

Mapping Disparities in Medication Access, Utilization, and Spending across the United States

Maitreyi Sahu

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Joseph L. Dieleman, Chair

Laura Dwyer-Lindgren

Annie Haakenstad

Program Authorized to Offer Degree:

Health Metrics Sciences

Chapter 3 ©Copyright 2025

Sahu M, Wagner TD, Thomson A, Beauchamp M, Campbell JD, Crosby S, DeJarnatt D, Lescinsky H, Salih RK, Taylor K, Weil M, Dwyer-Lindgren L, Haakenstad A, Scott JW, Stergachis A, Essien UR, Dieleman JL. “Prescription drug utilization and spending by race, ethnicity, payer, health condition, and US state.” JAMA Health Forum. Aug 2025.

Published by the JAMA Network

Licensed under the Creative Commons Attribution (CC BY 4.0) license

All other materials ©Copyright 2025

Maitreyi Sahu

University of Washington

Abstract

Mapping Disparities in Medication Access, Utilization, and Spending across the United States

Maitreyi Sahu

Chair of the Supervisory Committee:

Joseph L. Dieleman

Department of Health Metrics Sciences

Achieving equitable access to medicines requires measuring disparities in pharmaceutical utilization and spending and understanding the mechanisms that drive them. Yet in the United States, disparities in medication access, use, and spending—by race, geography, and disease—remain inadequately measured and insufficiently addressed. This dissertation develops new tools and applies them at scale to quantify these inequities, leveraging large-scale spending, prevalence, and geospatial data to examine underlying mechanisms and support targeted health policy action.

Chapter 2 introduces a novel probabilistic framework to map prescription drugs to likely health conditions in retail pharmaceutical claims data, addressing a major limitation of existing administrative datasets. Unlike approaches based on US Food and Drug Administration–approved indications or clinical guidelines, it captures real-world prescribing practices, enabling measurement of both on- and off-label use. The resulting drug-condition map supports national, state, and county-level estimates of condition-specific utilization and spending across the US.

Chapter 3 applies this drug-condition map to generate state-level estimates of pharmaceutical utilization and spending by race and ethnicity, producing the most comprehensive geographic mapping of medication inequities to date, covering 143 conditions, four payer types, and all 50 states and DC. Findings reveal persistent underutilization among Black populations relative to need, and substantial variation across states and conditions. Decomposition analyses highlight the relative contributions of disease prevalence, prescribing intensity, spending per prescription.

Chapter 4 evaluates the role of physical access to community pharmacies in shaping these disparities over a ten-year period across the US. Using over a decade of geolocated pharmacy data, I develop a metric incorporating both driving and walking proximity to assess access across urban and rural census block groups nationwide. I then examine how access relates to medication use, using a disparity index building on findings from Chapter 3. I find that inadequate pharmacy access is associated with wider racial disparities in medication use across health conditions.

Together, these chapters offer a data-rich framework for understanding and addressing inequities in access, use, and spending on prescribed medicines across the United States.

TABLE OF CONTENTS

Positionality.....	vi
Glossary.....	vii
Definitions.....	viii
Chapter 1 Introduction	1
1.1 Specific Aims	2
1.2 Study Context	2
1.3 Significance	3
1.4 Funding.....	3
1.5 Outline	4
Chapter 2 Developing a Probabilistic Map to Link Drugs to Diagnoses	5
Chapter 3 Prescription Drug Utilization and Spending by Race, Ethnicity, Payer, Health Condition, and US State	39
Chapter 4 Pharmacy Access, Utilization of Medicines, and Racial Disparities	54
Chapter 5 Conclusion	65
Chapter 6 Additional Contributions	68
5.1 Geographical variation in health spending and performance across the US	68
5.2 Racial and ethnic disparities in health spending across the US.....	68
5.3 Health care value and market consolidation in the US.....	69
5.4 Access to community pharmacies in the US	69
5.5 Alignment between pharmaceutical products and disease burden globally	69
5.6 Improving treatment delivery for HIV/AIDS medicines in South Africa	70
5.7 Mass Drug Administration (MDA) of albendazole in India, Benin, & Malawi.....	71
5.8 Health insurance claims data in Sub-Saharan Africa	72
5.9 Large collaborative efforts.....	72
Appendix A: Supplementary Materials for Chapter 2	74
Appendix B: Supplementary Materials for Chapter 3.....	84
Appendix C: Supplementary Methods for Chapter 4.....	176
References	178

DEDICATION

This thesis is dedicated to my mother, Meenakshi Srinivasan Sahu,
who dedicated her doctoral thesis on the Gum Nebula to me.

ACKNOWLEDGEMENTS

This thesis was deeply inspired by a session I attended at the 2022 AcademyHealth Annual Research Meeting titled ‘*Achieving Pharmacoequity – a Policy Prescription for Reducing Health Disparities.*’¹ The panel, featuring Drs. Utibe Essien, Stacie Dusetzina, and Dima Qato, shaped much of my thinking on the intersection of pharmaceutical policy and health equity. While not formal mentors or members of my committee, I have had the opportunity to meet and learn from each of them, and their research and scholarship have had a profound influence on this work.

Committee

I am fortunate to have had a brilliant and supportive doctoral committee, and I am deeply grateful to them for their thoughtful feedback and guidance throughout the development of this dissertation. Joe Dieleman, my dissertation chair, opened the door for me to explore the world of US health policy after years focused on global health, offering a compelling topic and opportunity that shaped this work. He consistently had a good idea for how to estimate just about anything, encouraged me to expand my technical and coding skills, and provided patient, generous support throughout. Laura Dwyer-Lindgren engaged deeply in the technical work, always provided detailed feedback, and even helped to establish the simulation framework. Her rigor and expertise in racial disparities strengthened this work immensely. Annie Haakenstad not only helped shape the framing of my dissertation, but also pushed me to think more carefully about the policy implications of my work and challenged me to ensure that each of my analytic choices aligned with meaningful policy questions. Finally, Andy Stergachis, my graduate student representative, offered helpful feedback on all my papers even though it was not required. He introduced me to the field of pharmaceutical policy, connecting me with colleagues in the University of Washington (UW) School of Pharmacy and beyond. I am grateful for his mentorship and his generous sharing of his network.

Program

I thank the Department of Health Metrics Sciences program staff, especially Carleen Ratcliffe and Kris Larson, for their support and guidance in navigating the university administrative processes.

US Disease Expenditure (DEX) Team

This dissertation would not have been possible without the work of the DEX team, which produces the county-level estimates of health spending and utilization that underpin this research. Led by Joe Dieleman, the team includes Meera Beauchamp, Sawyer Crosby, Drew DeJarnatt, Haley Lescinsky, Azalea Thomson, Max Weil, Kayla Taylor, and Goli Tsakalos. I am profoundly grateful for their collaboration, support, and generosity.

In particular, I worked closely with Max Weil for nearly four years to co-develop the probabilistic model linking prescription drugs to diagnoses. Azalea Thomson and Meera Beauchamp led the

development of the race and ethnicity imputation models that formed the basis of the parent study used in my second aim. Drew DeJarnatt and Haley Lescinsky supported elements of data processing and visualization for the pharmacy access analysis. I also thank the many former DEX team members whose past work contributed to the infrastructure on which this research depends.

While this dissertation builds on data and infrastructure developed through team science, the analysis plans, methodological choices, and interpretation are my own. I produced nearly all visualizations and was solely responsible for all writing.

Funding Support and Collaboration

I thank our collaborators at the National Pharmaceutical Council—especially Tyler Wagner and Jonathan Campbell—for their feedback and support in developing our health equity project.

Other Mentors

I have been lucky to learn from many generous academic mentors over the years. Through the UW Department of Global Health, I have worked closely with Professors Ruanne Barnabas (now at Massachusetts General Hospital), Adam Szpiro (UW Biostatistics), and Paul Revill (University of York)—all of whose support and guidance have been instrumental in my training.

I am also grateful to mentors who have supported me across institutions and stages of my career. I thank Günther Fink (Swiss TPH), Dick Chaisson, Larry Moulton, Adrienne Shapiro, and Céline Gounder (Center for Tuberculosis Research, Johns Hopkins University). I also appreciate my colleagues and friends from Swiss TPH, especially the postdoctoral fellows and collaborators who served as informal mentors. Their encouragement during a formative period, when my confidence in pursuing a doctorate was still emerging, has had a lasting impact on my academic path.

Friends and Family

First, I thank my parents, Meenakshi and Kailash Sahu, who have inspired, encouraged, and supported me through every stage of my education. Their example and their belief in the power of science, learning, and purpose have profoundly shaped the way I engage with the world.

I have also been fortunate to be part of an incredible community of doctoral students and friends in Seattle and beyond. I'm especially grateful for the support of my twelve doctoral classmates—including Nok Chhun, Katrin Fabian, Melody Wang, Jiawei He, and Aldina Mesic—as well as the junior members of the HMS doctoral program. Our mutual support certainly helped me to grow and succeed in the program. In addition, the many supportive friends I've made throughout my training and work in global health continue to inspire me.

Finally, I am deeply grateful to my partner, Friedrich Knuth, whose steadiness, wit, and grace have carried me through the demands of the final stretch of doctoral journey. He encourages me to think more deeply, live more intentionally, and be a better scientist—and life with him reminds me that the best adventures are the ones we share.

LAND ACKNOWLEDGEMENT

I conducted this work while based at the University of Washington, located on the ancestral lands of the Duwamish Tribe, the first people of Seattle, whose traditional territory includes what is now the greater Seattle area.

Although the Duwamish signed the 1855 Treaty of Point Elliott, ceding their land, the US government still has not fulfilled its treaty obligations and, as of 2025, continues to deny them federal recognition—excluding the Duwamish from tribal trust lands, educational and economic development funding, and access to guaranteed healthcare through the Indian Health Service.²

I honor with gratitude the land itself and the Duwamish People, past and present.³

Positionality

I am a woman of color and an immigrant to the United States. I spent my early childhood in Europe, in the Netherlands, Spain, and Germany—where universal access to social and health services tended to receive broader support. At seven, I moved to Baltimore—a city with stark disparities—where I remained through early adulthood. I was fortunate to begin my public health education at Johns Hopkins University, a leading medical research institution, though located in a Baltimore neighborhood with a long history of health disparities and mistrust towards the medical system.

Following this training, I spent several years working with health programs in underserved communities—including in South Sudan, Sierra Leone, and Malawi. I am deeply motivated by my family background, especially my father’s upbringing in a coastal village in Odisha, India, where smallpox was still circulating during his childhood—in a region that was among the last in the world to eliminate the disease. It has been a privilege to study at the University of Washington, where Dr. William Foege, who helped lead the smallpox eradication efforts in India, also trained.

While these experiences have shaped my commitment to health equity, I acknowledge that, as an Indian-American with two PhD-educated parents, I do not come from a group that has historically faced the highest levels of systemic mistreatment or exclusion in the US healthcare system. I have benefited from significant forms of privilege, including educational capital, economic security, and native English fluency. I also recognize that I may not experience racialization the same way as other minoritized groups—including the wide range of experiences among Asian immigrants and Asian Americans.⁴

I entered doctoral studies without prior academic experience in US racial health disparities. I am aware of critiques of ‘health equity tourism’ and the risks of opportunistic engagement in this space.⁵ I approach this work with a deep commitment to learning from and being guided by longstanding experts—both those on my committee and others whose work shapes our approach—including the community leaders, health equity scholars, policy advocates, and journal editors who have helped define the discourse in this field.

Glossary

ACS	American Community Survey
ATC	Anatomical Therapeutic Chemical (classification system)
CBG	Census Block Group
CRS	Coordinate Reference System
CMS	Centers for Medicare & Medicaid Services
DALY	Disability-Adjusted Life Year
DEX	US Disease Expenditure Project
FDA	US Food and Drug Administration
ICD	International Classification of Diseases
IHME	Institute for Health Metrics and Evaluation
IHS	Indian Health Service
IRA	Inflation Reduction Act
IRB	Institutional Review Board
MEPS	Medical Expenditure Panel Survey
NCHS	National Center for Health Statistics
NCPDP	National Council for Prescription Drug Programs
NDC	National Drug Code
NHANES	National Health and Nutrition Examination Survey
NHEA	National Health Expenditure Accounts
PBM	Pharmacy Benefit Manager
PDE	Medicare Part D Event Files
RMSE	Root Mean Squared Error
SHEA	State Health Expenditure Accounts
TAF	Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files
UI	Uncertainty Interval
US	United States of America

Definitions

This dissertation relies on the following conceptualizations, previously defined in the literature.

Race

A social construct used to categorize people based on socially determined boundaries, including physical features and ancestry. These boundaries are context-specific and can shift over time.⁶

Ethnicity

A categorization based on shared cultural traits such as language, religion, and traditions.^{7,8} In US public health contexts, this term is often used to describe Hispanic or Latino origin only.⁹

Health disparity

A measurable and avoidable difference in health outcomes between socially advantaged and disadvantaged groups, often shaped by policy.^{10,11} Used interchangeably with *health inequality*.¹⁰

Health inequity

A health difference that is not only unnecessary and avoidable, but also unfair and unjust.^{12,13}

Health care disparity (*a core focus of this dissertation*)

Derived from the definition of health disparity, a measurable and avoidable difference in access to or use of health services and facilities between groups, which may be shaped by policy.^{10,11,14}

Health care inequity (*a core focus of this dissertation*)

When use of health services (including access and allocation) does not correspond to health needs, particularly when socially disadvantaged groups receive less care than required.¹⁰

Structural racism (*a core focus of this dissertation*)

A system in which laws, policies, practices, and norms reinforce racial inequities, operating across sectors including health care to produce adverse outcomes for marginalized groups.¹⁵

Pharmacoequity (*a guiding framework for this dissertation*)

First coined by Essien et al. (2021), the goal of ensuring that everyone, regardless of race and ethnicity, socioeconomic status, or resources, has access to the highest-quality medications.¹⁶

Health equity

As defined by Whitehead (1992), the goal that everyone should have a fair chance to reach their full health potential, and no one should be disadvantaged in doing so, if avoidable.¹²

Chapter 1 Introduction

The United States (US) faces persistent racial and ethnic disparities in both life expectancy^{17–19} as well as use of prescription medications,^{20,21}—an important determinant of longevity.²² Despite escalating pharmaceutical expenditures across the U.S.,^{23–25} spending is unequally distributed: minoritized populations spend less on prescription medications compared with the population average, even after adjusting for age.²⁶ In addition, American Indian/Alaska Native and Black populations experience disproportionately worse health outcomes, including higher mortality^{17,19} and greater disease severity.^{27,28} These patterns point to pharmaceutical inequities, i.e., systematic underuse by marginalized populations relative to need,¹⁰ which may contribute to poorer outcomes.

A growing literature points to structural and service-level factors driving racial and ethnic disparities in access to and utilization of medicines, including *differential prescribing* (e.g., due to access to health providers,^{29,30} type of provider visited,^{31–33} and implicit bias or discrimination^{34–37}) and *differential uptake and adherence* (e.g., due to insurance coverage,^{38,39} pharmacy access,^{40–43} differential pricing across pharmacies,^{43–45} out-of-pocket cost burden,^{46–49} quality of care,⁵⁰ and health system distrust^{51–53}). These disparities persist even with similar provider access and insurance coverage,⁵⁴ reflecting broader inequities in social determinants such as food security,^{55,56} transportation,^{57,58} and housing.⁵⁵ Marginalized populations face compounding barriers—such as limited access to racially-concordant care;^{59–64} and, more broadly, residential segregation due to redlining that has concentrated communities in areas with poor pharmacy access,^{65,66} pharmacy closures,^{67,68} and limited drug availability (including stockouts);^{69,70}—which are rooted in and reinforced by historical and ongoing *structural racism* across health and non-health domains.^{71,72}

Recent calls to advance *pharmacoequity* highlight the need for more granular, actionable data.^{16,73} Yet efforts to measure disparities in pharmaceutical access and utilization have been constrained by data limitations and methodological gaps. Pharmaceutical claims data capture rich information on drugs and costs but lack information on the diagnosis for which a drug is prescribed—making it difficult to assess whether spending and use align with health needs. The complexity of off-label prescribing and multi-use medications further limits the utility of existing classification systems and one-to-one drug-condition maps. This dissertation addresses these limitations by introducing a probabilistic approach to link drugs to diagnoses, enabling more granular, county- and state-level estimates as well as a measure of inequality that reflects both access and need.

Pharmacy access is an emerging concern, with more pharmacies closing than opening nationally in recent years.^{68,74} The concept of “pharmacy deserts,” adapted from food access literature,^{75,76} highlights persistent geographic access gaps, particularly in predominantly Black and Hispanic neighborhoods.^{40,41} Recent pharmacy closures have been linked to immediate declines in prescription fills for cardiovascular medications⁷⁷ and anticonvulsants,⁷⁸ though the long-term impact remains unclear. The rise of telepharmacy may alleviate some of these dynamics, though its impact on disparities remains unclear.^{79–81} No prior study has systematically assessed how pharmacy access contributes to racial and ethnic inequities in medication use across conditions.

1.1 Specific Aims

This dissertation develops new tools and generates novel estimates to evaluate disparities in retail pharmaceutical access, utilization, and spending.

- **Aim 1 (Chapter 2):** Develop a *probabilistic mapping method* to link drugs to diagnoses using large-scale retail pharmaceutical claims data, enabling condition-level estimates of utilization and spending, including off-label use.
- **Aim 2 (Chapter 3):** Apply this map to produce the most comprehensive estimates to date of *pharmaceutical utilization and spending by race, ethnicity, payer, and health condition*, covering the entire US population across all 50 states and the District of Columbia, adjusted for population age and disease burden.
- **Aim 3 (Chapter 4):** Evaluate how *geographic access to community pharmacies* affects condition-specific, prevalence-adjusted medication use and racial disparities in use at the US county- and state-levels, using 2010-2019 pharmacy location and population data.

1.2 Study Context

This dissertation was embedded within—and contributed directly to—the US County Disease Expenditure (DEX) project at the Institute for Health Metrics and Evaluation (IHME), which estimated health care spending across 3,110 US counties from 2010 to 2019 by age, sex, payer, health condition, and type of care—including retail pharmaceuticals.^{82,83} It also incorporates complementary work from IHME’s US Health Disparities (USHD) project, which produced county-level estimates of mortality and disease burden by race and ethnicity.^{17,84}

While the DEX project previously produced national estimates of retail pharmaceutical spending,⁸⁵ including analyses by race and ethnicity,²⁶ several key gaps remained. First, although a prior attempt had been made,⁸⁶ no viable method existed to leverage large-scale insurance claims data to probabilistically assign prescriptions to their corresponding health conditions. Second, subnational estimates of pharmaceutical use and spending disaggregated by race and ethnicity were unavailable, obscuring geographic variation. Third, existing estimates of pharmaceutical spending by race and ethnicity were not fully standardized for age, geography, or disease prevalence—prior work standardized for just 20 conditions at the national level.²⁶ This dissertation addresses that limitation by leveraging USHD disease burden estimates to facilitate comparisons. Finally, no prior analysis linked prescription drug utilization to pharmacy access infrastructure.

This dissertation expands the methodological scope and empirical outputs of the DEX project by developing a probabilistic drug-condition map, generating race- and ethnicity-specific estimates of pharmaceutical use and spending at the state level, and introducing a measure of pharmacy access as a structural determinant of utilization.

1.3 Significance

Access to medications in the US remains unequal. Racial and ethnic disparities in prescription drug use persist even after adjusting for age and disease burden, reflecting not just differences but *pharmaceutical inequities*—systematic underuse relative to clinical need.²⁶ These health care inequities may contribute to significant health disparities including preventable hospitalizations, disease progression, and premature mortality among minoritized populations.^{17,87}

Efforts to understand and advance pharmacoequity have long been hampered by data and methodological limitations. Most pharmaceutical claims from health insurance databases lack diagnosis information, making it difficult to assess whether medication use aligns with health needs. Existing classification systems fail to account for off-label use or multi-use drugs, while national spending and utilization estimates mask substantial racial and geographic variation.

This dissertation directly addresses these gaps. First, it introduces a **probabilistic drug-condition mapping approach** that enables scalable linkage between pharmaceutical claims and the conditions for which drugs are prescribed. This probabilistic map supports the generation of condition-specific estimates of pharmaceutical spending and utilization across the US population, disaggregated by state, county, demographic group, payer, and health condition.

Second, it produces **the most granular estimates to date of pharmaceutical utilization and spending by race, ethnicity, payer, and condition, adjusted for age and disease burden** across all 50 states and D.C. This study links pharmaceutical utilization and disease prevalence estimates by race and ethnicity from the US County Disease Expenditure (DEX) and Health Disparities (USHD) projects to identify racial/ethnic groups, health conditions, and US states with the greatest medication access gaps.

Third, it evaluates **geographic access to community pharmacies** as a structural determinant of medication use. It introduces two novel metrics: (1) a granular measure of pharmacy access incorporating both driving and walking distances, and (2) a state-level disparity index, adapted from the Gini coefficient, quantifying whether medication use aligns with need. Building on evidence of pharmacy deserts in Black and Hispanic neighborhoods,^{40,41} this study examines how pharmacy availability shapes racial and ethnic disparities in condition-specific medication use.

Together, this dissertation develops scalable tools to quantify pharmaceutical inequities by race, place, and payer, and highlights structural policy levers—such as targeted pharmacy support in underserved areas—to address them. While focused on racial and ethnic disparities, the methods are adaptable to other dimensions of inequality, laying the foundation for a more equity-driven pharmaceutical policy framework.

1.4 Funding

This dissertation was completed as part of the US County Disease Expenditure (DEX) project, which received funding from the Peterson Center on Healthcare and Gates Ventures. Additional support for Chapters 3 and 4 was generously provided by the National Pharmaceutical Council.

1.5 Outline

This dissertation is structured as three stand-alone papers, each corresponding to one of the three aims described above.

- **Chapter 2** introduces a probabilistic method to assign prescriptions to diagnoses using large-scale insurance claims, a foundational tool for the analyses that follow. This method contributed to the DEX project.
- **Chapter 3** applies this method to generate population-standardized estimates of pharmaceutical utilization and spending by race, ethnicity, payer, and health condition across all US states. Building on outputs from the DEX race-ethnicity project, this chapter evaluates national and state-level patterns of pharmacoequity.
- **Chapter 4** examines a key structural determinant—pharmacy access—and evaluates its relationship to condition-specific medication use and racial and ethnic disparities. This study links the race- and ethnicity-specific pharmaceutical estimates from Chapter 3 to a novel metric of geographic access to community pharmacies.
- Additionally, **Chapter 5** summarizes 30 more academic manuscripts I contributed to or led during my doctoral studies, spanning topics related to health care spending and value, pharmacy access, and the burden and treatment of infectious diseases.

Figure 1.5: Dissertation Contributions to the US Disease Expenditure (DEX) Project

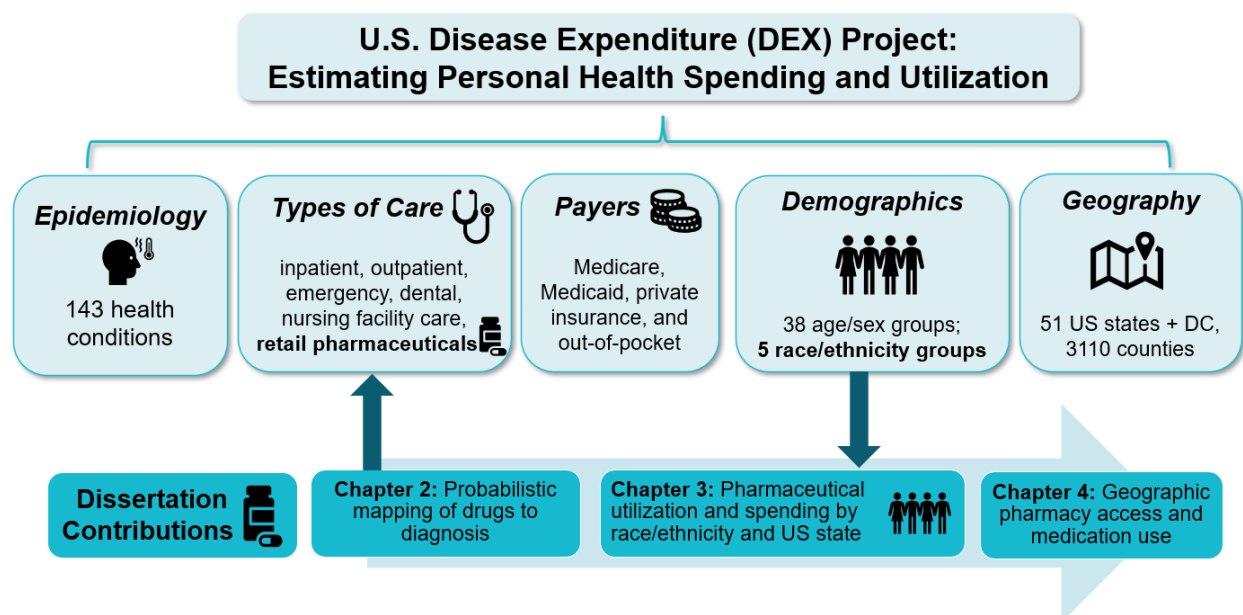


Figure 1.5. This figure illustrates how Chapters 2–4 extend the US Disease Expenditure (DEX) Project.

