertainty intervals are provided with all results	Main text, methods appendix, and online data tools (data visualization tools, data query tools, and the Global Health Data Exchange, link to the GHDx to be provided upon publication.)
17	Interpret results in light of existing evidence. If updating a previous set of estimates, describe the reasons for changes in estimates.	Discussion of methodological differences between GBD estimates and other available evidence provided in the paper and methods appendix	Main text (Methods and Discussion) and methods appendix
18	Discuss limitations of the estimates. Include a discussion of any modelling assumptions or data limitations that affect interpretation of the estimates.	Discussion of limitations was provided	Main text (Limitations) and methods appendix

S15. References

1. Dieleman JL, Cao J, Chapin A, et al. US health care spending by payer and health condition, 1996-2016. *JAMA*. 2020;323(9):863. doi:10.1001/jama.2020.0734

2. Bureau UC. 2019 Population Estimates by Age, Sex, Race and Hispanic Origin. The United States Census Bureau. Published June 25, 2020. Accessed November 12, 2020. https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html

3. Centers for Disease Control and Prevention. Bridged-Race Population Estimates. Accessed November 12, 2020. https://wonder.cdc.gov/bridged-race-population.html

4. GBD 2019 Demographics Collaborators. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950-2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020;396(10258):1160-1203. doi:10.1016/S0140-6736(20)30977-6

5. Centers for Disease Control and Prevention. U.S. Census Populations With Bridged Race Categories. National Vital Statistics System. Published July 6, 2020. Accessed November 13, 2020. https://www.cdc.gov/nchs/nvss/bridged_race.htm

6. Kolenikov, Stanislav. Resampling variance estimation for complex survey data. *The Stata Journal*. 2010;10(2):165-199. doi:10.1177/1536867X1001000201

7. Kolenikov, Stanislav. *Survey Bootstrap and Bootstrap Weights. Summer North American Stata Users' Group Meetings 2008 19, Stata Users Group. Revised 28 Aug 2008*.

8. Boos DD. Introduction to the bootstrap world. *Statistical Science*. 2003;18(2):168-174.

9. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York, NY: Chapman and Hall/CRC Press; 1994.

10. Agency for Health Care Research and Quality (AHRQ). Medical Expenditure Panel Survey (MEPS) summary tables. Accessed December 17, 2020. https://meps.ahrq.gov/mepstrends/hc_use/

11. Stevens GA, Alkema L, Black RE, et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *Lancet*. 2016;388(10062):e19-e23. doi:10.1016/S0140-6736(16)30388-9