Together, these chapters build an integrated framework to quantify and address pharmaceutical inequities in the US.

Chapter 2 Developing a Probabilistic Map to Link Drugs to Diagnoses

This methods paper is being prepared for submission as:

- ❖ **Sahu M**, Weil M, Benson N, Dwyer-Lindgren L, Dieleman J. *Developing a probabilistic map to link drugs to diagnoses using MarketScan commercial health insurance claims*.

A prior version of this analysis for 2010-2019 informed pharmaceutical spending attribution in:

- ❖ Dieleman JL et al [incl. **Sahu M**]. *Tracking US Health Care Spending by Health Condition and County*. *JAMA*. 2025;333(12):1051-1061. doi:10.1001/jama.2024.26790
-

Abstract

Background

Understanding which condition a drug is intended to treat is critical for analyzing medication spending and utilization by health condition. However, pharmaceutical insurance claims do not record treatment intent, limiting their utility for disease-specific analyses.

Objective

To develop a probabilistic map linking drugs to their likely indications, including both on- and off-label usage, and demonstrate its application to state-level insurance claims data.

Methods

We constructed a cohort of prescription fills and temporally adjacent outpatient and hospital visits using MarketScan insurance claims (2010–2022), restricted to known on- and off-label indications from Merative’s Micromedex. We defined four estimators and used simulation to evaluate theoretical performance. We then trained random forest and logistic regression models to estimate the conditional probability of a diagnosis given a drug, incorporating patient- and drug-level covariates. Model selection and performance were assessed using cross-validation and external benchmarking with the National Health and Nutrition Examination Survey (NHANES).

Results

Among the top 30 drugs by spending and use, the most common attributed health condition categories were (1) type 2 diabetes, (2) attention-deficit/hyperactivity disorder, and (3) skin and subcutaneous diseases for spending, and (1) upper respiratory infections, (2) endocrine, metabolic, and immune disorders, and (3) skin and subcutaneous diseases for utilization. Condition-specific attribution enabled analysis of off-label prescribing across drugs.

Conclusion

This scalable mapping approach enables condition-level analysis of retail pharmaceutical claims. It supports improved study of pharmacoepidemiology, off-label use, and health equity, and provides a foundational tool for estimating prescription drug spending and access.

I. Introduction

Prescription drug spending in the US rose to \$449.7 billion in 2023—an 11.4% increase from the previous year, making it the fastest-growing major category of health care spending.²³ Understanding which health conditions drive this spending is essential for assessing whether medication use aligns with population health needs, evaluating therapeutic value, and promoting equity in access to medicines. While insurance claims and pharmaceutical sales data provide detailed information on prescription drug use and costs, condition-specific analyses are limited.

This limitation reflects a fundamental data gap: pharmaceutical claims capture drug identifiers, such as National Drug Codes (NDCs), but not the clinical indication. Medical claims list up to four diagnoses per encounter using the International Classification of Diseases (ICD) system and can be linked to prescriptions via the beneficiary ID, but these diagnoses are associated with the visit—not to a specific medication. As a result, clinical intent can be difficult to infer, especially in patients with multiple health conditions. This limits efforts to quantify condition-specific medication use, measure off-label prescribing, and assess disparities in treatment.

Claims databases offer a rich, longitudinal source of information on drug utilization, patient demographics, and spending. However, their full potential for pharmacoepidemiologic analysis remains unrealized without a scalable method to assign prescriptions to their clinical indications. This gap is especially relevant for the Disease Expenditure (DEX) project, which aims to disaggregate US health care spending—including retail pharmaceutical spending—by condition, payer, geography, and demographic group.^{26,85}

To address this challenge, we developed a probabilistic model to estimate the likelihood that a prescription corresponds to a specific diagnosis. Using MarketScan insurance claims (2010–2022), we constructed a dataset of drug fills and temporally co-occurring diagnoses within a 30-day lookback window. The model estimates the conditional probability of each diagnosis given the drug, producing a probabilistic map from NDCs and generic drug names to health conditions.

In this paper, we describe our methodological approach, validate its performance using cross-validation and external benchmarks, and demonstrate its application with real-world data. By probabilistically linking prescriptions to diagnoses, this work enables condition-specific analysis of pharmaceutical use, thereby supporting more equitable, evidence-based pharmaceutical policy.

II. Review of Prior Efforts and Classification Systems

To inform model development, we reviewed existing drug and diagnosis classification systems and prior attempts to link prescriptions to clinical indications. Standard classification schemes such as the Anatomical Therapeutic Chemical (ATC) system provide a structure for organizing pharmaceutical and diagnostic codes into therapeutic classes, but do not provide a direct or probabilistic mapping between drugs and health conditions.⁸⁸ Tools like RED BOOK⁸⁹ and RxNorm⁹⁰ support intermediate mappings (e.g., NDC to generic name or therapeutic class), but do not capture treatment intent.

We also examined prior efforts to link prescriptions to diagnoses, including survey-based approaches,⁹¹ proprietary episode groupers,⁹² and early modeling efforts.⁸⁶ These methods were either limited in scope, lacked generalizability, or did not account for the complexity of multi-morbidity and off-label prescribing. A more detailed review of the classification systems described above as well as linkage strategies is provided in Appendix Section A.

III. Data Sources

We constructed a dataset linking retail pharmaceutical claims to diagnosis codes from recent outpatient and inpatient visits. **Table 2.1** summarizes the data sources, their use in this study, and their coverage between 2010 and 2022.

Table 2.1: Overview of data sources, their usage in this study, and coverage (2010-2022)

Source	Dataset	Population	Use	Years	N
MarketScan Commercial Claims and Encounters ⁹³	Retail Pharmaceutical Claims	Commercially insured; some Medicare Advantage	Model development and results presentation	2010-2022	4,043,822,393
MarketScan Commercial Claims and Encounters ⁹³	Outpatient and Inpatient Visit Claims	Commercially insured; some Medicare Advantage	Model development and results presentation	2010-2022	12,202,920,961
Merative Micromedex ⁹⁴	Drug Indications (FDA and non-FDA uses)	N/A	Restriction of dataset to only known on- or off-label uses	Snap-shot (2024)	N/A
MarketScan RED BOOK ⁸⁹	RED BOOK Drug Database	N/A	Linking National Drug Codes (NDC) to generic drug names	2010-2022	N/A
National Health and Nutrition Examination Survey (NHANES)	Prescription Medications Files	Nationally representative survey	Model benchmarking	2013-2023	143,988
US Disease Expenditure Project (DEX) ⁸²	ICD to DEX Condition Mapping	N/A	Mapping ICD-coded diagnoses to a reduced set of health conditions	2010-2019	N/A

i. MarketScan Commercial Claims and RED BOOK^{89,93}

Our primary source was the Merative MarketScan Commercial Claims and Encounters database (2010-2022), which includes over 250 million enrollees covered by private insurance and

Medicare Advantage. We began over 4 billion retail pharmaceutical claims and over 12 billion medical visit claims from outpatient and inpatient encounters (**Table 2.1**). MarketScan records include demographic (age, sex) and geographic (US state) information.⁹³

Our analysis focused on retail prescription drugs (i.e, drugs dispensed at pharmacies) and excluded inpatient or in-office administration, which together account for approximately 40% of pharmaceutical use and 30% of spending.⁷ Each claim includes an 11-digit NDC identifying the drug's manufacturer, formulation, and packaging, which we linked to its generic name, therapeutic class, and brand/generic status using the MarketScan RED BOOK.⁸⁹

To identify plausible clinical indications, we linked drug claims to recent encounters (outpatient or inpatient visits), which contained up to four diagnosis codes (ICD-9/10) recorded per visit.

To keep data processing manageable, we limited our analysis to 806 selected drugs that together accounted for approximately 90% of total spending, out-of-pocket spending, and prescription fills in the MarketScan database (out of 9,452 drugs in total). While we did not apply formal inclusion criteria, we curated the list of 806 drugs to capture drugs that were either commonly prescribed or associated with high total or per-prescription spending (e.g., specialty drugs). **Table 2.2** presents drug-level coverage statistics by year, showing the representativeness of our selected sample.

Table 2.2: Share of total spending, out-of-pocket spending, and prescription fills captured by 806 selected drugs in MarketScan, 2010-2022

Year	Total Spending	Out-of-pocket Spending	Prescription Fills
2010	89.4%	87.8%	86.0%
2011	90.3%	88.7%	87.1%
2012	90.9%	89.4%	87.9%
2013	91.4%	89.8%	88.5%
2014	92.2%	90.2%	89.1%
2015	93.1%	90.8%	89.6%
2016	94.2%	91.6%	90.1%
2017	94.2%	91.7%	90.5%
2018	93.2%	91.3%	91.1%
2019	91.8%	90.2%	91.1%
2020	90.9%	89.5%	91.3%
2021	89.7%	89.0%	92.0%
2022	87.3%	87.3%	90.9%
Total	91%	90%	89%

ii. *US Disease Expenditure Project (DEX)*⁸²

We used the health condition classification system developed by the US Disease Expenditure (DEX) project, which comprehensively mapped all ICD-9 and ICD-10 codes to a standardized list of 148 mutually exclusive health conditions and 28 broader health condition categories. This

mapping provides a consistent framework for analyzing healthcare spending, utilization, and burden by health condition. In our study, we applied the DEX condition hierarchy to diagnosis codes from MarketScan claims, reducing the number of labels to 143 distinct health conditions, which served as classification targets in our modeling.

iii. *Merative Micromedex*⁹⁴

To identify known clinical indications, we used Micromedex⁹⁴, a comprehensive drug reference database that catalogs both FDA-approved and non-FDA uses of medications. For each drug, using its generic name, we extracted all known indications and manually mapped them to the 143 health conditions from the DEX project. This allowed us to restrict our analysis to drug–diagnosis pairs with documented on- or off-label use. We further restricted diagnoses to only known uses from Merative Micromedex, which lists accepted on- and off-label uses given the generic drug name.⁹⁴

iv. *National Health and Nutrition Examination Survey (NHANES)*

Finally, we used the National Health and Nutrition Examination Survey (NHANES) Prescribed Medicines files (2013–2023) for external benchmarking.⁹⁶ NHANES is a nationally-representative survey that collects detailed information on health status, medical conditions, and prescription drug use through interviews and physical examinations. To ensure representativeness of the national population, we applied the provided person-level survey weights which account for NHANES’s complex sampling design, which includes intentional oversampling of specific subpopulations.

The prescribed medicines component includes self-reported data on medications taken in the past 30 days, linked to self-reported indication, allowing us to benchmark our internally-validated probabilities against real-world survey data. Specifically, patients are asked: ‘In the past 30 days, have you taken any prescription medications?’ as well as the reason why they were taking it. Up to three diagnoses can be listed for the self-reported indication, which are then reported both in plain language and as ICD codes, which are assigned by the interviewer.

IV. Data Processing

i. *Final Testing and Training Dataset*

Figure 2.2 summarizes the data pipeline used to construct the final training and testing dataset. We began by merging pharmaceutical claims from MarketScan with drug details from RED BOOK, including generic name, dosage, strength, package size, and route of administration. Because pharmaceutical claims lack diagnosis information, we linked them to medical visit claims using a multi-step process designed to reduce noise.

First, we retained only original prescriptions (excluding refills) to reduce noise, as these are most likely to reflect new prescribing decisions. Claims with negative or null total or out-of-pocket payments were excluded. Outpatient and inpatient visit claims—each listing up to four

diagnoses—were reshaped to long format and filtered to include only diagnoses that mapped to our list of 143 DEX health conditions (which were linked using the DEX ICD map). Then, using patient IDs, we linked pharmaceutical and medical claims, retaining only matches within 30 days prior to the prescription fill. When multiple visits matched, we selected the one closest in time. This resulted in a dataset with one row per claim-diagnosis pair (up to four rows per claim).

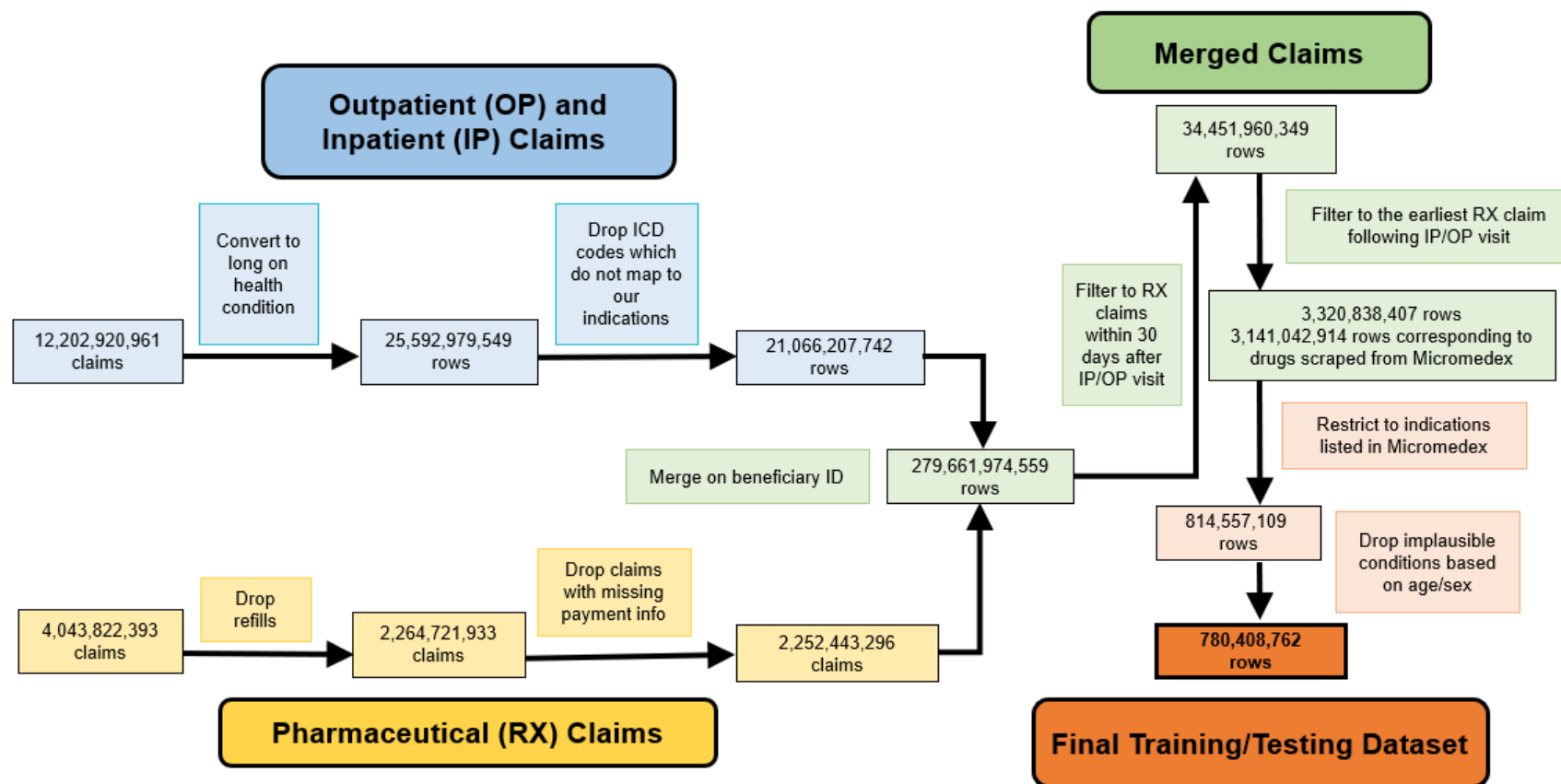
To further reduce noise from unrelated co-morbidities (e.g., linking insulin to hypertension solely due to common co-occurrence) we restricted the dataset to clinically-plausible indications. The Merative Micromedex Compendium catalogs accepted on- and off-label uses along with supporting evidence.⁹⁴ First we manually mapped plain-language indications listed in Micromedex to our standardized list of 143 conditions (Appendix A2). Then we restricted the analytic dataset to these conditions, which reduced the dataset from 3.14 billion to 814 million rows. Finally, we excluded implausible age/sex-condition combinations (e.g., cervical cancer in men), as detailed elsewhere,⁸² yielding a final training/testing dataset of 780.4 million rows.

The final dataset includes many co-diagnosed claims, i.e., those with more than one diagnosis per drug. **Table 2.3** summarizes the co-diagnoses across the dataset, and **Figure 2.3** shows the share of multi-diagnosed claims among the most commonly used and highest-spend drugs.

ii. Benchmarking Dataset

We processed the nationally representative NHANES data using the same approach as for MarketScan. Prescribed medicines which were already linked to self-reported indications were then converted to long format and filtered to retain only clinically plausible uses based on Micromedex, thereby excluding spurious self-reports (e.g., back pain for buprenorphine). For the eight tracer drugs listed below, we mapped the reported ICD codes to DEX health conditions and health condition categories for comparison. When analyses were restricted to individuals taking a given drug, estimates reflected the weighted subpopulation of US individuals who reported use of that drug, rather than the general population.

Figure 2.2: Data processing flowchart for generating the final training and testing dataset



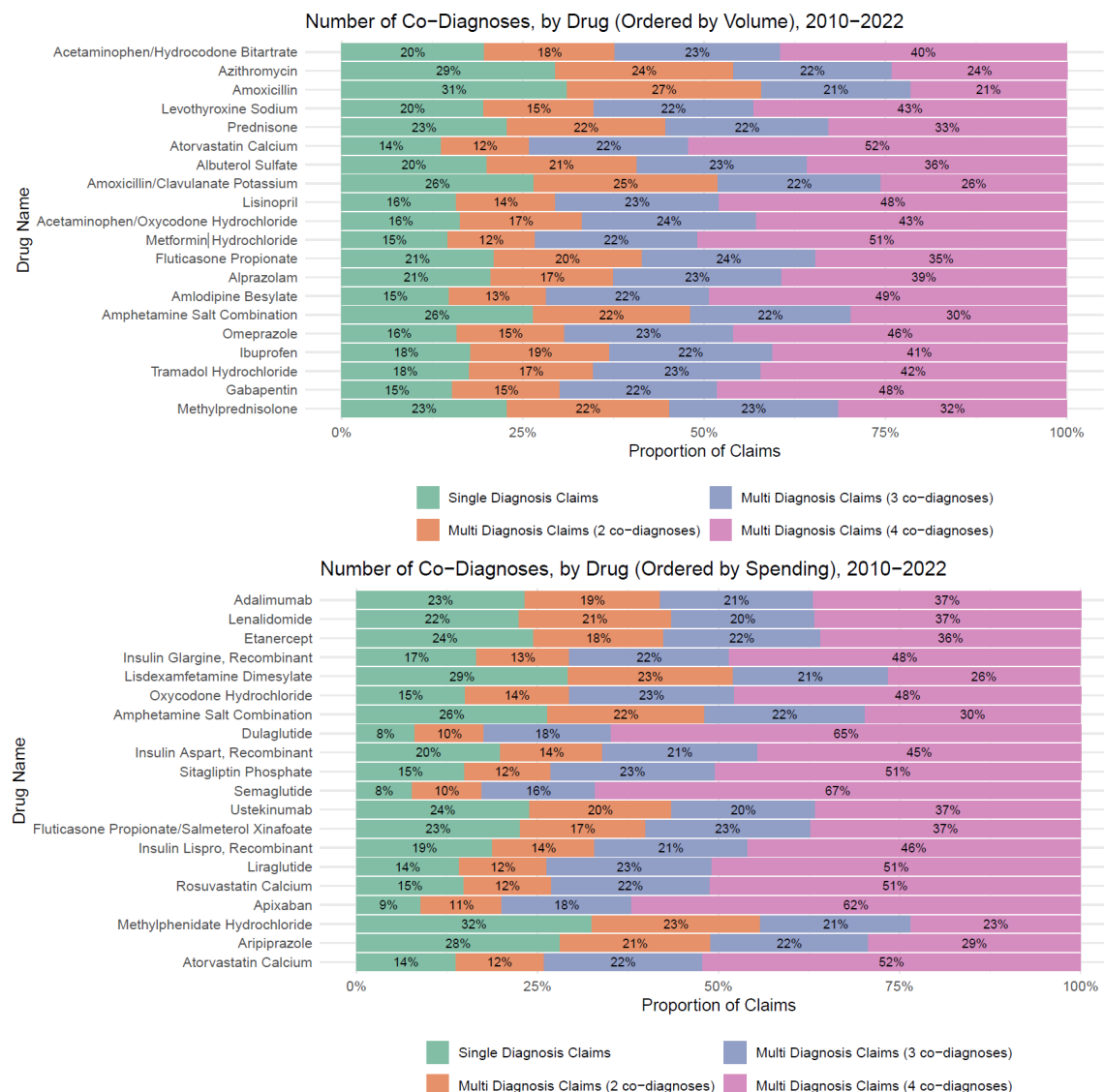
Note: The full processed dataset includes only drugs whose generic names could be mapped to entries in the Micromedex database. After restricting to known on- and off-label drug–indication pairs, the dataset was reduced from 3.14 billion to 814 million rows, approximately one-quarter of its original size.

Table 2.3: Proportion of Single and Multi-Diagnosis Claims in MarketScan, 2010–2022

	Single Diagnosis Claims	Multi- Diagnosis Claims (2)	Multi- Diagnosis Claims (3)	Multi- Diagnosis Claims (4)
Weighted Average (% of claims) <i>(volume-weighted across 806 drugs)</i>	20.9%	18.1%	22.4%	38.5%
Weighted Interquartile Range (IQR)	17.0% - 23.7%	14.7% - 20.7%	21.9% - 23.2%	32.7% - 44.9%
Full Range [min, max]	0% - 100%	0% - 100%	0% - 51.6%	0% - 68.8%

Note: This table summarizes our 806 selected drugs. Means and IQRs are weighted by total claim counts.

Figure 2.3: Proportion of Multi-Diagnosis Claims by Drug in MarketScan, 2010–2022



Notes:

- Panel A (top) shows the top 20 drugs, ordered by total prescription fills; Panel B (bottom) shows the top 20 drugs, ordered by total payments in MarketScan 2010-2022.

V. Assumptions Embedded in Data Processing: Simulation-Based Evaluation

We begin by outlining a two-diagnosis scenario in which a prescription drug may be indicated for either of two conditions (e.g., Metformin is indicated for both diabetes and obesity, which often co-occur; however, when both are coded, treatment intent is typically more likely to be diabetes).

Let the total number of claims be $N = A + B + C + D$, where:

- **A**: claims with diagnosis A only, prescribed for A
- **B**: claims with diagnosis B only, prescribed for B
- **C**: claims with both A and B, prescribed for A (B is present but not the true intent)
- **D**: claims with both A and B, prescribed for B (A is present but not the true intent)

Our framework assumes that each prescription has a *single true treatment intent*, consistent with clinical workflows in which prescriptions are generally written to treat a primary diagnosis, even when multiple diagnoses are coded. Since treatment intent is not directly observable in retail pharmaceutical claims data, when both A and B are coded, we cannot distinguish claims in group C (prescribed for A) from group D (prescribed for B).

Our objective is to estimate the true treatment intent, i.e., the probability P^d that a given claim was prescribed to treat diagnosis d .

- For diagnosis A:

$$P^A = \frac{A + C}{A + B + C + D}$$

- For diagnosis B:

$$P^B = \frac{B + D}{A + B + C + D}$$

As described in our data processing section, we transform the data to a claim–diagnosis format (one row per diagnosis) to support modeling. Before introducing the machine learning approach, we define four attribution strategies based solely on diagnosis presence, each reflecting a distinct set of assumptions about how to distribute treatment intent across co-diagnosed claims:

1. The row-based estimator, $\tilde{P}$
2. The equal-weight estimator, $\tilde{P}_{1/n}$
3. The single-diagnosis estimator, $\tilde{P}_{single-dx}$
4. The bounds-based estimator, $[\tilde{P}_{lower}, \tilde{P}_{upper}]$

These approaches do not use covariates or model treatment intent; rather, they make explicit the assumptions embedded in diagnosis-level attribution when no predictive modeling is used. We use simulation to evaluate how these assumptions perform under varying conditions and highlight their structural and clinical limitations. The machine learning model we describe later then builds on these processed data, serving as a smoothing and imputation layer.

(I) Row-based estimator, $\tilde{P}$

i. Definition

The naïve estimator derived from the processed dataset would assume that every diagnosis row reflects treatment intent. We denote this estimator $\tilde{P}$, and define it as:

- For diagnosis A:

$$\tilde{P}^A = \frac{A + C + D}{A + B + 2C + 2D}$$

- For diagnosis B:

$$\tilde{P}^B = \frac{B + C + D}{A + B + 2C + 2D}$$

ii. Conditions for equality with the true treatment intent, P

Algebraically, in this two-diagnosis scenario, $P^A = \tilde{P}^A$ if and only if:

$$AC + C^2 = BD + D^2$$

This condition is also necessary and sufficient for $P^B = \tilde{P}^B$. The row-based estimator over-attributes treatment to both diagnoses because co-diagnosed claims get double-counted in the data structure with one row per claim-diagnosis. It is unbiased when the total contribution to over-attribution from diagnosis A (captured by $AC + C^2$) equals that from diagnosis B (captured by $BD + D^2$), a somewhat more complex requirement than simple proportionality.

iii. Algebraic derivation of equality between P^A and $\tilde{P}^A$

The condition above can be derived by equating the true parameter and the naïve estimator:

$$P^A = \tilde{P}^A$$

$$\frac{A + C}{A + B + C + D} = \frac{A + C + D}{A + B + 2C + 2D}$$

Cross-multiplying: $(A + C)(A + B + 2C + 2D) = (A + C + D)(A + B + C + D)$

Expanding both sides:

- *Left-hand side:* $A^2 + AB + 3AC + 2AD + BC + 2C^2 + 2CD$
- *Right-hand side:* $A^2 + AB + 2AC + 2AD + BC + C^2 + 2CD + BD + D^2$

Subtracting common terms yields the equality condition: $AC + C^2 = BD + D^2$

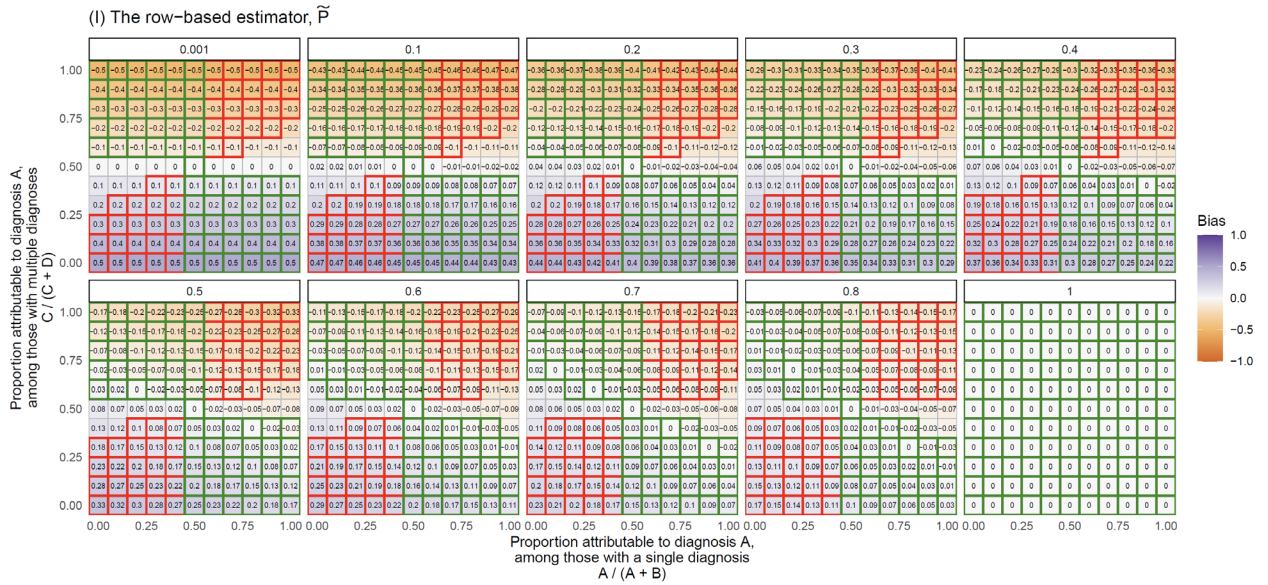
iv. *Simulation grid examining bias for the row-based estimator, $\tilde{P}$*

To evaluate how $\tilde{P}$ performs under varying conditions, we present a structured simulation grid (also referred to as an estimator bias map) comparing $\tilde{P}^A$ to the true treatment intent P^A in the two-diagnosis scenario. We varied three key simulation parameters:

- **X-axis:** The proportion of intended diagnosis A in single-diagnosis claims: $A/(A+B)$
- **Y-axis:** The proportion of intended diagnosis A in co-diagnosed claims: $C/(C+D)$
- **Facets (panels):** The proportion of all claims with a single diagnosis: $(A+B)/N$

The facets and x-axis are observable from our data, but the y-axis is not observable.

Figure 2.4: Simulation of row-based attribution bias, varying the proportion of claims with only a single diagnosis (0 = all co-diagnosed; 1 = all single-diagnosis)



*Colors represent the bias $\tilde{P}^A - P^A$. Purple = over-attribution to A; Orange = under-attribution; White = no bias (perfect agreement); Green outline = highest performing estimator among the 3 studied (I – III); Red outline = lowest performing.

The simulation shows that when co-diagnosed claims are rare, the naïve estimator $\tilde{P}^A$ approximates P^A . However, as the proportion of co-diagnosed claims increases, bias increases as well. The row-based estimator is least biased when $C/(C+D) \approx 0.5$, meaning when treatment intent is equally split between co-diagnoses. Bias increases as the distribution of intended diagnoses becomes more skewed, particularly when co-diagnosed claims are frequent.

(II) The equal-weight estimator, $\tilde{P}_{1/n}$

i. Definition

We next consider a second estimator that avoids double-counting by equally distributing attribution across all diagnoses coded on a claim. Instead of treating each expanded diagnosis row equally, this estimator assigns a weight of $1/n_i$ to each diagnosis for claim i , where n_i is the number of diagnoses present for that claim. In the processed MarketScan dataset, up to four diagnoses can be coded per claim, so $n_i \in \{1, 2, 3, 4\}$. This approach ensures that each claim contributes exactly one unit of attribution.

In the two-diagnosis scenario ($n = 2$), this is equivalent to assuming that each co-diagnosed claim contributes 0.5 units to A and 0.5 units to B:

- For diagnosis A:

$$\tilde{P}_{1/2}^A = \frac{A + \frac{1}{2}(C + D)}{A + B + C + D}$$

- For diagnosis B:

$$\tilde{P}_{1/2}^B = \frac{B + \frac{1}{2}(C + D)}{A + B + C + D}$$

ii. Conditions for equality with the true treatment intent, P

Algebraically, in the two-diagnosis scenario, $P^A = \tilde{P}_{1/2}^A$ if and only if:

$$C = D$$

This condition is also necessary and sufficient for $P^B = \tilde{P}_{1/n}^B$. Intuitively, the equal-weight estimator is unbiased if and only if treatment intent in co-diagnosed claims is exactly balanced.

iii. Algebraic derivation of equality between P^A and $\tilde{P}_{1/2}^A$

This condition can be derived by equating the true parameter and the equal-weight estimator:

$$\begin{aligned} P^A &= \tilde{P}_{1/2}^A \\ \frac{A + C}{A + B + C + D} &= \frac{A + \frac{1}{2}(C + D)}{A + B + C + D} \end{aligned}$$

Multiplying both sides by the denominator: $A + C = A + \frac{1}{2}(C + D)$

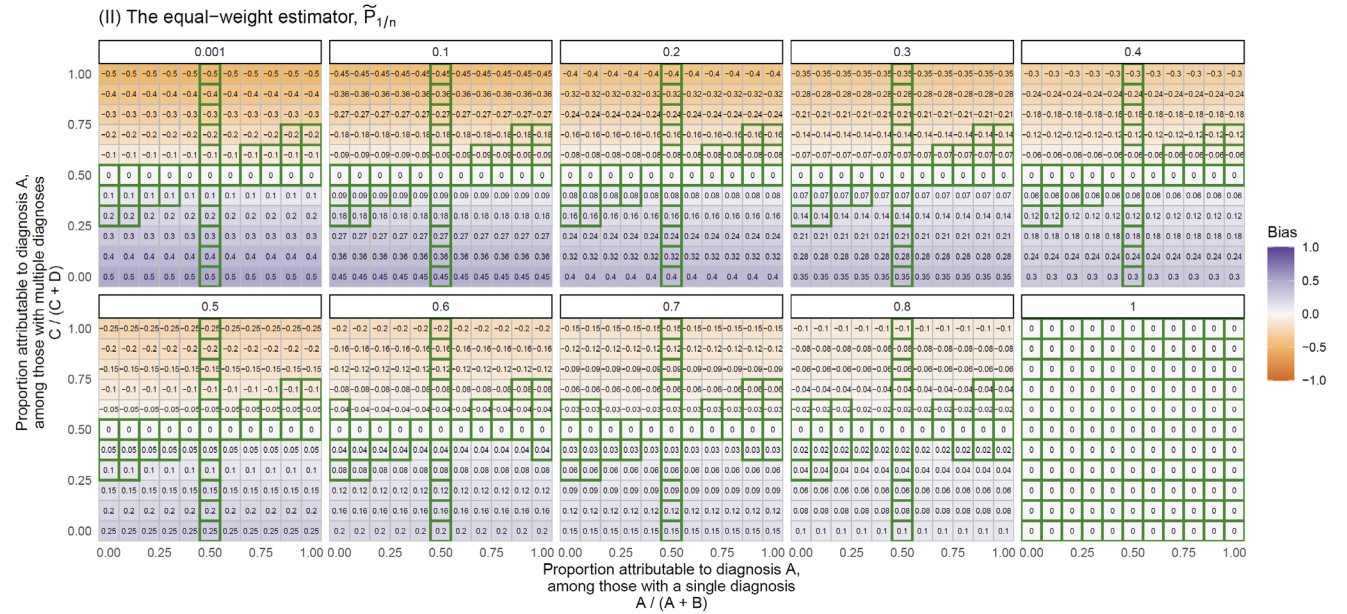
Subtract A from both sides and multiply by two: $C = \frac{1}{2}(C + D) \rightarrow 2C = C + D$

Subtracting common terms yields the equality condition: $C = D$

iv. **Simulation grid examining bias for the equal-weight estimator, $\tilde{P}_{1/n}$**

We simulated how $\tilde{P}_{1/2}$ performs under varying conditions for two diagnoses.

Figure 2.5: Simulation of equal-weight attribution bias, varying the proportion of claims with only a single diagnosis (0 = all co-diagnosed; 1 = all single-diagnosis)



*Colors represent the bias $\tilde{P}_{single-ax}^A - P^A$. Purple = over-attribution to A; Orange = under-attribution; White = no bias (perfect agreement); Green outline = highest performing estimator among the 3 studied (I – III); Red outline = lowest performing.

The equal-weight estimator assumes that co-diagnosed claims should be attributed equally across diagnoses. This estimator performs best, exhibiting zero bias, when the true treatment intent is also equally split, meaning when $C / (C + D) = 0.5$ in the two-diagnosis scenario. This condition directly aligns with the estimator's core assumption.

Outside this equivalence condition, the estimator's performance declines, particularly in panels with a greater number of co-diagnosed claims. In these scenarios, the equal-weight estimator often performs only marginally better than the row-based approach.

While the equal-weight estimator is simple, its assumption that all co-diagnoses are equally means its performance depends heavily on how treatment intent is actually distributed.

(III) Single-Diagnosis Estimator, $\tilde{P}_{single-dx}$

i. Definition

We also consider a third estimator which is restricted to only A and B (i.e., single-diagnosis claims only) – thereby reducing potential noise from co-diagnoses, but also excluding much of the available data. This third estimator, $\tilde{P}_{single-dx}$, can be defined as follows.

- For diagnosis A:

$$\tilde{P}_{single-dx}^A = \frac{A}{A + B}$$

- For diagnosis B:

$$\tilde{P}_{single-dx}^B = \frac{B}{A + B}$$

ii. Conditions for equality with the true treatment intent, P

Algebraically, in the two-diagnosis scenario, $P^A = \tilde{P}_{single-dx}^A$ if and only if:

$$\frac{A}{B} = \frac{C}{D}$$

This condition is also necessary and sufficient for $P^B = \tilde{P}_{single-dx}^B$. Thus, the single-diagnosis estimator is only unbiased when co-diagnosed claims are either (a) absent or (b) their treatment intent is proportional to that observed in single-diagnosis claims.

iii. Algebraic derivation of equality between P^A and $\tilde{P}_{single-dx}^A$

This condition can be derived by equating the true parameter and the single-diagnosis estimator:

$$P^A = \tilde{P}_{single-dx}^A$$

$$\frac{A + C}{A + B + C + D} = \frac{A}{A + B}$$

Cross-multiplying: $(A + C)(A + B) = A(A + B + C + D)$

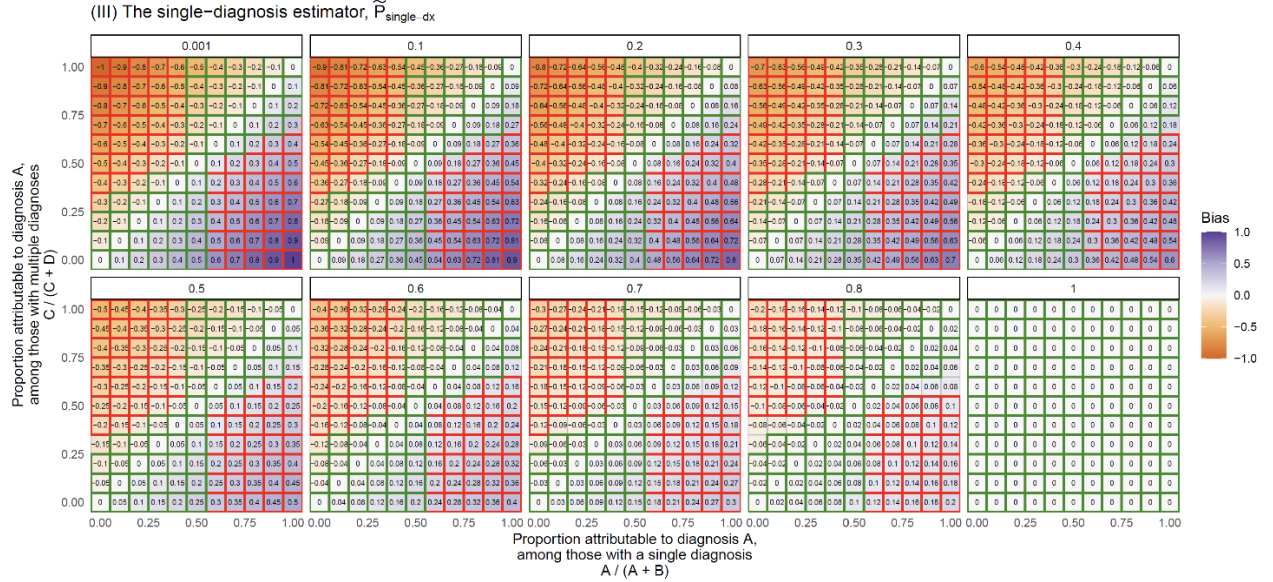
Expanding both sides: $A^2 + AB + AC + BC = A^2 + AB + AC + AD$

Subtracting common terms yields the equality condition: $AD = BC \rightarrow \frac{A}{B} = \frac{C}{D}$

iv. *Simulation grid examining bias for the single-diagnosis estimator, $\tilde{P}_{single-dx}$*

Here we simulate how $\tilde{P}_{single-dx}$ performs under varying real-world conditions. We present a estimator bias map comparing $\tilde{P}_{single-dx}^A$ to the true treatment intent P^A in the two-diagnosis scenario, varying the same simulation parameters described above. The results are shown below.

Figure 2.6: Simulation of single-diagnosis attribution bias, varying the proportion of claims with only a single diagnosis (0 = all co-diagnosed; 1 = all single-diagnosis)



*Colors represent the bias $\tilde{P}_{single-dx}^A - P^A$. Purple = over-attribution to A; Orange = under-attribution; White = no bias (perfect agreement); Green outline = highest performing estimator among the 3 studied (I – III); Red outline = lowest performing.

The bias grid shows that the single-diagnosis estimator $\tilde{P}_{single-dx}$ is unbiased only when the distribution of treatment intent among co-diagnosed claims is equal to that of single-diagnosis claims, i.e., when $A/(A+B) = C/(C+D)$, and bias is minimized when the treatment intent for co-diagnosed claims is similar to those for single-diagnosis claims.

However, when treatment intent in co-diagnosed claims diverges from the single-diagnosis distribution, the estimator becomes biased, either under- or over-attributing prescriptions to diagnosis A. This bias increases as the proportion of co-diagnosed claims grows, since more of the total rows are excluded from the estimator. When all claims are co-diagnosed (i.e., the panel on the top left), the estimator fails completely.

(IV) The bounds-based estimator, $[\tilde{P}_{lower}, \tilde{P}_{upper}]$

i. Definition

Finally, we define a theoretical bound for the estimator, which we denote $[\tilde{P}_{lower}, \tilde{P}_{upper}]$, reflecting the full range of possible values for the treatment intent associated with a diagnosis when no information is available about how co-diagnosed claims should be attributed. Unlike the point estimators (I – IV), this approach does not make any underlying assumptions.

For the two-diagnosis scenario, we define the bounds for A as follows:

- **Lower bound:** All co-diagnosed claims (C + D) are attributed to diagnosis B

$$\tilde{P}_{lower}^A = \frac{A}{A + B + C + D}$$

- **Upper bound:** All co-diagnosed claims (C + D) are attributed to diagnosis A

$$\tilde{P}_{upper}^A = \frac{A + C + D}{A + B + C + D}$$

The true treatment proportion for diagnosis A lies within the interval:

$$P^A \in \left[\frac{A}{A + B + C + D}, \frac{A + C + D}{A + B + C + D} \right]$$

The same logic applies to diagnosis B:

$$P^B \in \left[\frac{B}{A + B + C + D}, \frac{B + C + D}{A + B + C + D} \right]$$

These bounds rely on the extreme assumption that each co-diagnosed claim is attributable entirely to one of the listed diagnoses, which guarantees that the true value lies within the resulting interval. However, when co-diagnoses are common, the bounds can become very wide and difficult to interpret, as multiple diagnoses may have overlapping or indistinguishable plausible ranges, offering limited practical information. Accordingly, we treat these bounds as a theoretical reference rather than a practical estimator. Nonetheless, they may still be useful for sensitivity analyses or conservative reporting.

Of note, the midpoint of this interval corresponds exactly to the equal-weight estimator, $\tilde{P}_{1/n}$:

$$\tilde{P}_{1/2}^A = \frac{A + \frac{1}{2}(C + D)}{A + B + C + D}$$

ii. Conditions for equality with the true treatment intent, P

These bounds represent extreme attribution scenarios. The true treatment proportion equals the lower bound when no co-diagnosed claims are attributed to A, and equals the upper bound when all co-diagnosed claims are attributed to A.

Algebraically, in the two-diagnosis scenario,

- $P^A = \tilde{P}_{lower}^A$ if and only if $C = 0$
- $P^A = \tilde{P}_{upper}^A$ if and only if $D = 0$

The reverse holds for B:

- $P^B = \tilde{P}_{lower}^B$ if and only if $D = 0$
- $P^B = \tilde{P}_{upper}^B$ if and only if $C = 0$

Essentially, the bounds cancel to the true treatment proportions when we remove the ambiguity from the co-diagnosed claims. This attribution uncertainty drives the interval width.

iii. Algebraic derivation of equality with P^A

First, we derive when $P^A = \tilde{P}_{lower}^A$:

$$\frac{A + C}{A + B + C + D} = \frac{A}{A + B + C + D}$$

Multiply both sides by the denominator: $A + C = A \rightarrow C = 0$

Next, we derive when $P^A = \tilde{P}_{upper}^A$:

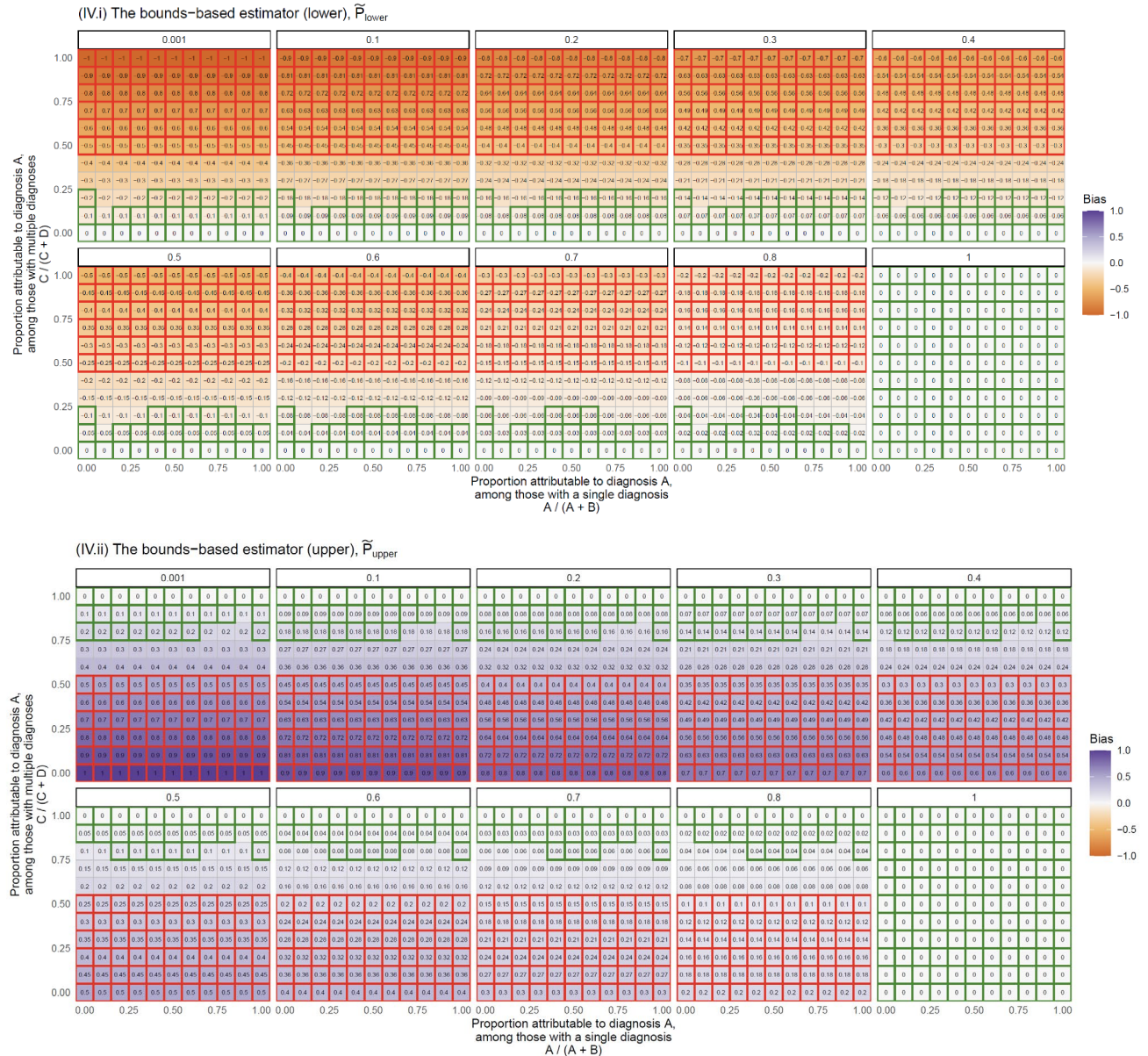
$$\frac{A + C}{A + B + C + D} = \frac{A + C + D}{A + B + C + D}$$

Multiply both sides by the denominator: $A + C = A + C + D \rightarrow D = 0$

iv. Simulation grid examining bias for the bounds-based estimator, $[\tilde{P}_{lower}, \tilde{P}_{upper}]$

Here we present the simulated bias grids for $\tilde{P}_{lower}$ and $\tilde{P}_{upper}$ in the two-diagnosis scenario. The lower bound always either under-estimates or is equal to the true treatment intent for A, while the upper bound always either over-estimates or is equal to the true treatment intent for A, serving as a valid interval for the true treatment intent, P .

Figure 2.7: Simulation of bound-based attribution bias, varying the proportion of claims with only a single diagnosis (0 = all co-diagnosed; 1 = all single-diagnosis)



*Colors represent the bias $\tilde{P}_{single-dx}^A - P^A$. Purple = over-attribution to A; Orange = under-attribution to A; White = no bias (perfect agreement); Green outline = highest performing estimator among the 5 examined (I – III; IV.i, IV.ii); Red outline = lowest performing.

Structural Challenges and Clinical Assumptions

In addition to the algebraic equalities and bias grid shown above, these estimators have some structural and clinical assumptions that affect their validity.

Structural Challenges

Expanding claims into claims-diagnosis rows leads to three structural challenges:

- Double-counting: Co-diagnosed claims contribute to multiple diagnoses
- Inflated denominators: Diagnoses for multi-diagnosis claims get more weight
- Violation of additivity: Each claim no longer contributes one unit of attribution, distorting interpretability of the probabilities

The weighted estimators ($\tilde{P}_{1/n}$ and $\tilde{P}_{single-dx}$) address these structural challenges essentially by assigning fractional weights, ensuring each claim contributes one unit of attribution. In addition, the bounds-based estimator $[\tilde{P}_{lower}, \tilde{P}_{upper}]$ avoids these structural issues by assigning only one diagnosis or the other for all co-diagnosed claims.

Clinical Assumptions

Each estimator makes some assumptions about treatment intent:

- Any listed diagnosis in the processed data is a true indication (assumed by all estimators)
- Any co-diagnosed conditions are equally likely to drive treatment (assumed by $\tilde{P}$, $\tilde{P}_{1/n}$)
- Single-diagnosis patterns are not biased for co-diagnosed attribution ($\tilde{P}_{single-dx}$)
- The bounds-based estimator does not make strong assumptions since there is an interval.

Table 2.4: Structural and clinical assumptions underlying the naïve estimators (I – IV)

Estimator	Structural			Clinical		
	Double-Counting	Inflated Denominator	Additivity Violation	Equal co-diagnoses	Single-dx as prior	No strong assumptions
$\tilde{P}$	☑	☑	☑	☑	--	--
$\tilde{P}_{1/n}$	--	--	--	☑	--	--
$\tilde{P}_{single-dx}$	--	--	--	--	☑	--
$[\tilde{P}_{lower}, \tilde{P}_{upper}]$	--	--	--	--	--	☑

Key takeaway: No attribution estimator performs best across all scenarios—but each performs well under certain conditions. All estimators, except the bounds-based approach, rely on strong and unverifiable assumptions about how treatment intent is distributed. The row-based estimator introduces additional structural challenges by double-counting co-diagnoses and violating additivity, complicating the interpretation of bias. Our simulation framework demonstrates that estimator performance depends on both the prevalence of co-diagnosed claims and the extent to which our assumptions about treatment intent for co-diagnosed claims align with the truth.

Choice of Estimator for this Study

Since our goal is to select a single estimator that can be applied uniformly across all drug–diagnosis pairs, we evaluate estimator performance using **Mean Absolute Error (MAE)** across our full simulation grid. This approach summarizes the average magnitude of bias.

The MAE is calculated as:

$$MAE_e = \frac{1}{N} \sum_{i=1}^N | \tilde{P}_e^{(i)} - P^{(i)} |$$

where:

- $\tilde{P}_e^{(i)}$ is the estimated probability under estimator e at simulation grid point i
- $P^{(i)}$ is the true treatment intent at simulation grid point i
- N is the total number of grid points (i.e., all combinations of the proportion attributable to diagnosis A among single-diagnosis and co-diagnosed claims, across all levels of co-diagnosis prevalence from 0 to 1 in increments of 0.1).

We compute MAE for each estimator:

1. Across the full simulation grid. Because MAE cannot be calculated when the co-diagnosis prevalence is exactly zero, we used a small nonzero value (0.001) as an approximation.
2. At a co-diagnosis rate of 79.1% (i.e., 20.9% single-diagnosis claims), which reflects the average observed in our dataset of 806 drugs (see **Table 2.3**).

This allows us to evaluate which estimator performs best both under general conditions and for our dataset. Results are presented below in **Table 2.5**.

Table 2.5: Mean Absolute Error (MAE) of each estimator vs. true treatment intent, P

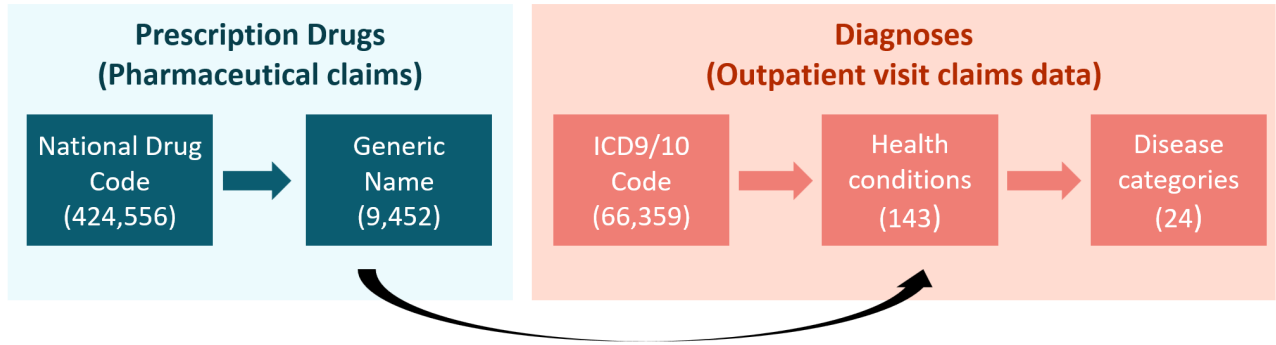
Estimator	Description	MAE (Full Grid)	MAE (20.9% Single-Dx)
$\tilde{P}$	Attributes full co-diagnosed claims to both A and B equally	0.141	0.218
$\tilde{P}_{1/n}$	Assigns equal weight to each diagnosis in co-diagnosed claims (e.g. 0.5, for 2 diagnoses)	0.136	0.216
$\tilde{P}_{single-dx}$	Only includes claims with a single diagnosis	0.182	0.288

Across the full simulation grid, **the equal-weight estimator $\tilde{P}_{1/n}$ exhibits the lowest MAE (0.136)**, slightly outperforming the row-based estimator (0.141), with the single-diagnosis estimator performing worst overall (0.182). At the empirically observed 20.9% single-diagnosis rate, all estimators showed increased error, though the relative performance ranking remains the same. Thus, we use the **equal-weight estimator $\tilde{P}_{1/n}$** for our primary analysis.

VI. Machine Learning Model Framework and Performance

Following data processing and estimator selection, we developed a machine learning model to estimate the probability that a prescription drug, identified by its generic name, was used to treat each of 143 health conditions, defined to comprehensively represent all ICD-9/10 codes as described previously by Dieleman and colleagues.⁸² This approach enabled the generation of a generalizable probabilistic map to link pharmaceutical claims to health conditions. Specifically, machine learning allows us to smooth noisy patterns and predict likely clinical indications by leveraging observed drug features (e.g., dosage, route of administration, spending, etc.), conditional on the assumptions embedded in the selected estimator.

Figure 2.7. Model objective: linking drugs (using their generic names) to one of 143 conditions



As shown in **Figure 2.7**, we structured the model to link each prescription drug, identified by its generic name, to one of 143 health conditions derived from diagnosis codes in medical visit claims. This section details the machine learning models evaluated.

Model Framework

We trained multiple versions of a model to estimate the conditional probability that a given prescription claim for drug d , characterized by a vector of covariates $\mathbf{x}_d = (x_1, x_2, \dots, x_N)$, was used to treat diagnosis k . Separate models were trained for each of 8 mutually exclusive age-sex groups (g). The general form of the model is:

$$\hat{P}_{dk,g} = Pr(Y = k \mid d, \mathbf{x}_d, g)$$

Where:

- $\hat{P}_{dk,g}$: Estimated probability that drug d , in group g , with features $\mathbf{x}_d$, was prescribed for diagnosis k
- d : Generic drug name
- $\mathbf{x}_d$: Vector of additional drug-level features (branded/generic indicator, dosage, package size, and route of administration)
- g : Age–sex group (male/female $\times$ 0–20, 21–45, 46–64, 65+)

We evaluated three model types: (1) a frequency-based model, (2) a multinomial logistic regression model, and (3) a random forest classifier.

1. *Frequency-based model*

The simple frequency-based model is not a machine learning model, but rather an empirical frequency table. Here, the estimated conditional probability $\hat{P}_{dk,g}$ takes the form:

$$\hat{P}_{dk,g} = \frac{N_{dk,g,x_d}}{N_{d,g,x_d}}$$

Where:

- N_{dk,g,x_d} : Number of claims for drug d in group g associated with diagnosis k within covariate stratum x_d
- N_{d,g,x_d} : Total number of claims for drug d in group g within covariate stratum x_d

2. *Multinomial logistic regression*

We implemented a multinomial logistic regression model with a softmax output. No reference category was specified, as the model estimates unnormalized scores for each diagnosis and normalizes them via the softmax transformation across all $K = 143$ diagnoses. The predicted probability is given by:

$$\hat{P}_{dk,g} = \frac{\exp(\alpha_{k,g} + \beta_{k,g}^T x_d)}{\sum_{i=1}^K \exp(\alpha_{i,g} + \beta_{i,g}^T x_d)}$$

Where:

- $\alpha_{k,g}$: Intercept term for diagnosis k , group g
- $\beta_{k,g}^T x_d$: Coefficient vector for diagnosis k , group g corresponding to covariate vector x_d

The model was trained using a Stochastic Average Gradient solver with up to 300 iterations.

3. *Random forest*

Finally, we implemented a random forest classifier, which estimates the predicted probability as the average across T decision trees:

$$\hat{P}_{dk,g} = \frac{1}{T} \sum_{t=1}^T P_{dk,g}^{(t)}$$

Where:

- T : number of trees in the forest (set to 1,000)
- $P_{dk,g}^{(t)}$: Class probability predicted by tree t for group g

Model Cross-Validation and Selection

To evaluate model performance in the absence of a gold standard, we focused on how well the estimated probabilities from the model matched the observed empirical frequencies, rather than on row-by-row classification accuracy.

Data Splitting and Cross-Validation

We randomly split the processed MarketScan dataset into an 80% training set and a 20% test set. To ensure model stability, we repeated this process three times using different random seeds. Performance metrics reported reflect the average across these three runs.

Instead of evaluating binary classification accuracy, we assessed how well the model’s predicted probabilities in the training data matched the observed frequencies in the test data. Specifically, for each age–sex group g , we computed model-predicted probabilities $\hat{P}_{dk,g}$ for all combinations of drug d and diagnosis k , and compared them to empirical frequencies observed in the test set. We used linear regression to evaluate the alignment between these predicted and observed values, across drug–diagnosis–covariate strata.

Performance Metrics

We extracted two goodness-of-fit measures from the regression:

- Root Mean Squared Error (RMSE)
- R^2 (coefficient of determination)

Each metric was calculated in two forms:

- **Unweighted:** treating all drugs equally
- **Weighted by prescription frequency:** giving more weight to commonly prescribed drugs

We pre-specified the frequency-weighted RMSE as our primary performance metric, reflecting the model’s ability to accurately estimate probabilities for the most frequently used drugs.

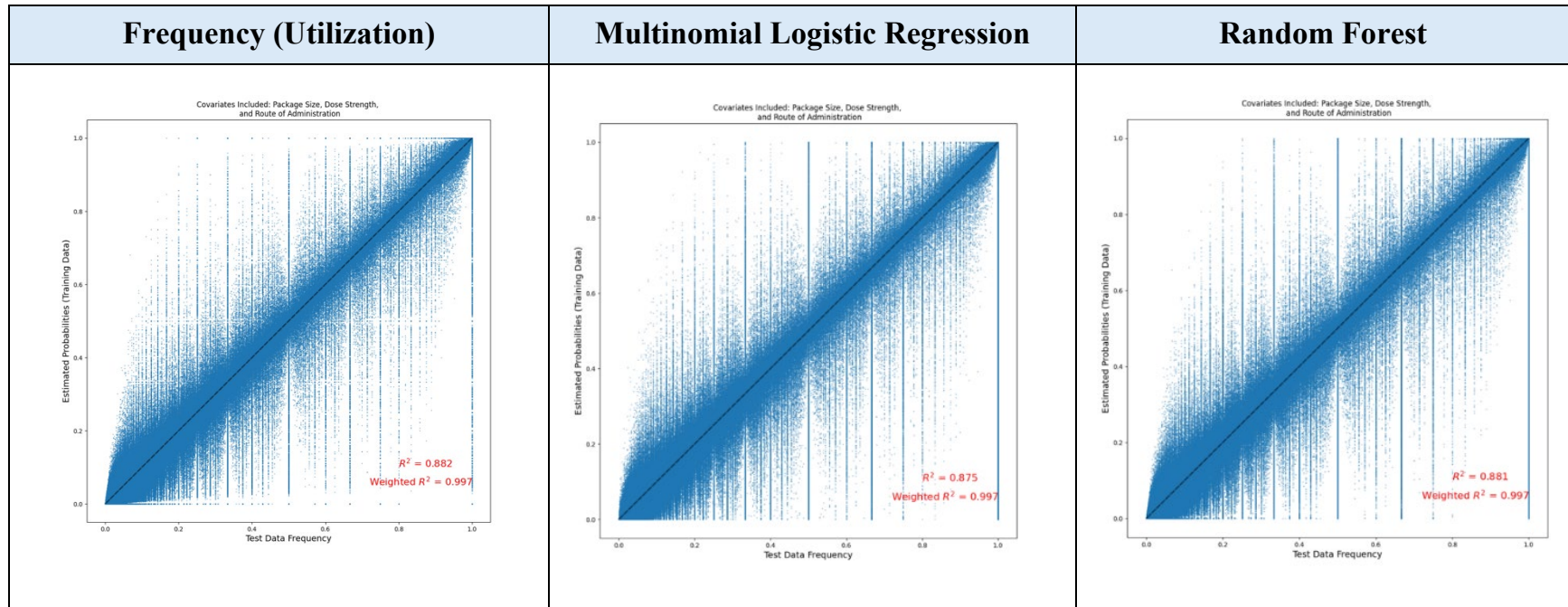
Results Summary

Table 2.6 shows performance across the three models. All models performed similarly across RMSE and R^2 , particularly when weighted by prescription frequency.

Table 2.6. Comparison of model fit metrics across models (average across 3 runs)

Metric	Weighting	Frequency Map	Multinomial Logistic Regression	Random Forest
<i>RMSE</i>	<i>Frequency</i>	<i>0.017</i>	<i>0.017</i>	<i>0.017</i>
RMSE	Unweighted	0.103	0.107	0.105
R^2	Frequency	0.997	0.997	0.997
R^2	Unweighted	0.883	0.875	0.881

Figure 2.8. Predicted versus observed probabilities across model types in cross-validation (average across 3 runs)



Notes:

- Scatterplots compare the predicted probabilities from each model (y-axis) to the observed test set frequencies (x-axis).
- Each point represents a drug–diagnosis–covariate group combination.
- Weighted R^2 emphasizes model accuracy for high-volume prescriptions.
- All three models exhibit strong agreement with observed data.

Figure 2.8 further visualizes these comparisons. Notably, the simple frequency-based model performed nearly as well as the machine learning models, despite its simplicity.

Model Selection

Given the comparable performance, we selected the **random forest model** for downstream analyses. Unlike logistic regression, random forests do not assume a logit-linear relationship between drugs and diagnoses, making them more flexible for this probabilistic mapping task.

2.3 Benchmarking of tracer drugs with NHANES

Across all drugs, we calculated the marginal probabilities for further presentation and benchmarking. Specifically, this reflects, for each health condition, the expected fraction of claims used to treat that condition—computed as a weighted average across all drug claims. For a given drug d and condition k , the marginal probability is:

$$P(k \mid d) = \frac{1}{N_d} \sum_{i=1}^{N_d} \hat{P}_{i,k}$$

Where:

- N_d is the total number of claims for drug d
- $\hat{P}_{i,k}$ is the predicted probability that claim i was for condition k

Table 2.7. Tracer drugs, common name, and reasoning for selection

Tracer Drug	Therapeutic Class	Common Brand Name(s)	Rationale
Adalimumab	Immunosuppressant (TNF inhibitor)	Humira	Among the highest-cost drugs; prescribed for multiple autoimmune diseases; tests model performance for low volume, high-cost drug.
Albuterol Sulfate	Bronchodilator (Beta-agonist)	ProAir, Ventolin, Proventil	Used for asthma and COPD; closely linked to respiratory conditions; check of accuracy for unambiguous indications.
Atorvastatin Calcium	Lipid-lowering agent (Statin)	Lipitor	Very high volume and clearly indicated for hyperlipidemia / prevention of heart disease.
Buprenorphine/Naloxone	Opioid dependence treatment	Suboxone	Used for opioid use disorder; patients often have complex co-morbidities; public health relevance
Fluoxetine Hydrochloride	Antidepressant (SSRI)	Prozac	Commonly used for both depression and anxiety, tests model performance for mental health diagnoses with high co-occurrence.
Gabapentin	Neuropathic pain / Anticonvulsant	Neurontin	Prescribed for a wide range of off-label conditions; tests how models handle diffuse and overlapping prescribing intent.
Insulin Glargine, Recombinant	Antidiabetic (Long-acting insulin)	Lantus, Basaglar	Prescribed for both type 1 and type 2 diabetes
Metformin Hydrochloride	Antidiabetic (Biguanide)	Glucophage	High utilization, low cost, typically indicated for type 2 diabetes.

Out of 806 total drugs, we selected eight **tracer drugs**. These tracer drugs were chosen to represent a variety of prescribing patterns, therapeutic areas, and levels of utilization and spending. **Table 2.7** summarizes the selected tracer drugs and their rationale for inclusion. Across the eight tracer drugs, we compared across six estimation approaches:

1. Frequency Map (utilization-based; row-based estimator)
2. Multinomial Logistic Regression (row-based estimator)
3. Random Forest (row-based estimator)
4. Random Forest (equal-weighted estimator)
5. Random Forest (single-diagnosis estimator)
6. NHANES (external benchmarking; row-based estimator)

Figure 2.9 shows the estimated marginal probabilities for tracer drugs across models at the condition level. The cause-level version is provided in **Appendix S2**.

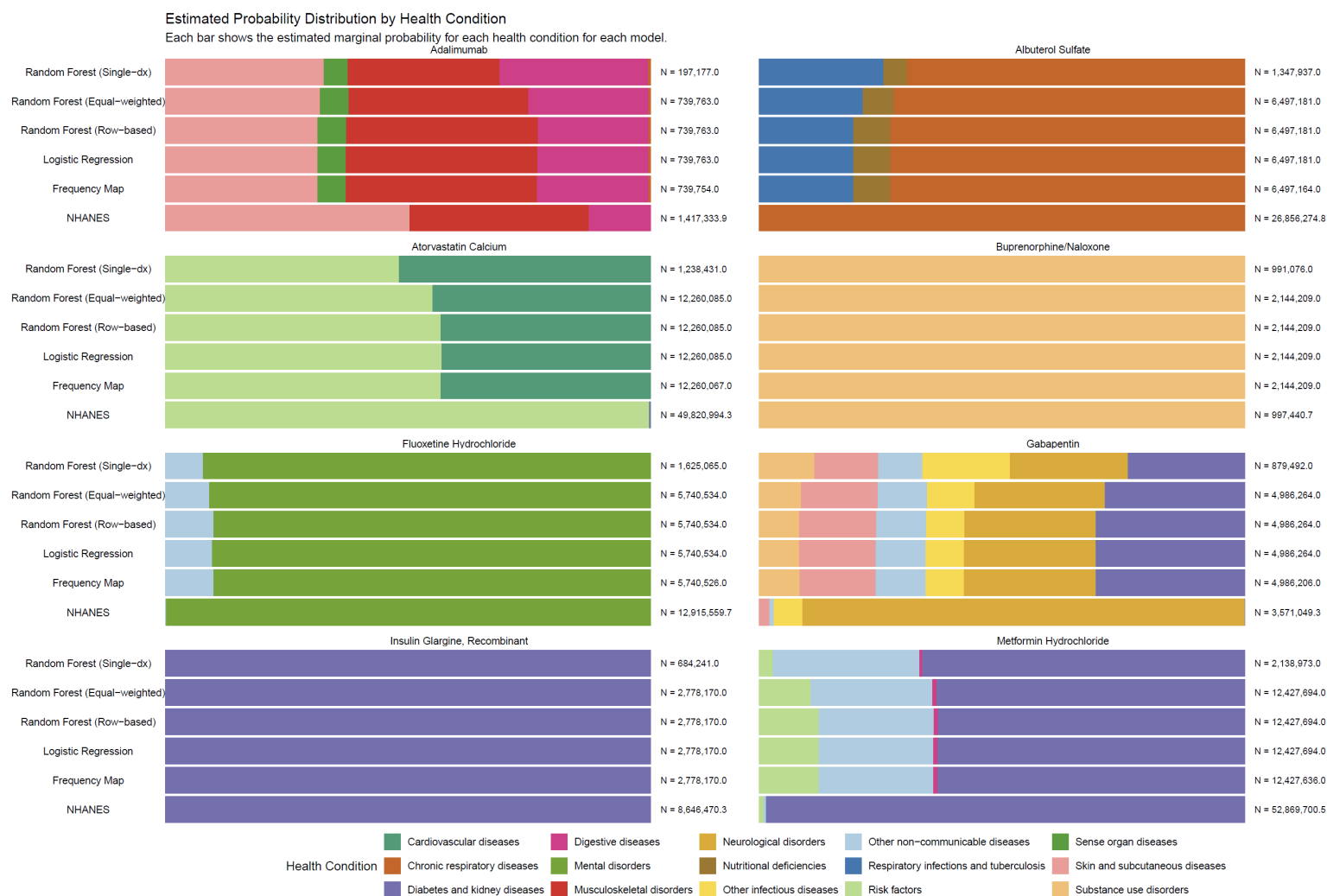
First, we see that using MarketScan data allows for a broader set of conditions to be examined—particularly for low-utilization, multi-indication drugs such as adalimumab (Humira) and gabapentin—compared to NHANES, which often has limited sample sizes for these medications. In some cases, sample sizes are so small that they fail to capture the full range of health conditions for which the drug is indicated.

Second, we observe that the choice of model—logistic regression, random forest, or simple frequency map—has relatively little effect on the estimated marginal probabilities averaged across all claims. This is somewhat expected, as the main advantage of machine learning models is in capturing covariate-driven complexities at the individual level, not at the aggregate level. It may also reflect that all models perform similarly, as suggested by our cross-validation results.

Third, we find that the choice of estimator—row-based, single-diagnosis, or weighted—can substantially influence the results. This is especially true for when co-morbidities are common and diagnosis coding practices vary for the single-diagnosis compared with co-diagnosed claims.

Finally, the choice of restrictions for the set of plausible indications appears to exert the strongest influence on the final probabilities. For example, applying Micromedex restrictions at the DEX cause level removes many implausible mappings but may fail to eliminate spurious links within broad categories. A more detailed description of each tracer drug is given below.

Figure 2.9: Estimated marginal probabilities for tracer drugs across models (at the health condition category level)



Adalimumab [Humira]

Humira, one of the highest-cost specialty drugs, has low volume but is prescribed for a wide range of conditions. On-label uses listed in Micromedex include rheumatoid arthritis, Crohn's disease, sarcoidosis, and uveitis, as well as on- and off-label uses for psoriasis. NHANES, with only 40 observations from 2013–2023, assigns Humira to skin, musculoskeletal, and digestive disorders—missing additional known uses like sense organ diseases (e.g., uveitis). In contrast, the MarketScan-based models leverage hundreds of thousands of claims to capture a more complete distribution of indications. Of note, the single-diagnosis model attributes a higher share to digestive diseases (e.g., Crohn's, ulcerative colitis), while models incorporating co-diagnoses distribute a greater share to skin, sense organ, and musculoskeletal conditions. Overall, this highlights the value of large-scale administrative claims data for characterizing real-world prescribing patterns, especially for expensive, low-frequency drugs like Humira.

Albuterol Sulfate Inhaler

Albuterol is a bronchodilator approved for asthma and chronic obstructive pulmonary disease (COPD), with off-label uses including coronavirus and hyperkalemia (high potassium). In NHANES (2012–2013), it is assigned only to asthma and COPD. MarketScan captures additional COVID-related use during 2020–2022, mapped under respiratory infections. MarketScan models also attribute 4–8% of use to nutritional deficiencies—likely an overestimate, as this broad category may reflect common co-morbidities rather than true treatment for hyperkalemia, a rare, typically inpatient use. This potential over-attribution is lower in the single-diagnosis model. These findings highlight MarketScan's ability to capture emerging indications like COVID-19, while also revealing sensitivity to condition granularity and Micromedex constraints, which are applied at the DEX condition level rather than the ICD level. As a result, some potentially spurious diagnosis codes which fall within the same DEX condition may be inadvertently included.

Atorvastatin Calcium [Lipitor]

Atorvastatin is a widely prescribed lipid-lowering agent, approved for hyperlipidemia and cardiovascular event prevention, with off-label uses including generalized atherosclerosis and heart failure. In NHANES, it is assigned exclusively to hyperlipidemia. In contrast, the MarketScan model assigns it to both hyperlipidemia and heart disease. This highlights the model's sensitivity to mapping decisions—specifically, whether preventive drugs should be attributed to the condition they prevent or limited strictly to their labeled preventive indication.

Buprenorphine/Naloxone [Suboxone]

Buprenorphine/Naloxone, a schedule III controlled substance which treats narcotic dependence, is assigned exclusively to opioid use disorder in both NHANES and the MarketScan models. However, MarketScan draws on over 2 million claims (or nearly a million for the single diagnosis

model), while NHANES relies on just 20. Notably, before applying Micromedex restrictions, 1 of 21 NHANES respondents (~5%) reported using it for low back pain—illustrating the limitations of self-reported data, especially for stigmatized conditions like substance use. While MarketScan offers the advantage of a much larger sample size (and thus less stochastic noise), both datasets are subject to potential systematic biases. MarketScan lacks representativeness, as it includes only the privately insured population, while NHANES is nationally representative but vulnerable to misreporting due to social desirability bias.⁹⁷

Fluoxetine Hydrochloride [Prozac]

Prozac is an antidepressant approved for a range of mental health conditions, including anxiety, depression, post-traumatic stress disorder, bulimia nervosa, and obsessive-compulsive disorder. It is also used on- and off-label for premenstrual dysphoric disorder and hot flashes, both categorized under gynecologic disorders. In NHANES, all reported uses of Prozac are attributed to mental health conditions. In contrast, the MarketScan models assign some use to gynecologic disorders—likely due to the DEX condition-level restrictions, which allow for mapping to unrelated comorbidities within that category. This may lead to over-assignment of Prozac use to gynecologic categories, despite mental health being the primary indication in most cases.

Gabapentin

Gabapentin is commonly prescribed to treat nerve pain and is approved for epilepsy (as an anticonvulsant) and postherpetic neuralgia. It also has a range of off-label uses, including neuralgia, hot flashes, pruritus (itchiness), diabetic neuropathy, and alcohol dependence. In NHANES, among 155 respondents reporting Gabapentin use, 74% linked it to neurologic conditions, 14% to epilepsy, 9% to postherpetic neuralgia, and <5% to other conditions such as diabetic neuropathy and pruritus. The MarketScan models, leveraging over 4 million claims, assign use more diffusely across these categories—likely reflecting a mix of true clinical variation and noise from co-morbid conditions. Notably, the single-diagnosis model attributes a higher share to alcohol use disorders and infectious conditions (e.g., postherpetic neuralgia), and a lower share to diabetes—suggesting that models incorporating co-diagnoses may over-assign to diabetes due to its high background prevalence. This highlights the challenge of distinguishing true indications from correlated diagnoses, particularly for drugs with diverse usages.

Insulin Glargine, Recombinant

Insulin Glargine is used to treat both type 1 and type 2 diabetes. While type 2 diabetes is far more prevalent, type 1 patients typically generate higher prescription volume due to lifelong dependence on insulin. Appendix A3 shows the version of **Figure 2.9** further disaggregated by health condition. In NHANES, 97% of the 464 respondents using Insulin Glargine reported it for type 2 diabetes, with only 3% reporting use for type 1. In contrast, the MarketScan models assign 60–70% of use to type 2 diabetes, with the remainder to type 1. The single-diagnosis model shows the

lowest share for type 2 (60.7%), suggesting that single-diagnosis claims may not fully capture the distribution seen in co-diagnosed claims.

These differences likely reflect underlying contrasts in the populations: MarketScan claims reflect prescription volumes, while NHANES represents person-level self-reported use. Thus, NHANES may underrepresent type 1 in terms of prescriptions (since type 1 patients have higher prescription volume). Thus, MarketScan may better capture prescribing intensity across real-world settings.

Metformin Hydrochloride

Metformin is a widely used, low-cost medication primarily indicated for type 2 diabetes, with FDA approval for both type 2 diabetes and, more recently, chronic kidney disease. Off-label, it is commonly used to treat polycystic ovary syndrome (PCOS), obesity, and cirrhosis—particularly among younger adults and women of reproductive age. In NHANES, 80.6% of respondents reported using Metformin for type 2 diabetes and 18.6% for elevated blood glucose, with almost no reported use for other conditions. These self-reports may underrepresent valid off-label uses, particularly in younger women. In contrast, the MarketScan models—drawing on over 20 million pharmacy claims—capture broader use patterns, including assignment to gynecologic conditions (e.g., PCOS) and obesity. However, the single-diagnosis model rarely attributes use to obesity alone, suggesting that co-diagnosed models may over-assign to obesity when it co-occurs with diabetes or PCOS. This reflects both the model’s strength in detecting diffuse, real-world prescribing patterns and its sensitivity to common co-morbidities.

Summary of Benchmarking Exercise

The comparison of models and benchmarking against NHANES highlights both the strengths and limitations of the MarketScan modelling approach. First, MarketScan incorporates far more data than a survey-based approach, including millions of claims and the ability to integrate covariates—producing more granular estimates. Second, it reflects real-world prescribing behavior, which is often the target of analysis, rather than population-level disease prevalence that is disconnected from prescription volume (as illustrated in the insulin glargine example). Third, MarketScan avoids the pitfalls of self-report, which can be biased in terms of both medication recall and indication disclosure—especially for stigmatized conditions.

However, the approach has challenges. In addition to the assumptions required for handling multi-diagnosis claims, when Micromedex restrictions are applied at the condition category level rather than the ICD level, the model may capture unrelated co-morbidities, leading to misattribution—particularly for drugs with diffuse off-label use or common co-occurring conditions. Ultimately, while MarketScan supports more precise and scalable drug-to-diagnosis mapping, careful examination of restrictions and attribution logic is critical.

VII. Results

This section presents the primary results from the random forest model using the equal-weighted estimator, applied to the MarketScan commercially insured population from 2010 to 2022.

Figure 2.10 displays the health condition categories assigned to the top 30 drugs by spending (left) and by prescription volume (right) across 13 years of MarketScan data. Among the highest-spending drugs, the top three health conditions (out of 143) by pharmaceutical spending were:

- Type 2 diabetes: averaging \$1.75 billion annually, driven by drugs such as Insulin Glargine, Dulaglutide, Insulin Aspart, Sitagliptin, Semaglutide, Insulin Lispro, Liraglutide, Empagliflozin, and glucose meter test control strips.
- Attention-deficit/hyperactivity disorder (ADHD): averaging \$666 million annually, including Lisdexamfetamine Dimesylate (Vyvanse), Amphetamine Salt Combination (Adderall), and Methylphenidate Hydrochloride (Ritalin/Concerta).
- Skin and subcutaneous diseases: averaging \$650 million annually, with top drugs including Adalimumab (Humira), Etanercept (Enbrel), and Fluticasone Propionate.

Among the highest-volume drugs, the top three health conditions by prescription fills were:

- Upper respiratory infectious: averaging 7.7 million prescription fills annually, driven by drugs such as Azithromycin, Amoxicillin, Prednisone, Methylprednisolone, Cephalexin, and Albuterol Sulfate.
- Endocrine, metabolic, and immune disorders: averaging 5.8 million prescription fills annually, largely by Levothyroxine Sodium, which treats thyroid disorders.
- Skin and subcutaneous diseases: averaging 5.6 million prescription fills annually, driven by drugs such as Cephalexin, Fluconazole, and Fluticasone Propionate.

Figure 2.11 shows, for the same set of drugs, the estimated distribution of retail pharmaceutical use—by spending (left) and volume (right)—disaggregated by FDA label status (on- vs. off-label) and by whether the drug was branded or generic.

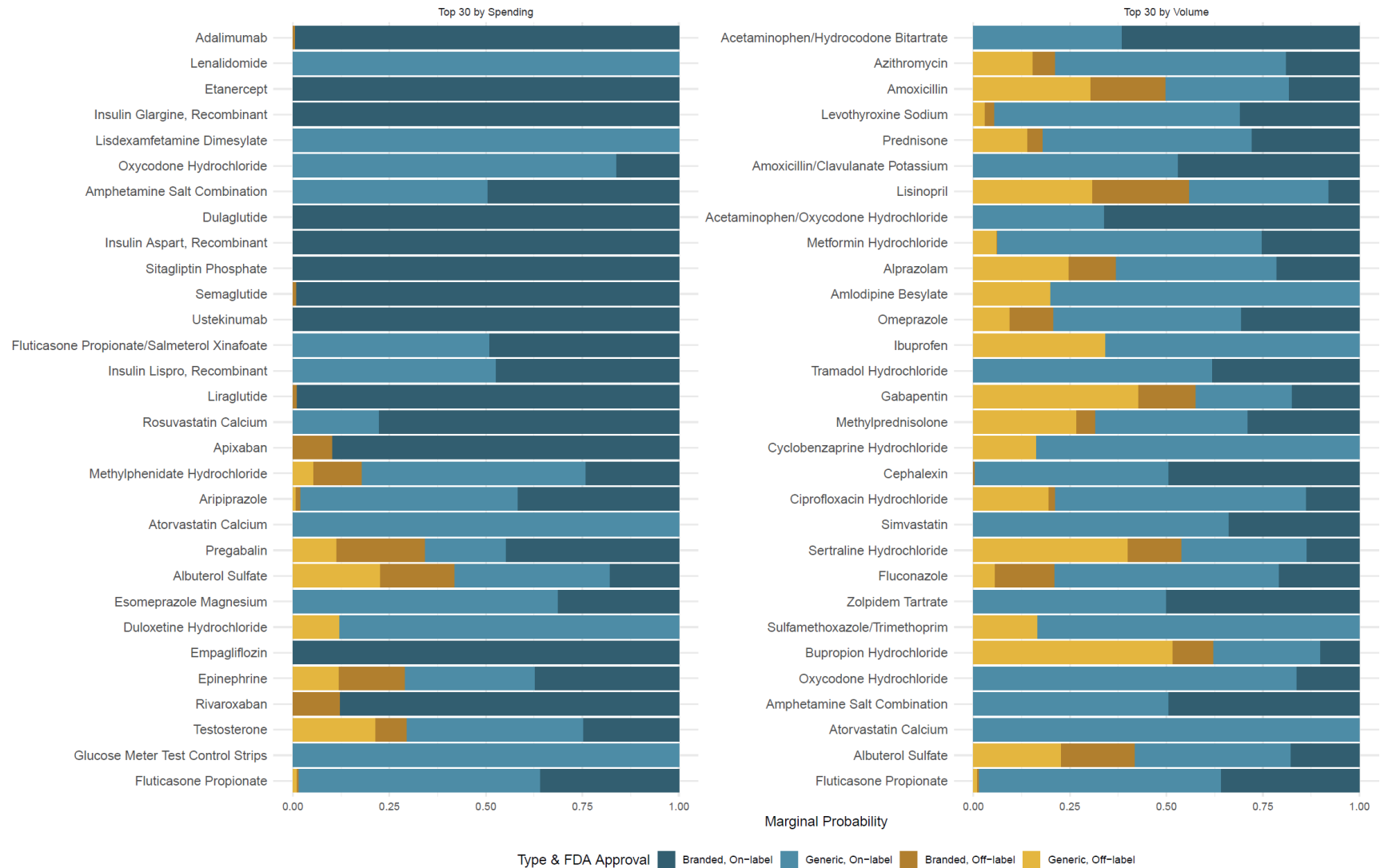
In general, high-volume drugs show a greater share of off-label and generic use, while high-spend drugs tend to be branded and more likely to be used on-label. In addition, there is more generic off-label use than branded off-label use across this set of drugs. However, there are notable exceptions—such as Albuterol Sulfate, which appears in both the high-spend and high-volume categories and shows substantial off-label use, including among branded products.

There is relatively little off-label use among the highest-spend drugs, which may reflect stronger manufacturer incentives to pursue FDA approval for additional indications, particularly for blockbuster drugs with large market potential.

Figure 2.10: Distribution of health conditions for top 30 retail drugs by spending and volume, 2010–2022



Figure 2.11: Distribution of off-label and generic prescribing for top 30 retail drugs by spending and volume, 2010–2022



Discussion

In an ideal world, every prescription claim would include its clinical indication, carried through from the electronic health record to the pharmacy claim. In the absence of this information, this study demonstrates that it is feasible and valuable to use a probabilistic model to map each drug to the full range of conditions for which it is likely prescribed, using large-scale claims data. The final model we develop, using a random forest with an equal-weighted estimator, generates condition-specific prescribing estimates that are both clinically plausible and analytically useful.

Applications

Our methodology was designed with practical applications in mind. It has already been implemented as part of the US Disease Expenditure (DEX) project, which models spending by health condition across geography, demographic groups, and time.^{82,83} Additionally, it serves as the foundation for work quantifying inequities in pharmaceutical utilization and spending by race and ethnicity—an effort that would not be possible without first mapping drugs to their likely indications. By enabling condition-level estimates of drug use, our model supports disease-standardized comparisons that are critical to equity-focused policy and research.

This work also makes a foundational contribution by addressing a long-standing limitation in the use of retail pharmaceutical claims for pharmacoepidemiology and pharmaceutical policy research. The model allows for probabilistic attribution of spending and utilization data to specific conditions, providing a clearer picture of real-world prescribing, including the measurement of off-label prescribing which is widespread but not well understood.⁹⁸

Limitations

Despite its strengths, the model has key limitations. First, there is no gold standard providing the true clinical indication for pharmaceutical claims data. Validation relies on indirect benchmarks, including internal cross-validation and comparison with NHANES—a self-reported survey covering a different population, limiting its comparability. Attribution is also sensitive to the granularity of diagnostic categories. Additionally, the Micromedex restrictions used were captured at a single time point, preventing assessment of changes in labeling or clinical evidence over time.

Recommendations for future iterations of this model include: (1) additional clinical vetting of top drug-condition pairs; (2) mapping Micromedex’s plain-language indications to ICD codes rather than broad DEX categories to reduce misattribution from co-morbidities within the DEX cause; and (3) expanding beyond the current set of 806 drugs to the full set of 9,452 drugs.

Conclusion

We develop a mapping approach that leverages large-scale pharmaceutical claims data to link prescriptions to their likely clinical indications—unlocking a vast repository of data for use in health equity, regulatory, and condition-specific pharmacoepidemiologic analysis.

Chapter 3 Prescription Drug Utilization and Spending by Race, Ethnicity, Payer, Health Condition, and US State

This chapter was recently accepted for publication as:

Sahu M, Wagner TD, Thomson A, Beauchamp M, Campbell JD, Crosby S, DeJarnatt D, Lescinsky H, Salih RK, Taylor K, Weil M, Dwyer-Lindgren L, Haakenstad A, Scott JW, Stergachis A, Essien UR, Dieleman JL. *Prescription drug utilization and spending by race, ethnicity, payer, health condition, and US state. JAMA Health Forum*. Accepted May 5, 2025.

Key Points

Question

How do retail prescription drug utilization and spending vary by race, ethnicity, health condition, payer, and US state?

Findings

In this 2019 cross-sectional study of 143 health conditions across all US states, per capita pharmaceutical use was highest among White populations and lowest among Asian/Pacific Islander and Hispanic populations. However, after standardizing for age and disease prevalence (for 52 conditions with available data), prescription fills were substantially lower for Black populations relative to the all-population mean.

Meaning

Disparities in medication use relative to disease burden persist – especially underutilization among Black populations – highlighting the need for targeted efforts to advance pharmacoequity.

Abstract

Importance: Achieving equitable access to medicines requires understanding how pharmaceutical use and spending vary by race and ethnicity across the US.

Objective: To quantify variation in prescription drug utilization and spending per capita and per prevalent case for the 2019 US population by race, ethnicity, health condition, payer (Medicare, Medicaid, private, and uninsured), and state.

Design: We extended the US Disease Expenditure Project to incorporate disaggregation by race and ethnicity for state-level retail prescription drug utilization and spending—in addition to 143 health conditions, 38 age/sex groups, and four payers (Medicare, Medicaid, private insurance, and out-of-pocket). We estimated prescriptions dispensed and spending for retail pharmaceuticals. four mutually exclusive racial and ethnic groups (Asian/Pacific Islander, Black, Hispanic, and White). We standardized estimates by population size, population age, and—where data permitted—by disease burden (52 conditions). We used Das Gupta decomposition to estimate the relative contribution of three factors (disease prevalence, prescriptions per prevalent case, and spending per prescription) on observed disparities in age-standardized per-capita pharmaceutical spending.

Results

In 2019, age-standardized pharmaceutical utilization and spending per person with a given disease was substantially lower than the all-population mean for Black populations, close to the mean for Hispanic populations, and often higher than the mean for Asian/Pacific Islander and White populations. These trends—particularly those for the Black population—were generally consistent across 52 health conditions, but varied widely across payers and US states. The decomposition analysis for these 52 conditions showed that differences in per-capita pharmaceutical spending across race/ethnicity groups were primarily explained by disease prevalence for Black populations (driving increased per-capita spending), and by utilization rates per prevalent case for Hispanic populations (also driving increased spending). In contrast, differences in drug price or product type (spending per prescription) contributed less to observed spending disparities.

Conclusions and Relevance

Racial and ethnic disparities in medication use persist, most notably the underutilization of medicines relative to disease burden among Black populations. These patterns vary by state, highlighting the need for local- and condition-specific approaches to advancing pharmacoequity in the US.

Introduction

Life expectancy in the United States (US) varies considerably by race and ethnicity: in 2019, Black and American Indian/Alaska Native individuals experienced life expectancies nearly four and seven years shorter than the national average, respectively.^{17,18} Prescription medications have contributed substantially to longevity, accounting for over one-third of the overall increase in US life expectancy between 1990 and 2015.²² However, long-standing racial and ethnic disparities in access to medicines have been documented.^{16,20,21} These inequities can contribute to suboptimal management of health conditions and, in turn, more severe health outcomes—such as the disproportionately high rates of lower-limb amputations observed among Black individuals with diabetes.^{27,28} Accordingly, the concept of *pharmacoequity*, defined as “ensuring that all individuals, regardless of race and ethnicity, socioeconomic status, or availability of resources, have access to the highest-quality medications,” has gained attention in healthcare discussions.^{16,99}

Ensuring equitable access to medicines begins with understanding how pharmaceutical utilization and spending vary across racial and ethnic groups, and identifying the mechanisms driving these differences. While health insurance claims data and national surveys offer valuable insights into patient access, drug pricing, and financial burden,¹⁰⁰ prior studies using these datasets for disparities research have typically focused on payer-specific prescription drug spending,^{101–103} or national-level health expenditure.²⁶ To date, no analysis has systematically evaluated disparities in pharmaceutical use and spending across race, ethnicity, payer, health condition, and US state. In addition, earlier work did not account for key differences across groups—such as variation in population age structure, disease burden, and regional demographic distribution (e.g, across states)—which could bias or obscure disparities. Addressing these limitations requires a more granular, standardized approach which accounts for both demographic and epidemiologic variation across racial and ethnic groups and geography.

To address these gaps, the present study leverages a novel dataset that integrates and standardizes health spending data to be representative across demographics (age, sex, race, and ethnicity), payers, 143 health conditions, and all US states, —enabling a deeper level of granularity and comparability for measuring racial and ethnic disparities in access/uptake and spending on medicines.⁸² Our study objectives were to (1) estimate pharmaceutical utilization and spending in 2019 for five race and ethnicity groups, four payers, and 143 health conditions across all US states; (2) standardize race-specific pharmaceutical utilization and spending data by age and, for 52 conditions with available prevalence or incidence data, by disease burden; and (3) examine how racial and ethnic differences in age-standardized pharmaceutical spending per capita are driven by variation across groups in disease burden, prescription fills per prevalent case (reflecting access, uptake, and adherence), and spending per prescription (reflecting prices and product mix).

Methods

This study used outputs from the US Disease Expenditure (DEX) project, which produced estimates of US health care spending and utilization for 143 health condition categories –

comprehensive of all spending – at the national, state, and county levels for 2010 to 2019.⁸² The DEX project integrated a broad set of health insurance, survey, and population data, which were scaled to align with official state-level spending totals.¹⁰⁴ In a companion study, national- and state-level DEX estimates (across all types of care) were further disaggregated for five mutually exclusive racial and ethnic groups: Hispanic, and (all non-Hispanic) American Indian/Alaska Native, Asian/Pacific Islander, Black, and White populations.¹⁰⁵ The present study focused on 2019 retail prescription drug utilization and spending by race and ethnicity, reported per capita and per prevalent case, including a more granular assessment across four payers, 143 health conditions (including 52 with available disease burden data), and 50 US states and DC.

Data

Prescription drug utilization and spending

The DEX project leveraged over 27.3 billion public and commercial claims and survey retail pharmaceutical data from 2010 to 2019, including Medicare Part D, Medicaid, private insurance (MarketScan, Kythera, and Health Care Cost Institute), and the Medical Expenditure Panel Survey (MEPS) – a national-level survey covering all payers. Details on these datasets and how they were each used are provided in Appendix S1. Each prescription drug dataset was linked to beneficiary demographic data, including age and sex, and—where available—self-reported race and ethnicity. Race/ethnicity data were available in four pharmaceutical datasets, as detailed in Appendix S1.2, each with varying degrees of missingness: Medicare (3%), Medicaid (18%), Kythera (88%), and MEPS (0.5%). For Medicare and Medicaid, we imputed missing race/ethnicity as described in S1.2. Given the high level of missingness for Kythera, we did not perform imputation and instead relied on the 12% of records with available information. Nonetheless, our private insurance estimates included over 20 million records with usable race/ethnicity from Kythera, supplemented with nationally representative survey data from MEPS. We modelled estimates by race/ethnicity as described below and in Appendix S2.

This study included prescribed retail pharmaceuticals dispensed through community pharmacies, clinic pharmacies, mail-order pharmacies, and other outpatient settings—representing approximately 70% of total pharmaceutical spending in 2019.⁹⁵ It excluded inpatient-administered medications, office-administered drugs (e.g., chemotherapy), and institutional use such as long-term care. Data from the Indian Health Service, Tricare, and Veteran Affairs were not available and excluded. Because approximately 64% of American Indian/Alaska Native individuals receive care through the Indian Health Service,¹⁰⁶ we did not report race-specific estimates for this group except in the most aggregated result (**Figure 3.1**). However, they were included in the calculation of all-population averages.

Spending represented total expenditure across all payers and disaggregated by payer (including out-of-pocket spending). Utilization was measured by the number of prescriptions dispensed (i.e., fills).

Disease burden

For 52 health conditions, state-level disease burden by race and ethnicity was modeled using race-specific mortality data.⁸⁴ We then applied estimated incidence-to-mortality ratios for cancers and prevalence-to-mortality ratios for other conditions, as described in Appendix S1.3 and elsewhere.¹⁰⁵ For simplicity, we refer to this modeled burden as ‘prevalence’ or ‘prevalent cases’ throughout the manuscript.

Population denominators

To calculate per capita estimates, we used state population data by race and ethnicity from the US Census Bureau and the National Center for Health Statistics bridged-race population estimates.^{107,108} For payer-specific estimates, we used insurance coverage by race and ethnicity from the American Community Survey and State Health Access Data Center (SHADAC), as detailed in Appendix S1.4.¹⁰⁹

Analysis

Modelling

Pharmaceutical data underwent four modeling steps to generate comprehensive estimates of retail prescription drug utilization and spending, stratified by US state, health condition, payer, year, age category, sex, race, and ethnicity of the patient. These estimation methods are described in greater detail in Appendix S2 and elsewhere.^{82,105} Briefly, these steps included (1) disaggregating data by self-reported race/ethnicity with imputation for missing values, (2) linking drugs to health conditions using a random forest model and Merative Micromedex records,⁹⁴ (3) estimating utilization and spending rates through small area modeling, and (4) adjusting results to align with CMS State Health Expenditure Accounts (SHEA), which estimate prescription drug spending net of rebates and pharmacy discounts.^{104,110}

Uncertainty and Significance

In Step 3, we simulated 50 draws from a multivariate normal distribution. Calculations in Step 4 were performed separately for each draw. Then, estimates were calculated as the mean of these draws with 95% Uncertainty Intervals [UIs] calculated from the 2.5th and 97.5th percentiles. Race-specific estimates were considered statistically significantly different from the weighted all-population mean (calculated across racial and ethnic groups) when the 95% UI of the estimate did not overlap with the weighted mean.

Decomposition Analysis

For 52 health conditions with existing estimates of disease prevalence by race and ethnicity (Appendix S1.3), we examined drivers of differences in pharmaceutical spending by race/ethnicity using Das Gupta decomposition.^{111,112} Specifically, we decomposed differences between national-level spending (for the entire US population) and age-standardized race-specific spending (for each

race/ethnicity group) into three factors: (1) disease prevalence, (2) prescriptions per prevalent case, and (3) spending per prescription. We algebraically estimated the relative contribution of each factor to these differences:

$$\frac{Spending_d}{Population} = \frac{Prevalent\ Cases_d}{Population} * \frac{Prescription\ Fills_d}{Prevalent\ Cases_d} * \frac{Spending_d}{Prescription\ Fills_d}$$

where d is the health condition (52) and spending is specific to retail pharmaceuticals only. This analysis was conducted separately at national- and state-levels, and for each of the 52 health conditions.

Standardization

All comparisons (e.g., across race groups or states) were age- and sex-standardized using direct standardization with either the national population or payer-specific population as the reference group.

Reporting Standards and Ethics

This study followed the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER),¹¹³ as detailed in Appendix S5. Institutional Review Board (IRB) approval was not required because only aggregate, nonidentifiable data were used.

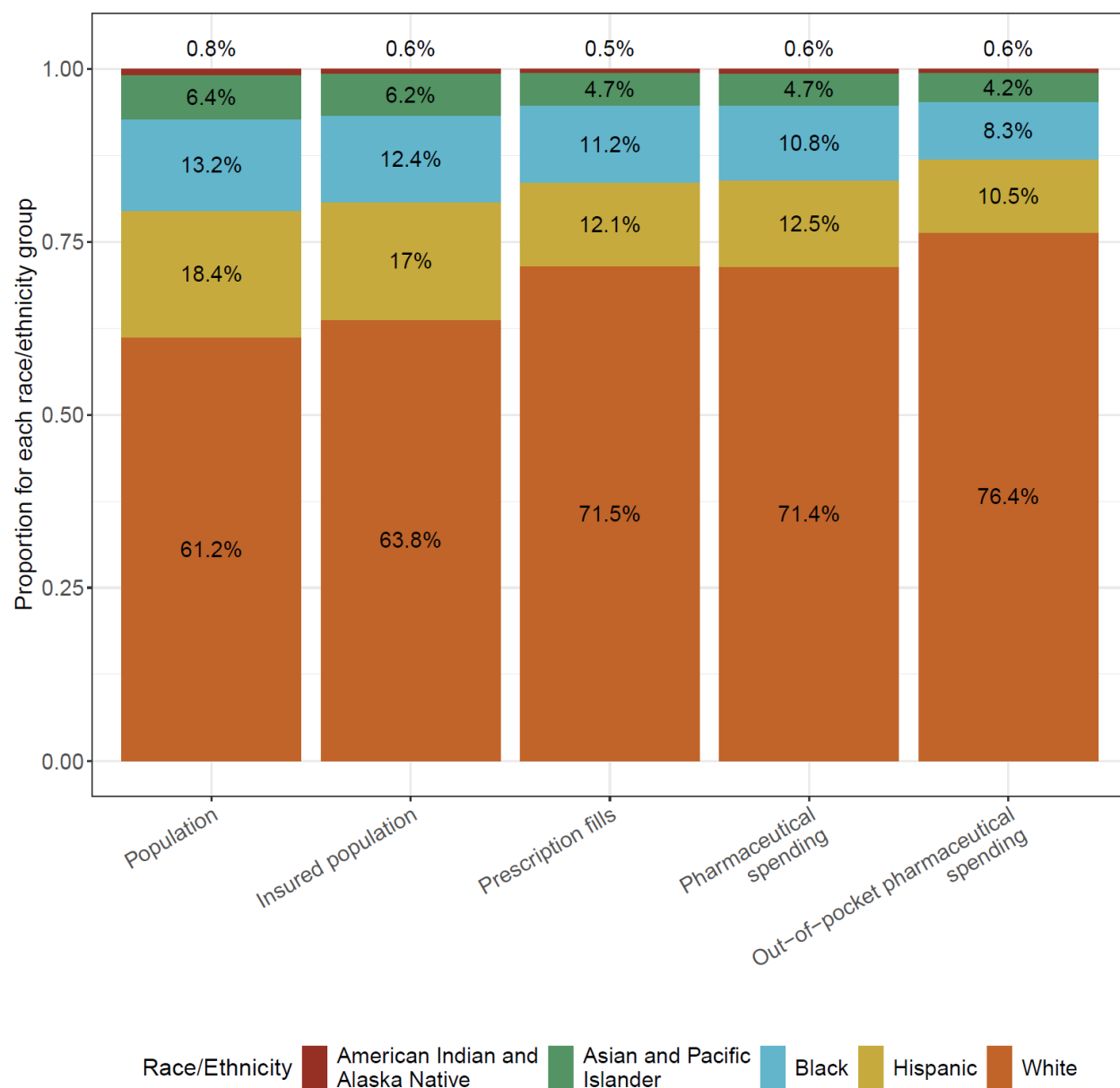
Results

We report retail pharmaceutical utilization and spending by race and ethnicity, organized as follows: (1) unadjusted national patterns, (2) age-standardized results, overall and by payer, (3) condition-specific trends, (4) geographic variation, and (5) decomposition of spending. Condition-specific and geographic results are age- and prevalence-standardized. Supplementary results are in Appendices S3-S4.

1. Unadjusted national patterns

In 2019, an estimated 6.7 (95% Uncertainty Interval [UI]: 5.9–7.5) billion prescriptions were filled, and \$331.4 (\$238.3–\$460.2) billion was spent on prescribed retail pharmaceuticals, including \$46.8 (\$35.4–\$62.3) billion in out-of-pocket spending. Before adjusting for age or disease prevalence, the 61.2% of the US population identifying as White accounted for disproportionately higher insurance coverage, prescription fills, total pharmaceutical spending, and out-of-pocket pharmaceutical spending (**Figure 3.1**). In contrast, Hispanic, Black, Asian/Pacific Islander, and American Indian/Alaska Native populations each accounted for a disproportionately lower share of these outcomes relative to their population size.

Figure 3.1: Proportion of population and pharmaceutical utilization and spending for each race and ethnicity group, 2019



Notes:

- These unadjusted proportions are not age- or prevalence-standardized
- Spending and prescription fills are estimated from our study across all payers and US states. Population data is from the US Census Bureau. Insurance coverage data is from the State Health Access Data Assistance Center (SHADAC).
- All groups other than Hispanic exclude individuals of Hispanic ethnicity.

2. Overall and payer-specific trends (age-standardized)

Overall, the White population had higher per capita retail pharmaceutical utilization and spending than other groups, even after adjusting for age differences (**Figure 3.2**). Private insurance accounted for 43.1% of retail pharmaceutical spending, followed by Medicare (32.3%), out-of-pocket spending (14.4%), and Medicaid (10.1%). For Medicare, utilization and spending were generally consistent across groups, with the exception of Black beneficiaries, who had marginally lower utilization. In contrast, for both private insurance and Medicaid, White beneficiaries had significantly higher prescription fills per person, while Asian/Pacific Islander, Black, and Hispanic beneficiaries had significantly lower fills and spending per person.

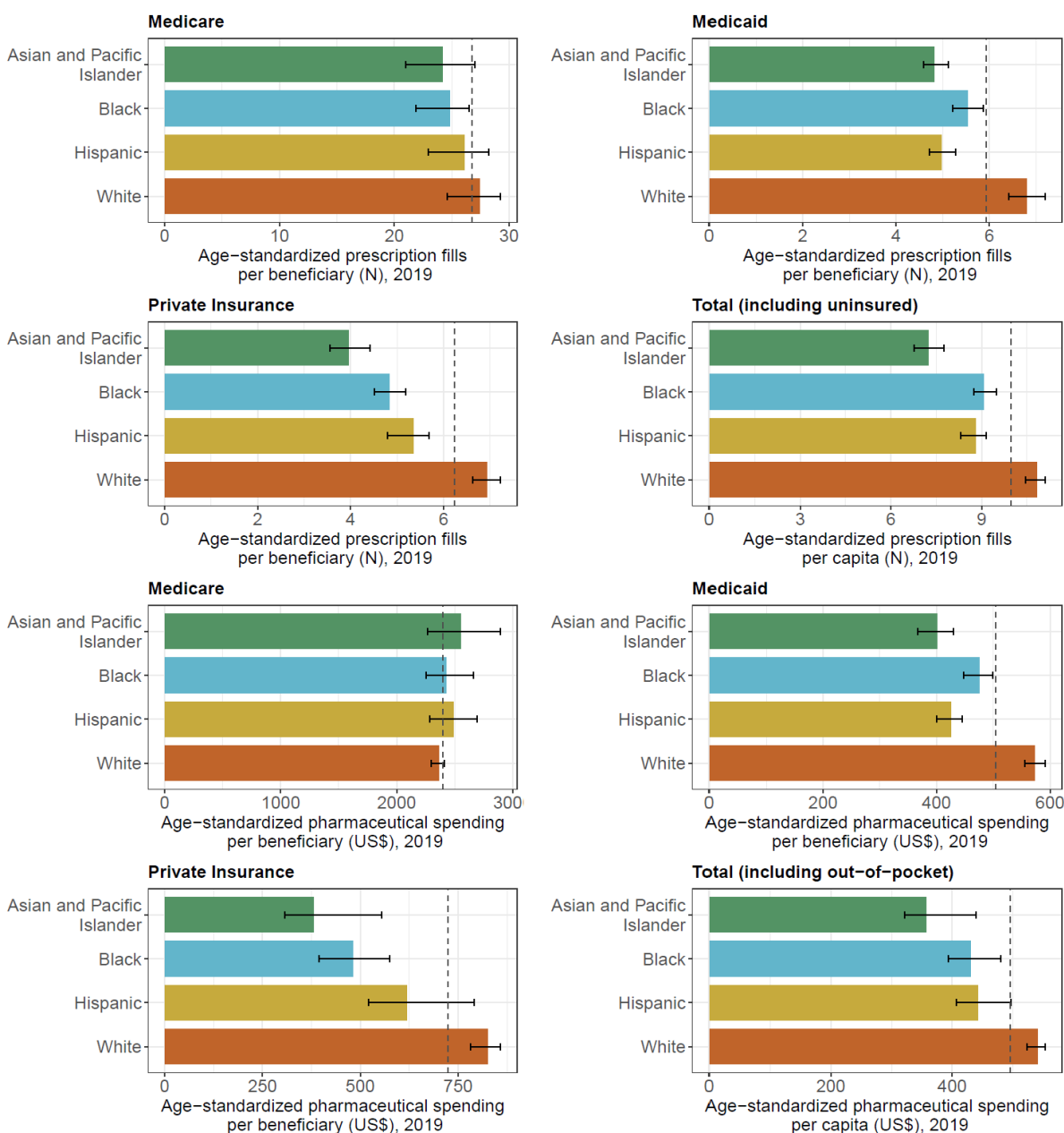
3. Disease-specific trends (age- and prevalence-standardized)

Figure 3.3 and **Appendix Figure B3.1** present age-standardized pharmaceutical utilization and spending per capita and per prevalent case for 52 health conditions with available race- and ethnicity-specific disease prevalence estimates. Compared with per capita estimates (**Figure 3 Panel A**), standardizing for prevalence revealed dramatically different trends (**Figure 3 Panel B**), with Black individuals having substantially lower than average prevalence-standardized utilization and Asian/Pacific Islander individuals having substantially higher than average prevalence-standardized utilization. For example, for type 2 diabetes (the condition with the highest total pharmaceutical spending and utilization), Black populations had markedly fewer prescriptions per prevalent case, while Asian/Pacific Islanders had substantially greater than average prescriptions per prevalent case. Full rankings of pharmaceutical utilization and spending – both per capita (for all 143 health conditions) and per prevalent case (for the 52 conditions with available prevalence data) – are provided in Appendix S3.

4. Geographic variation (age- and prevalence-standardized)

In 2019, the states/districts with the highest age-standardized per capita utilization of prescription drugs were District of Columbia, Alabama, West Virginia, Rhode Island, and Kentucky (**Figure 3.4 Panel A and Appendix Table 4.1**). The states with the lowest age-standardized per capita utilization of prescription drugs were (starting with lowest) Washington, Colorado, Wyoming, Alaska, and California. Spatial patterns across the 50 US states and District of Columbia differed by race and ethnicity (**Figure 3.4 Panel B**). Differences across race and ethnicity groups in age- and sex-standardized prescription drug utilization varied across US states. Pharmaceutical utilization exceeded the all-population mean in 16 states (31%) for Asian/Pacific Islander populations, 23 states (45%) for Black populations, 29 states (57%) for Hispanic populations, and 46 states (90%) for White populations. Geographic patterns varied similarly for pharmaceutical spending (**Appendix Figure B4.1**).

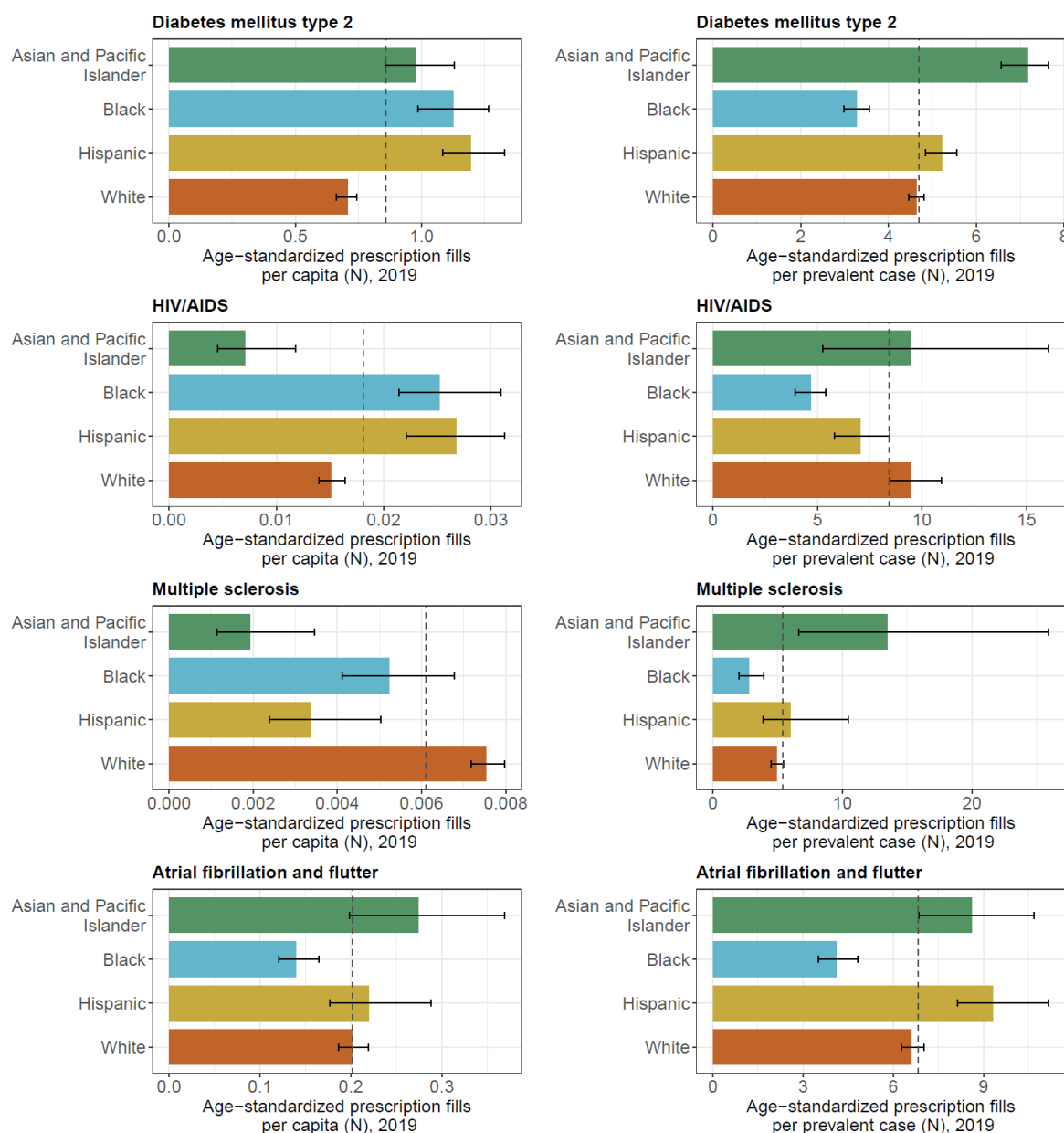
Figure 3.2: Age-standardized pharmaceutical utilization and spending by payer for each race and ethnicity group, 2019



Notes:

- This figure shows the age-standardized utilization and spending per capita (for the total) and per beneficiary (for Medicare, Medicaid, and private insurance) for each race and ethnicity group.
- The dashed line represents the all-population age-standardized mean prescription fills.
- Error bars represent the 95% Uncertainty Interval.
- All groups other than Hispanic exclude individuals of Hispanic ethnicity.
- The American Indian/Alaska Native population is excluded due to lack of data.

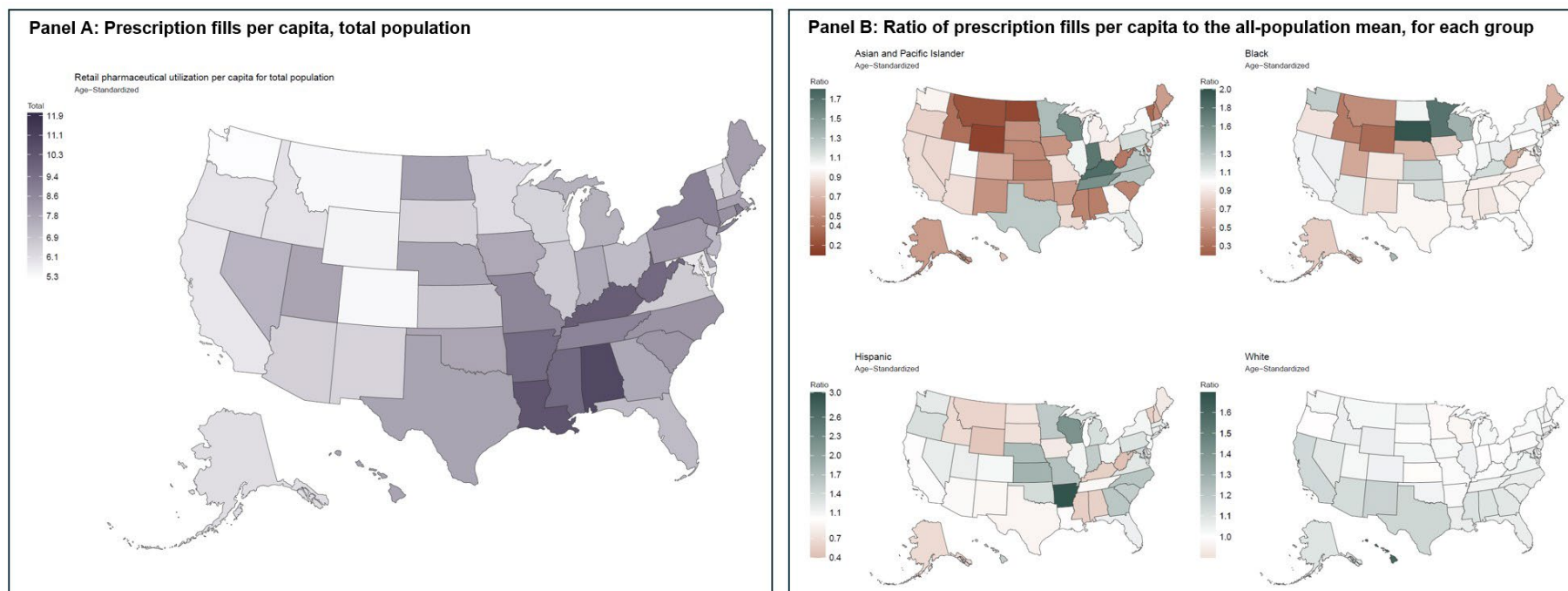
Figure 3.3: Age-standardized pharmaceutical utilization per capita and per prevalent case for four health conditions, 2019



Notes:

- This figure includes only four health conditions with the greatest pharmaceutical spending, out of 52 with available prevalence estimates. See Appendix Figure B3.1 for all 52 health conditions, including both pharmaceutical utilization and spending.
- Spending and utilization estimates are inclusive of preventive spending for that health condition.
- Dashed line is the all-population mean. Error bars show the 95% Uncertainty Interval.
- All groups other than Hispanic exclude individuals of Hispanic ethnicity.
- The American Indian/Alaska Native population is excluded due to lack of data

Figure 4: Age-standardized ratio of pharmaceutical utilization per capita to the all-population mean for each race and ethnicity group and US state, 2019



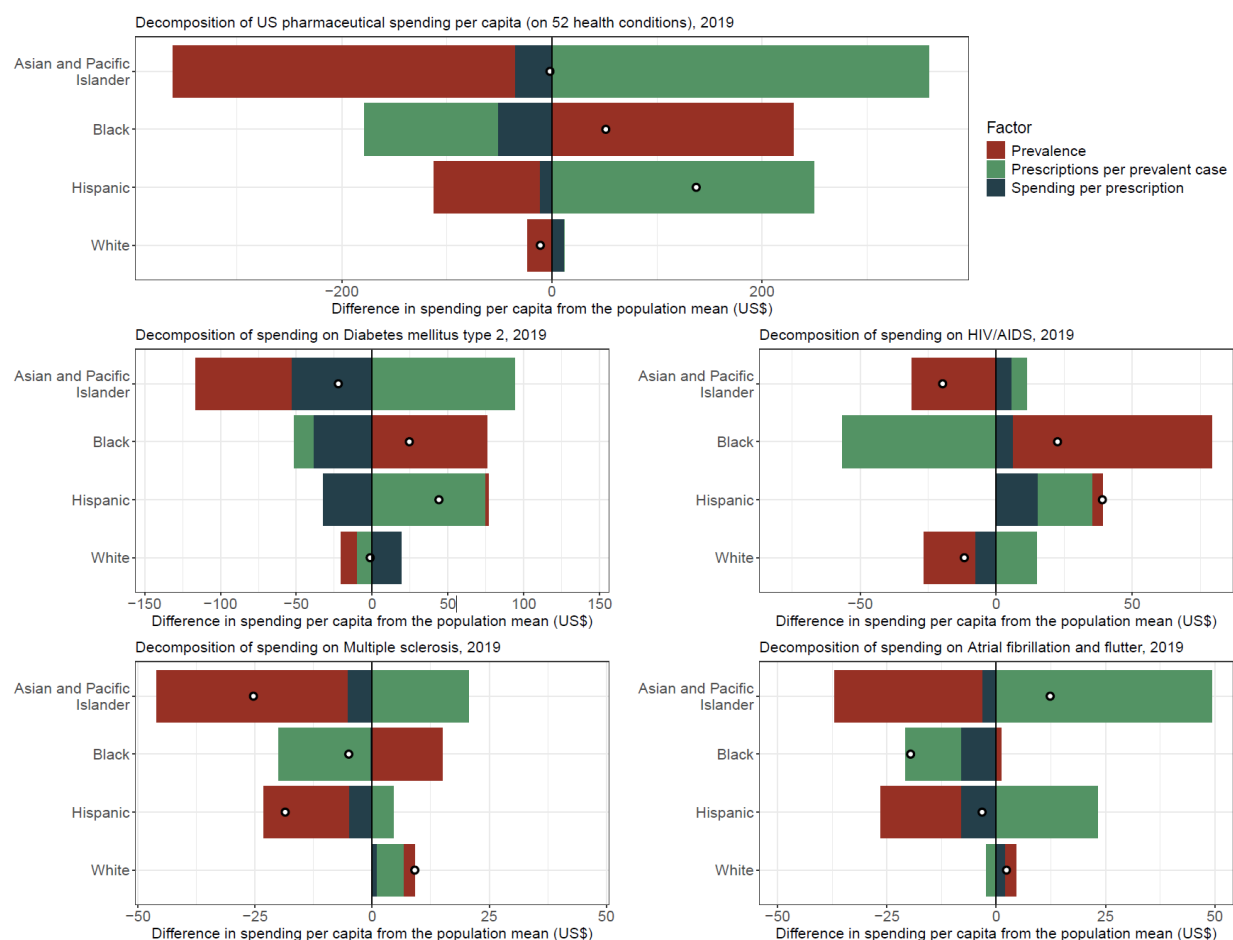
Notes:

- Panel A shows the total age-standardized pharmaceutical utilization per capita for the whole population of each state. Panel B shows the ratio of the race-specific age-standardized pharmaceutical utilization per capita compared to the corresponding value for the all-population mean for each state.
- State-level estimates and uncertainty intervals for prescription fills and spending per capita are shown in Appendix Table B4.1.
- All groups other than Hispanic exclude individuals of Hispanic ethnicity.
- The American Indian/Alaska Native population is suppressed due to lack of data, but is included in the all-population means.

5. Decomposition of pharmaceutical spending

Figure 5 and **Appendix Figure B3.2** show the national-level decomposition of retail pharmaceutical spending for 52 health conditions by race and ethnicity into prevalence, prescriptions per prevalent case, and spending per prescription; these 52 conditions represent 42% of pharmaceutical spending and 20% of prescription fills. Across these 52 conditions, Asian/Pacific Islander populations had lower than average spending per capita while the Black and Hispanic populations had higher than average spending per capita; White populations had spending per capita that was close to the all-population mean. For Asian/Pacific Islander populations, lower spending per capita appeared driven by lower prevalence, while the effect of prescriptions per prevalent case was higher than average. Notably, higher spending for Hispanic populations was driven by utilization per case, while higher spending for Black populations was driven by disease prevalence. Disease-specific results for 52 health conditions are presented in **Appendix Figure B3.2**, and state-level results across all 52 health conditions are shown in **Appendix Figure B4.2**.

Figure 5: Decomposition of national age-standardized pharmaceutical spending per capita, 2019



Notes:

- *This figure presents the results from the Das Gupta decomposition analysis, which shows the extent to which differences in each factor (disease prevalence, utilization, and spending per prescription) contributes to differences between the age-standardized race-specific pharmaceutical spending per capita and total national-level pharmaceutical spending per capita. The dot represents the total effect (in per capita terms), and each bar reflects the difference in age-standardized spending per capita due to each effect.*
- *The total figure (top panel) includes only 52 health conditions with available prevalence data, which account for 42% of pharmaceutical spending and 20% of prescription fills. Decomposition results for the remaining 48 health conditions are presented in Figure B3.2.*
- *All groups other than Hispanic exclude individuals of Hispanic ethnicity. The American Indian/Alaska Native population is excluded due to lack of data.*

Discussion

This study comprehensively describes inequities in retail prescription drug utilization and spending by race and ethnicity in 2019, leveraging a wealth of data to reveal heterogeneity across payers, health conditions and US states. After standardizing for age and disease burden, substantial disparities persisted, particularly affecting the Black population, highlighting ongoing barriers to achieving pharmacoequity.

Interpretation of findings

Our results confirm previous findings of higher per capita pharmaceutical utilization, spending, and out-of-pocket spending among White populations, driven in part but not completely explained by their older age.^{26,101,102} When disaggregated by payer (Medicare, Medicaid, private insurance), substantial inequalities emerged. While Medicare Part D showed little variation across race and ethnicity groups, Medicaid and private insurance revealed greater disparities in both spending and utilization. Our findings align with prior studies documenting higher per-capita pharmaceutical utilization and spending for White individuals as well as lower utilization and spending across payers,²⁶ though we see fewer differences than previously observed in Medicare.^{101,102} Prior studies did not comprehensively account for disease burden. After standardizing for disease prevalence, we found substantially lower prescription fills per prevalent case for Black populations and substantially *higher* than average prescription fills per prevalent case for Asian/ Pacific Islander populations. These disparities in access to medicines may help explain the more favorable health outcomes and higher life expectancy observed among Asian individuals and poorer outcomes for Black individuals.^{87,114,115}

Disease-specific analyses illustrate the importance of contextualizing disparities relative to disease burden. While compared with the all-population average, the Black population had higher per capita prescription drug utilization for most health conditions examined but markedly lower prescriptions per prevalent case, suggesting undertreatment relative to disease burden. In contrast,

the Asian/Pacific Islander population had lower than average per capita utilization for these conditions but higher than average per prevalent case utilization. Moreover, while the Hispanic population had lower than average per capita utilization for some conditions, this effect disappeared after accounting for differences in prevalence. Our decomposition of race-specific age-standardized spending compared with the all-population means similarly showed that Asian/Pacific Islander populations spent less per capita on pharmaceuticals due to lower prevalence; Black populations spent more per capita due to high prevalence but utilization was low with respect to prevalence; Hispanic populations spent more per capita than average due to higher utilization; and White populations spent close to average per capita but had higher spending per prescription compared with other groups. Linking these data to disease prevalence reveals patterns and disparities that would otherwise be obscured by analyzing utilization and spending alone.

Differences in access to and uptake of medicines across racial/ethnic groups can be influenced by differential prescribing (e.g., due to provider access, provider type visited,^{31–33} and discrimination^{34,36,37}) and differential uptake and adherence (e.g., due to insurance coverage,^{38,39} out-of-pocket cost burden,^{28,29} pharmacy access,⁴⁶ differential drug pricing across pharmacies,^{43–45} and health system distrust⁸). For example, a growing body of research on ‘pharmacy deserts’ has found poorer access to pharmacies in predominantly Black and Hispanic neighborhoods.^{32–116} Varying disease prevalence and healthcare access across geography may also likely contribute to our observed state-level variation in racial and ethnic differences in per-capita drug utilization. Previous research has highlighted the mixed success of insurance expansions in mitigating these gaps,¹¹⁷ with Medicaid expansion improving access to prescription drugs for some populations^{118–120} but the Medicare Part D coverage gap being particularly disruptive to Black and Hispanic populations.^{102,103}

Policy implications

Our findings highlight the need for tailored interventions to address racial/ethnic disparities in medication access and utilization. At the state level, prior studies have shown that Medicaid expansion demonstrated success in improving access, yet its incomplete adoption across the US limits its potential impact.^{118–120} Federal programs should aim to ensure equitable medication access while accommodating regional differences. While the long-term impact of the Inflation Reduction Act (IRA) on racial disparities is unknown, the IRA’s patient out-of-pocket caps and enhanced subsidies could further reduce financial barriers for vulnerable populations.^{99,121} Additionally, addressing structural inequities, such as pharmacy deserts and differential drug pricing, is critical to achieving pharmacoequity.⁴⁶ Efforts to enhance equity should also prioritize preventive care and early treatment, such as prophylaxis for HIV.¹²² Increasing equitable access to therapies such as GLP-1 receptor agonists and SGLT-2 inhibitors, could improve outcomes for type 2 diabetes and cardiovascular disease.^{123,124} Expanded insurance coverage, lower cost-sharing, and equitable provider incentives could mitigate these medication access disparities.

Limitations

This study had several limitations. First, it excluded data from the Indian Health Service, Tricare, Veterans Affairs, and incarcerated populations. Because 64% of American Indian/Alaska Native individuals access the Indian Health Service,¹⁰⁶ we did not report disaggregated estimates for this group – an important omission given known disparities.^{17,125} Second, race and ethnicity data were incomplete or unavailable in several datasets, especially private insurance claims, which have known limitations.¹²⁶ Nonetheless, our estimates (including those for private insurance) aligned with national survey data that better capture race and ethnicity.¹²⁷ Third, small population sizes for some groups (e.g., Asian and Pacific Islander individuals) contributed to wide uncertainty intervals. Fourth, demographic classifications were limited (e.g., the Asian and Pacific Islander group aggregates heterogeneous populations⁴) and variation in data collection across states (e.g., for Medicaid) may affect comparability. Fifth, drug-indication probabilities were derived from data without race/ethnicity, and we assumed these probabilities did not vary across groups; this may bias disease-specific estimates if prescribing patterns differ. Sixth, we only included retail prescriptions, excluding inpatient and in-office administered drugs, which account for roughly 30% of spending.⁹⁵ Seventh, our data ended in 2019 and did not capture the impact of the COVID-19 pandemic. Finally, because our analysis was conducted at the disease (rather than drug) level, we could not distinguish between treatment versus preventive use, account for differences in prescription length (e.g., 30- versus 90-day fills), or decompose the effects of drug prices versus product mix. We also did not examine the extent to which variation reflected differences in adherence,⁴⁶ provider behavior, or disease severity.

Conclusions

This study highlights persistent disparities in retail prescription drug utilization and spending by race and ethnicity across the US. In particular, we found that Black populations consistently had lower prescription drug utilization relative to disease prevalence. Geographic and payer-specific variations further highlight the need for targeted policy solutions. Advancing pharmacoequity will require coordinated action across federal, state, and local levels and the private sector to ensure equitable access to essential medications.

Acknowledgments

This work was completed with support from the National Pharmaceutical Council. We thank the members of the Institute for Health Metrics and Evaluation's US Health Disparities Team (in particular Michael Celone, Yekaterina Kelly, and Chris Schmidt) for their valuable insights and work on estimating disease burden by race and ethnicity.

Chapter 4 Pharmacy Access, Utilization of Medicines, and Racial Disparities

This chapter is in preparation for submission to a health policy journal as:

Sahu M, Wagner TD, DeJarnatt D, Beauchamp M, Crosby S, Lescinsky H, Taylor K, Thomson A, Weil M, Wittenauer R, Dwyer-Lindgren L, Haakenstad A, Stergachis A, Campbell JD, Dieleman JL. *Pharmacy Access, Utilization of Medicines, and Racial Disparities in the United States, 2010-2023*.

Abstract

Racial and ethnic disparities in access to medicines persist across the United States. Community pharmacies serve as critical access points for prescription drugs, yet their availability varies widely. This uneven distribution may contribute to disparities in medication use. Using census block group-level population data and geolocated pharmacies from 2010 to 2023, we estimated the share of each race and ethnicity group living within ten- and twenty-minutes of a pharmacy, by driving or walking, at the county and state levels. County-level pharmacy access varied substantially by race and ethnicity. In addition, utilization of medicines (after standardizing for age and disease prevalence) varied substantially by race and ethnicity. States with poorer pharmacy access showed wider racial and ethnic disparities in prescription fills across conditions. These findings suggest that uneven pharmacy access may exacerbate racial disparities in medication use. Policymakers seeking to improve access to medicines could consider targeted interventions to preserve or expand pharmacy infrastructure for underserved communities.

Introduction

Racial disparities in access to medicines persist across the United States (US), with Black individuals filling fewer prescriptions relative to their disease burden compared to other groups.^{128,129} In addition to dispensing medications, community pharmacies serve as critical health care access points for vaccinations and other basic health services.^{46,130–132} However, access to community pharmacies varies substantially across the US: county-level pharmacies per capita range from 0 to 13.6,¹³³ and 138 counties (4.4%) lack even a single pharmacy.^{134,135} Rural areas face particularly poor access, often compounded by limited public transportation and other mobility barriers.^{57,135,136} Within urban settings, studies on *pharmacy deserts* suggest that access in predominantly Black and Latinx neighborhoods is poorer compared with White and diverse neighborhoods.^{40–43} Access challenges may be further compounded by *pharmacy closures*, which have been on the rise, leading to an overall decline in pharmacies from 2018 to 2021⁶⁸ that may disproportionately affect communities already experiencing barriers, e.g., rural areas¹³⁶ or low-income urban areas.^{67,68} While geographic pharmacy access patterns are increasingly well described, direct evidence linking pharmacy access to medication use and its associated disparities remains limited.

Several studies have examined the relationship between access or distance to pharmacies and medication access, utilization, or adherence, with mixed findings. In rural areas, greater travel time has been linked to a lower likelihood of visiting a pharmacy; however, this trend does not hold in urban areas, where patients may bypass nearby independent pharmacies in favor of more distant chain stores.^{137,138} For example, a study of Medicaid patients with diabetes in Chicago found no association between distance to pharmacies and medication adherence.¹³⁹ On the other hand, studies on pharmacy deserts—typically defined as areas that are both low-income and low-access—have linked poor access to lower uptake of influenza vaccination and increased COVID-19 risk.^{140,141} Among elderly patients in Pennsylvania, medication adherence was modestly lower in pharmacy deserts compared to non-desert areas.¹⁴² Recent studies of pharmacy closures have found that closures are associated with immediate declines in prescription fills for cardiovascular medications⁷⁷ and anticonvulsants,⁷⁸ though long-term effects remain unclear. While a growing body of literature examines pharmacy deserts and closures, often in the context of racial and ethnic disparities,^{40,41,68} the direct relationship between pharmacy access and pharmaceutical utilization, including racial and ethnic disparities in utilization, has not been clearly established.

The US Disease Expenditure (DEX) project provides US state- and county-level estimates of pharmaceutical utilization across 143 health conditions for 2010 to 2019,⁸² with disease prevalence estimates available for 78 of those conditions.⁸³ A follow-on study estimated utilization by race and ethnicity across 50 US states and the District of Columbia, and for 52 out of 143 conditions with available race-specific disease burden estimates, estimated prescription utilization per incident or prevalent case—providing a novel dataset for examining pharmaceutical disparities.¹²⁸ In this study, we combine the DEX estimates with a historical dataset of geolocated US pharmacies

(2010-2023)¹⁴³ to examine how pharmacy access relates to utilization by race and ethnicity. Specifically, we (1) assess county- and race-specific trends in pharmacy access from 2010 to 2023, (2) construct a measure of inequality in pharmaceutical access by race and ethnicity, and (3) examine how pharmacy access contributed to racial and ethnicity disparities in medication utilization per prevalent case. We also explore how these associations vary by health condition. To our knowledge, this is the first study to directly link disparities in pharmacy access to disparities in medication use.

Study Data and Methods

Data

This analysis integrated multiple data sources to assess the relationship between pharmacy access and pharmaceutical utilization, including by race and ethnicity, across US counties and states. Our analysis was restricted to retail prescription drugs dispensed at pharmacies—excluding medications administered in hospitals, outpatient clinics, or other institutional settings—to capture the type of drug use most plausibly linked to geographic pharmacy access. The study covered five mutually exclusive race and ethnicity groups—the Hispanic, and (all non-Hispanic) American Indian/Alaska Native, Asian/Pacific Islander, Black, and White populations.

Pharmacy Access

Geolocated pharmacies were obtained from the National Council for Prescription Drug Programs (NCPDP), covering all active pharmacies on April 1 of each year, for fourteen years from 2010 to 2023.¹⁴³ We obtained census block group (CBG)-level population by race and ethnicity^{144,145} and census tract-level vehicle ownership rates, representing the percent of households with an available vehicle,¹⁴⁶ from the US Census Bureau’s 5-year American Community Survey (ACS) files. We then constructed a county- and state-level metric of pharmacy access, defined as the ‘percent of the population living within 20 minutes of a pharmacy,’ using the following steps: (1) we identified the geographic boundaries corresponding to 20-minute walking and driving distances around each pharmacy, (2) we obtained the geographic centroid of each CBG, (3) we used each centroid’s location to classify the entire CBG population as either within a 20-minute walk or outside a 20-minute drive of a pharmacy, (4) for populations in between a 20-minute walk and 20-minute drive, we applied household vehicle ownership rates to estimate access, and (5) we aggregated estimates to the county and state levels. Details on methods and input data are in Appendix C.

Pharmaceutical Utilization

This study used the DEX state-level estimates of retail prescription drug utilization by race and ethnicity for 143 health conditions from 2010 to 2019, derived using over 27.3 billion public and commercial pharmaceutical claims and survey records covering Medicare, Medicaid, private insurance, and the uninsured.¹²⁸ For 52 of these conditions with available race-specific prevalence

estimates at the state level, pharmaceutical utilization was linked to estimates of state-level disease incidence (for cancers) and prevalence (for all other causes)^{84,128} to estimate prescriptions per prevalent or incident case (hereafter referred to as ‘prescriptions per prevalent case’). To improve interpretability, low-burden health conditions were aggregated, resulting in a final analytic set of 28 grouped health conditions. For comparability across racial and ethnic groups, we age- and sex-standardized prescriptions per prevalent case using the total US population as the reference. Uncertainty in these estimates was captured by simulating 50 draws from a multivariate normal distribution (as described in Chapter 3), which we carried through all analysis. Full details of the estimation process are published elsewhere.^{82,128}

Pharmaceutical Disparities

To examine pharmaceutical disparities, we estimated state-level disparities in age-standardized prescriptions per prevalent case by calculating the Gini coefficient¹⁴⁷ for each state and condition, treating the population as if it were evenly distributed across racial and ethnic groups. The Gini coefficient is the most popular measure of income inequality, while in this case it was used to measure equitable utilization of pharmaceuticals per prevalent case. In addition to the cause-specific version, we also computed the weighted mean of these Gini coefficients across health conditions within each state, using total disability-adjusted life years (DALYs) as weights. We carried through the 50 draws from above, generating 50 draws of these disparity indices at the state-condition-year level to reflect uncertainty. To visualize the distribution of pharmaceutical utilization across racial and ethnic groups, we also plotted Lorenz curves.

Statistical Analysis

We conducted two separate regression analyses at the state level: (1) assessing the association between pharmacy access and pharmaceutical utilization per prevalent case, disaggregated by race and ethnicity, and (2) examining the association between state-level pharmacy access and racial disparities in pharmaceutical utilization. We used panel regressions with year fixed effects to account for secular time trends. To address potential confounding, we controlled for income per capita, because states with higher income have both greater pharmacy access and higher utilization due to broader structural advantages and higher disposable income. We combined draw-level estimates from our panel regressions using Rubin’s Rules.¹⁴⁸

Statistical analyses were conducted with R version 4.5.0, using the *sf* package for spatial data processing and the *plm* package for panel regressions.^{149,150} We used Mapbox and qGIS to convert street addresses to coordinates, and driving and walking isochrones were generated using Mapbox. This study did not involve individual patient records, and was exempt from ethics review.

Sensitivity Analysis

We also assessed pharmacy access using a specification of 10-minutes rather than 20-minutes.

Limitations

Our study has several limitations. First, our utilization estimates, including those stratified by race and ethnicity, only extend through 2019, limiting our assessment of the impact of pharmacy access in more recent years. Second, we relied on state-level race-specific prevalence estimates, limiting our ability to run county-level analyses. Third, while our measure of pharmacy access incorporated both walking and driving distances—an improvement over simple Euclidean distance¹³⁵—we did not incorporate alternative conceptualizations used in the literature, including two-step floating catchment areas (which also capture demand and crowdedness),^{151,152} pharmacy deserts,⁴⁰ or commuting-based measures that reflect proximity to workplaces rather than residences. Moreover, we only captured geographical access, not payer-level differences. Fourth, while our estimates were relatively granular—using census block groups for population and census tracts for vehicle ownership (the most detailed level available for that variable)—we did not analyze data at the census block level, potentially missing more localized neighborhood-level patterns of racial and ethnic inequities in pharmacy access. We also did not examine telepharmacy, which in theory can improve access without additional brick-and-mortar infrastructure,^{79–81} nor did we assess pharmacy quality, such as stockouts, which may disproportionately affect lower-income communities.⁶⁹ Finally, we did not evaluate the downstream impact of pharmacy access on health outcomes or associated disparities.

Results

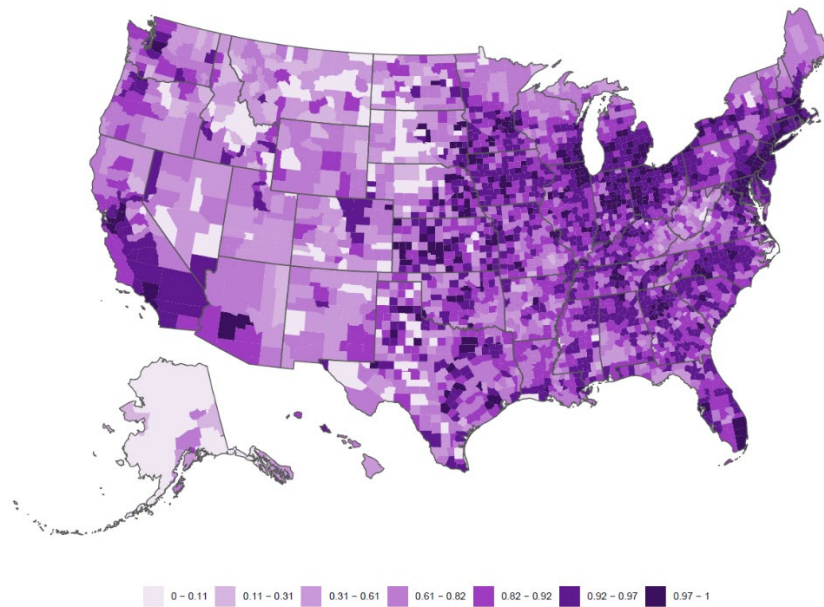
Pharmacy access varied substantially across US counties in 2019, with wide disparities by race and ethnicity (**Exhibit 4.1**). Nationally, 93.6% of the population lived within 20 minutes of a community pharmacy. In 2,746/3,110 counties (88.3%), greater than half of the total population lived within 20 minutes of a pharmacy.

Additionally, 20-minute access was markedly lower among American Indian/Alaska Native populations (69.7%), compared with Asian/Pacific Islander (97.6%), Black (95.6%), Hispanic (95.7%), and White populations (92.5%). Geographic access was lowest in rural counties, particularly in the rural Mountain West, and parts of the Midwest and South. From 2010 to 2019, change in pharmacy access over time was not consistent across counties, with counties experiencing a range of declines and increases in pharmacy access (**Exhibit 4.1, Panel B**).

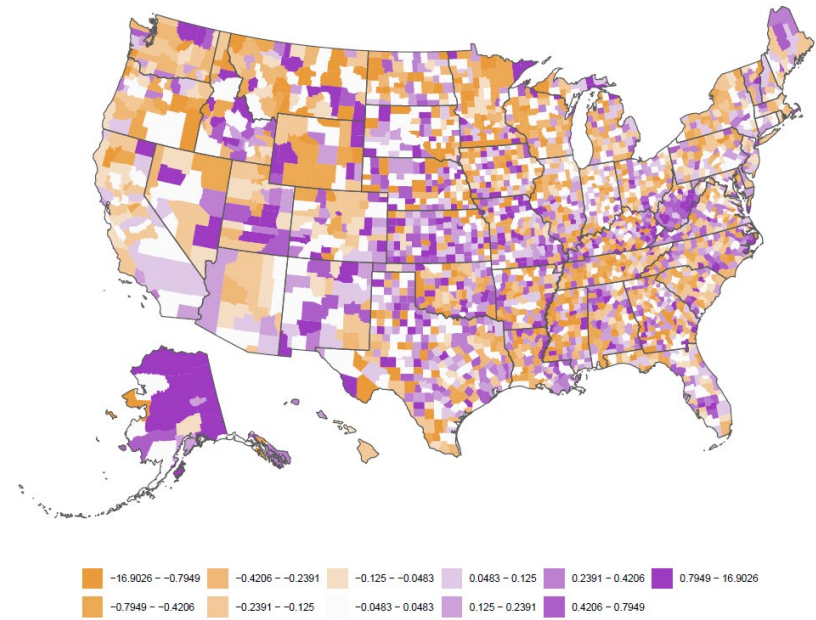
Racial and ethnic disparities in pharmaceutical utilization was substantial in 2019 (**Exhibit 4.3**). Black populations received fewer prescriptions per prevalent case relative to White populations across nearly all states and most conditions. Overall, disparities were highest in the Southeast, Midwest Mountain West, and Texas, though this varied by health condition and state. Lorenz curves illustrated the unequal distribution of utilization across racial and ethnic groups, with the steeper curves observed in for multiple sclerosis compared with diabetes, HIV/AIDS, and atrial fibrillation and flutter.

Exhibit 4.1: County-level variation in 20-minute pharmacy access (% of the population living within 20 minutes), 2010-2019

Panel A: 20-minute pharmacy access, 2019



Panel B: Change in 20-minute pharmacy access, 2010-19



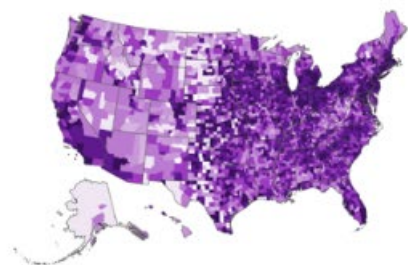
Notes:

- Percent of the population within 20 minutes of a pharmacy is defined as those living in Census Block Groups whose geographic centroids fall within a 20-minute walk, outside of a 20-minute drive, or (if in between) their vehicle ownership rate.
- Percent change over time is not applicable (NA) when the percent living within 20 minutes of a pharmacy in 2010 was 0.

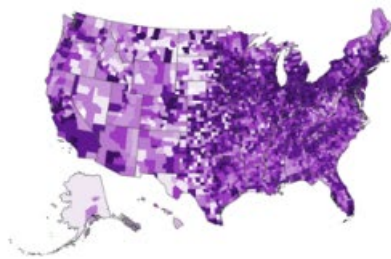
Exhibit 4.2: County-level variation in 20-minute pharmacy access by race and ethnicity, 2019

A: Pharmacy access for each race/ethnicity group

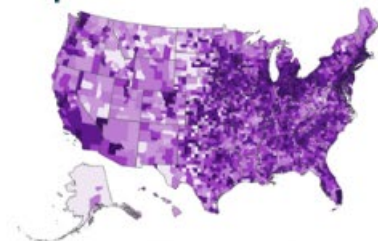
Asian/Pacific Islander



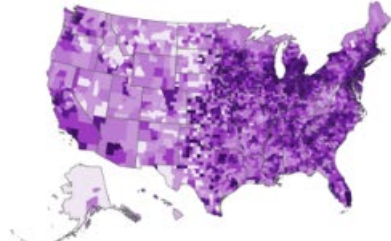
Black



Hispanic

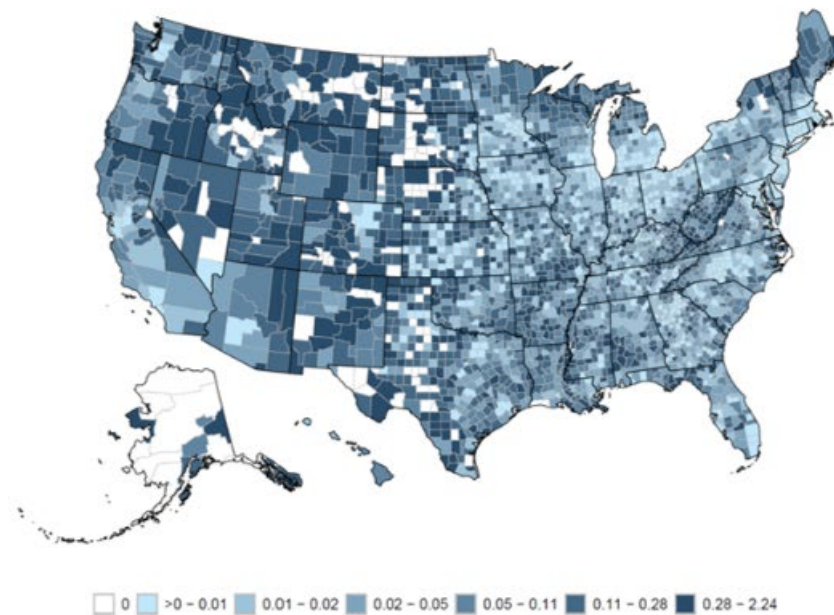


White



0 - 0.11 0.11 - 0.31 0.31 - 0.61 0.61 - 0.82 0.82 - 0.92 0.92 - 0.97 0.97 - 1

B: Coefficient of variation across race/ethnicity

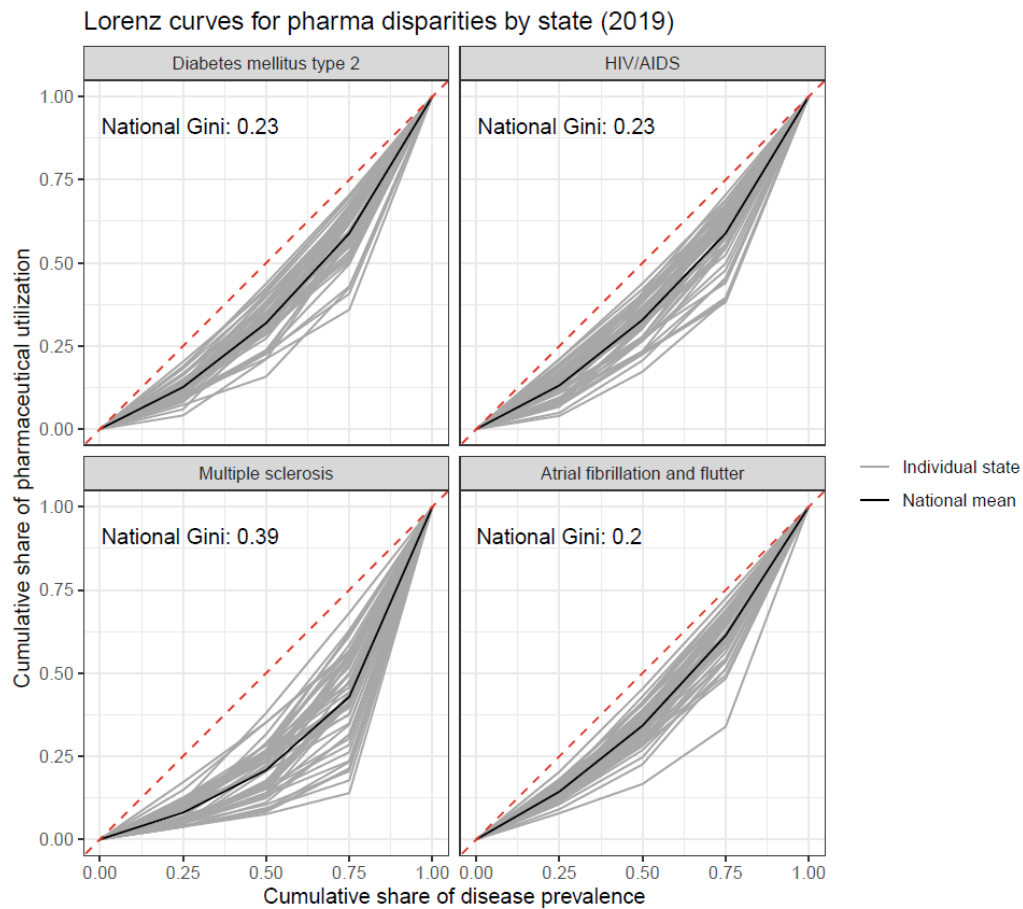


0 >0 - 0.01 0.01 - 0.02 0.02 - 0.05 0.05 - 0.11 0.11 - 0.28 0.28 - 2.24

Notes:

- *Percent of the population within 20 minutes of a pharmacy is defined as those living in Census Block Groups whose geographic centroids fall within a 20-minute walk, outside of a 20-minute drive, or (if in between) their vehicle ownership rate.*

Exhibit 4.3: Racial and ethnic disparities in retail pharmaceutical utilization across the United States, 2019



Disparity index by state (2019, RX only)

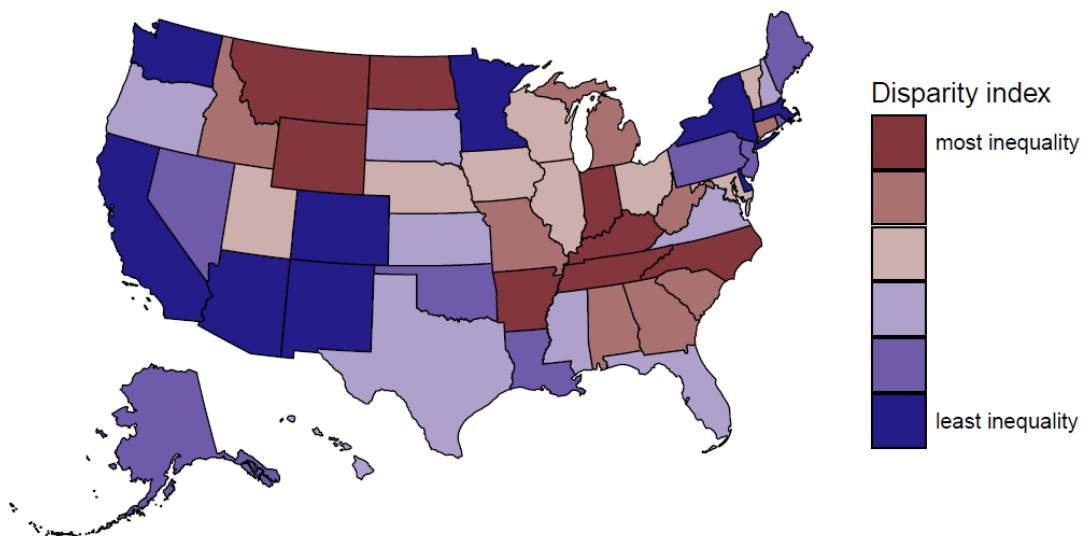
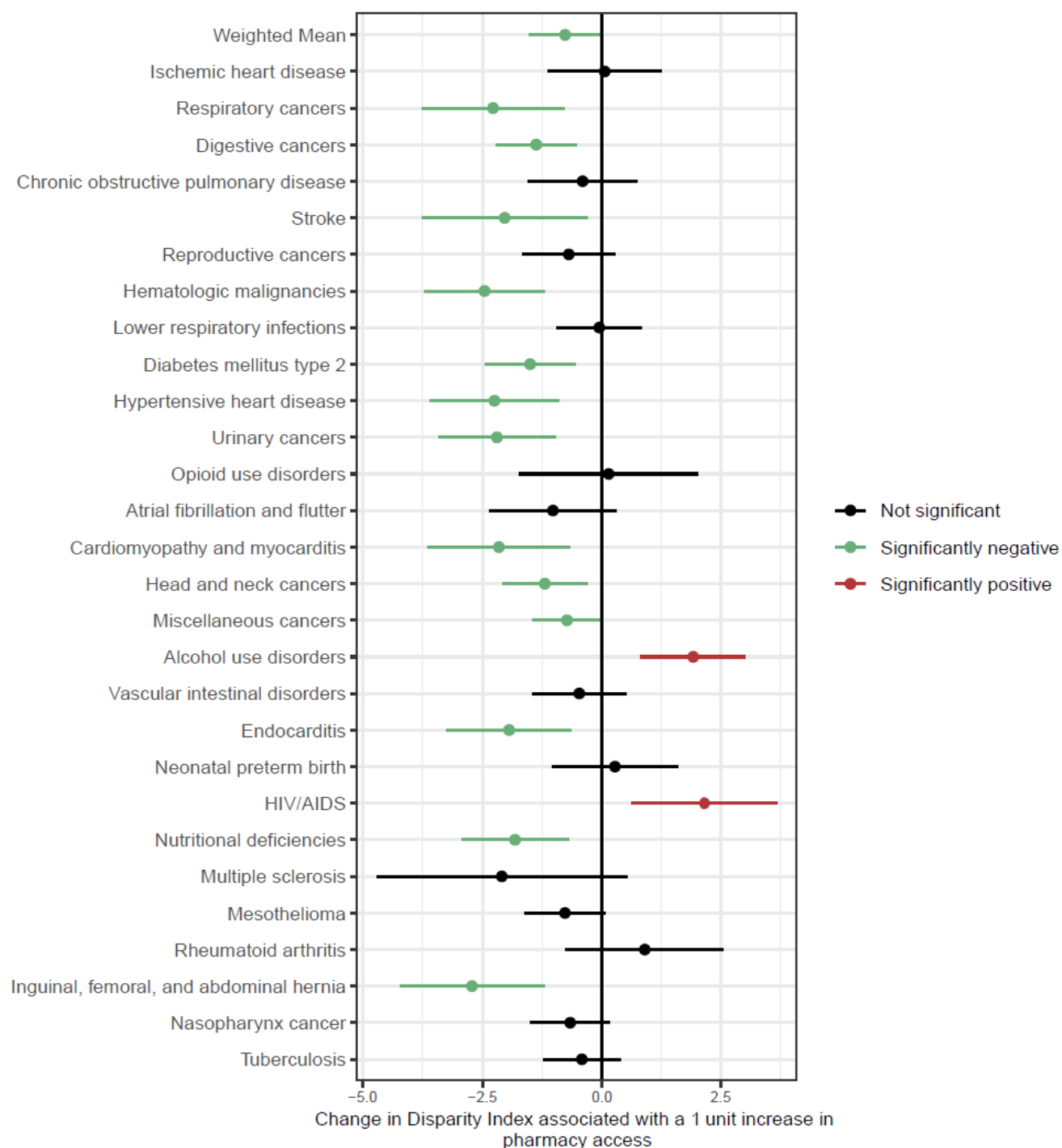


Exhibit 4.4: Associations between state-level access to community pharmacies (within 20 minutes) and racial and ethnic disparities in retail pharmaceutical utilization by health condition, 2010-2019



Notes: Panel regression includes a fixed effect for time and controls for income per capita.

Across states, the relationship between pharmacy access and pharmaceutical utilization varied significantly by race and ethnicity. In particular, Asian/Pacific Islander, Black, and Hispanic populations had the greatest associations between pharmacy access and pharmaceutical utilization, while this relationship was less pronounced for the White population. We also found that poorer pharmacy access was associated with wider racial and ethnic disparities in prescription drug utilization (**Exhibit 4.4**). In states with a lower share of the population living within 20 minutes of a pharmacy, disparity indices were significantly higher, suggesting a greater concentration of utilization among more advantaged groups. These associations were generally consistent across all health conditions studied, with the exception of HIV/AIDS and alcohol use disorders.

Discussion

Our study finds that pharmacy access varied widely across US counties and racial and ethnic groups from 2010 to 2023. The relationship between pharmacy access and pharmaceutical utilization varied greatly by race and ethnicity. In addition, states with poorer access to community pharmacies had wider racial disparities prescription drug use per prevalent case across a wide range of health conditions, after standardizing for age and disease burden. Using a novel metric incorporating both walk- and drive-time access as well as vehicle ownership, our study provides one of the most detailed analyses to date of the relationship between pharmacy access and medication use across the US.

Unlike prior studies that relied on simple pharmacy-to-population ratios or straight-line (Euclidean) distance,¹⁵³ our access measure accounts for travel time, transportation mode, and geography. This allows for more precise measurement of access-related barriers, particularly in communities with low vehicle access—an important contribution given that incorporating only drive-time has been shown to overestimate pharmacy access.¹⁵⁴ Thus far, studies evaluating the impact of pharmacy access on pharmaceutical uptake have been limited to those evaluating the immediate impact of a pharmacy closure on prescription fills.^{77,78} Our study leverages comprehensive estimates of pharmaceutical utilization from 2010-2019, standardized for age and disease burden, to examine the role of overall pharmacy access, and not just closures, across a range of geographies. Our significant associations suggest that improving pharmacy access could meaningfully increase medication use across a range of health conditions.

Several structural mechanisms may explain the associations we observed. Across urban and rural areas, neighborhoods with poor pharmacy access, sometimes referred to as ‘pharmacy deserts,’ are also more likely to have greater social vulnerability,^{134,155} poorer health insurance coverage,¹¹⁶ and poorer access to primary care physicians.¹⁵⁵ Historic and ongoing structural racism also contributes: for instance, residence in a historically redlined census tract has been associated with 9% lower odds of having a pharmacy within one mile.⁶⁵ Moreover, limited access may disproportionately impact the elderly, who not only have the highest disease burden but are less likely to drive and therefore more dependent on local pharmacy availability.¹⁵⁶ While some might expect mail-order pharmacies to compensate for these barriers, only 15–17% of prescriptions are

filled by mail¹⁵⁷ and some patients still prefer in-person pharmacy services compared with mail-order services.^{158,159} These services are less frequently used by communities of color, lower-income populations, and publicly insured patients, and do not substitute for critical in-person services like vaccinations.^{157,160}

Implications and Future Directions

Our findings suggest that pharmacy infrastructure is an essential component of equitable health care delivery. Unequal access to pharmacies across counties directly relates to geographic differences in pharmaceutical utilization, as well as racial and ethnic disparities in medication utilization. Future research should explore the downstream health consequences of these disparities, including their impact on preventable inpatient and emergency department visits, disease progression, and mortality. It will also be important to assess how recent trends, including the growth of telepharmacy and ongoing retail pharmacy closures, may be altering these dynamics.

Conclusions

Our study provides new evidence that pharmacy access plays a key role both in overall medication utilization and in equity of medication utilization. Improving geographic access to community pharmacies, particularly in underserved rural areas and communities of color, is critical for reducing persistent gaps in access to medicines. Policymakers should consider targeted investments and regulatory interventions to preserve or expand pharmacy infrastructure, including financial incentives to maintain pharmacy services in high-need areas. Efforts to improve medication access must account not only for insurance coverage and affordability but also for the physical accessibility of the sites where medications are dispensed and health services are delivered.

Acknowledgements

This work was completed with support from the National Pharmaceutical Council. We appreciate the valuable insights of Joey Mattingly, Kelly Anderson, and Jennifer Bacci regarding the National Council for Prescription Drug Programs (NCPDP) data used for pharmacy access measurement

Chapter 5 Conclusion

This dissertation advances the science and policy of pharmacoequity by mapping disparities in pharmaceutical utilization and spending—both through novel methodological tools (e.g. a probabilistic map linking drugs to diagnoses), and by presenting geographical patterns of inequity in the US prescription drug landscape. Across three aims, it offers a framework to identify, interpret, and address racial and ethnic disparities in medication access and use.

In **Aim 1**, I developed a probabilistic drug-to-diagnosis mapping approach, effectively a map of medication prescribing intent. This machine learning model estimates the probability that each prescription fill corresponds to a given diagnosis, enabling linkage of over 800 commonly used drugs to 143 condition categories encompassing all diagnoses. Using random forest classification with equal-weighted estimation, the model produces clinically plausible and analytically scalable results. This mapping enables prescription-level attribution to conditions in large-scale claims datasets, forming a foundation for disease-standardized analysis of medication use. This tool also supports future work linking drug spending to disease burden and tracking off-label use across branded and generic products. By enabling condition-specific analysis, this map allows for measuring pharmaceutical inequity not only in volume or cost, but in alignment with clinical need.

In **Aim 2**, I applied this methodological advance to national datasets to quantify racial and ethnic disparities in medication use and spending across payers, conditions, and states. After adjusting for age and condition-specific disease burden, I found persistent disparities—particularly affecting Black populations, who had lower prescription fills per prevalent case across a wide range of health conditions, indicating undertreatment relative to need. Decomposition analysis supported this interpretation: across 52 conditions, higher per capita spending among Black populations was largely driven by elevated disease prevalence, not higher utilization. Disparities varied markedly by payer: they were least pronounced in Medicare Part D and more substantial in Medicaid and private insurance, highlighting the need for payer-specific reforms—including Medicaid expansion, and structural interventions to address undertreatment.

Aim 3 shifted from measuring disparities to examining place-based structural drivers. I developed a novel measure of pharmacy access at the census block group-level, incorporating walk and drive times as well as vehicle ownership rates. I then linked this measure to disease-adjusted utilization across racial and ethnic groups (estimated in Aim 2). The results were clear: over a ten-year period, states with poorer pharmacy access exhibited wider racial disparities in medication use. These patterns reflected inequities in both pharmacy infrastructure^{40,41} and transportation access,¹⁶¹ both of which disproportionately affect Black and Hispanic communities. Achieving pharmacoequity

will require not only reducing prices and expanding coverage, but also addressing physical and logistical barriers to accessing medications.

Together, these three aims provide an integrated view of pharmaceutical inequity—mapping not only who receives medications, but why those patterns persist. By linking patient-level data to systemic factors, this work offers actionable insights for both policymakers and researchers.

Policy Implications

This dissertation highlights three priority areas for reducing racial and ethnic disparities in prescription drug access and advancing pharmacoequity:

1. **Medicaid and private insurance reform.** Disparities in medication use were most pronounced in private insurance and Medicaid—programs that serve many of the most marginalized populations. Recent policies, most notably the “One Big Beautiful Bill”, have introduced Medicaid work requirements, more frequent eligibility checks, and coverage caps, putting millions at risk of losing care.¹⁶² Rather than cutting Medicaid, policymakers should prioritize expanding Medicaid access and reforming benefit design in both Medicaid and employer-sponsored plans to ensure that coverage translates into equitable access and use.
2. **Align medication policy with disease burden.** Across 52 health conditions, prescription use often failed to reflect underlying disease prevalence—especially for Black populations. To address this, medication policy should explicitly account for condition-specific needs. This includes integrating disease burden and severity metrics into pharmaceutical benefit design,^{163,164} using value assessments to guide drug reimbursements,^{165,166} aligning incentives for research and development with unmet medical need,^{167,168} and access initiatives to ensure that high-value therapies reach the populations most affected.¹⁶⁹
3. **Improve pharmacy access.** Geographic access to pharmacies remains a major barrier, particularly in rural areas and communities of color. While mail-order pharmacy can help, it is not yet a full substitute for in-person services—especially for older adults, those without reliable internet, and patients needing vaccinations or consultations.^{157,160} Public investment and regulatory incentives should support pharmacy infrastructure in high-need areas.

Future Directions

In addition, this work opens several avenues for future research:

1. **Scaling drug-to-diagnosis mapping.** The probabilistic mapping approach developed in Aim 1 can be expanded to cover all 9,000+ prescription drugs in large claims databases. Future iterations could leverage AI-assisted mapping of free-text indications (e.g., in Micromedex).
2. **Tracking off-label and branded/generic use.** Systematic tracking of off-label prescribing—especially among generic vs. branded drugs—could improve transparency, safety surveillance, and regulatory oversight, while also revealing patterns of potential market distortion.

3. **Integrating therapeutic value into studies of pharmacoequity.** Future analyses should incorporate measures of clinical benefit or therapeutic value, enabling equity and spending evaluations to focus not just on who receives medications, but who receives high-value medications. This is critical for guiding resource allocation, identifying low-value or wasteful spending, and ensuring that investments in medications advance health outcomes.
4. **Evaluating the impact of pharmacy access.** Building on Aim 3, future work should evaluate the causal impact of pharmacy access interventions (e.g., new pharmacy openings, telepharmacy rollouts, or incentive programs) on prescription use and health outcomes—especially in underserved areas including rural and low-income urban areas.
5. **Examining other structural drivers of racial disparities.** While clinician bias^{34–37} and pharmacy access have gained attention, other structural drivers also merit further investigation. These include formulary design; pharmacy benefit manager (PBM) practices; drug pricing variation; and broader social determinants such as access to food, transport, and housing.^{55–58}
6. **Examining other dimensions of inequities.** Finally, beyond race and ethnicity, research can explore how medication use is shaped across other axes—including rural vs. urban geography, income level, disability, immigration status, and intersecting forms of social disadvantage.

Final Thoughts

This dissertation contributes not only new evidence on racial and ethnic disparities in medication use, but also new methods for measuring, interpreting, and addressing them. By developing a probabilistic approach to link prescriptions to clinical indication, standardizing utilization by disease burden, and integrating structural access metrics, this work lays the foundation for more precise, equitable, and accountable pharmaceutical policy.

If we are to build a health system where everyone—regardless of race, income, or geography—has access to the medications they need, we must ensure that access aligns with burden, not bias; and with need, not privilege. This work is one step toward that goal.

Chapter 6 Additional Contributions

Beyond the papers associated with my three thesis chapters described above, I had the opportunity to lead and contribute to 30 additional projects during my six years of doctoral training, including five as first- or co-first-author. Of these, 23 have been published or accepted. These projects span domestic and global studies of health care spending, equity, value, and access – with a special emphasis on expanding access to high-value medicines. This chapter briefly highlights the additional contributions completed during my time as a doctoral student.

5.1 Geographical variation in health spending and performance across the US

As part of the US County Disease Expenditure (DEX) project led by my advisor Dr. Joseph Dieleman, I contributed to studies assessing geographic variation in health spending and system performance. As a contributing member of this team for several years, I developed the strategy for estimating pharmaceutical spending and supported multiple analyses.

1. *US County Disease Expenditure Team [including **Sahu M**]. “Assessing drivers of variation in health care spending across US counties in 2019.” JAMA Health Forum. Feb 2025. Link*
2. *Johnson E, Wojtesta M, Crosby S, Duber H, Jun E, Lescinsky H, Nguyen P, **Sahu M**, Thomson A, Tsakalos G, Weil M, Haakenstad A, Mokdad A, Murray CJL, Dieleman J. “Varied health spending growth across US states was associated with income, price levels, and Medicaid expansion.” Health Affairs. Aug 2022.*
3. *US County Disease Expenditure Team [including **Sahu M**]. “Evaluating US County Health System Performance and Key Factors Associated with Performance.” (Under review)*

5.2 Racial and ethnic disparities in health spending across the US

As part of my work with the US DEX team, I also contributed to two studies disaggregating national and state-level health care spending by race and ethnicity. The second paper, led by Beauchamp and Thomson, served as the parent study for Chapter 3 of my dissertation.

4. *Dieleman JL, Chen C, Crosby S, Liu A, McCracken D, Pollock I, **Sahu M**, Tsakalos G, Dwyer-Lindgren L, Haakenstad A, Mokdad A, Roth G, Scott KW, Murray CJL. “US Healthcare Spending by Race and Ethnicity, 2002-2016.” JAMA. Aug 2021.*
5. *Beauchamp M, Thomson A, Crosby S, DeJarnatt D, Lescinsky H, McHugh T, **Sahu M**, Taylor K, Weil M, Scott JW, Essien UR, Mokdad A, Dieleman JL. “Tracking US State Health Spending by Race and Ethnicity.” (Under Review)*

5.3 Health care value and market consolidation in the US

In a collaboration between the US DEX team and Stanford's Clinical Excellence Research Center, I led the policy analysis and geospatial methods development for studies estimating state-level health care value and hospital market concentration using American Hospital Association data. This work has resulted in one published article and one manuscript under review on health care value, as well as a third manuscript in preparation on hospital market concentration, which I presented at the 2024 American Society of Health Economists Conference (ASHEcon).

6. Dieleman J, Kaldjian A, **Sahu M**, Chen C, Liu A, Chapin A, Scott KW, Aravkin S, Zheng P, Mokdad A, Murray CJL, Schulman K, Milstein A. "Estimating healthcare delivery system value for each US state and testing key associations." *Health Serv Res*. May 2020.
7. Lescinsky H*, **Sahu M***, Beauchamp M, Crosby S, Johnson EK, McHugh T, Scott JW, Thomson A, Weil M, Schulman K, Dieleman J, Milstein A. "Exploring state-level change in health care value over three decades in the United States, 1991-2020." (Under review)
8. **Sahu M**, Koos H, Lescinsky H, Milstein A, Schulman K, Dieleman JL. "Trends in hospital market concentration across the United States, 2002-2022: A comparison of methods to estimate the Herfindahl–Hirschman index." (In preparation)

5.4 Access to community pharmacies in the US

In collaboration with researchers at the University of Utah and University of Colorado, I contributed to two studies on pharmacy access and closures—now published or accepted in *JAMA* and *JAMA Health Forum*. My work included data processing, methods development, figure creation, and contributing to manuscript writing. I became involved with the research team through Dr. Angela Liu, a former colleague from the US Disease Expenditure (DEX) project.

9. Anderson KE, **Sahu M**, DiStefano MJ, Asche CV, Mattingly II TJ. "Pharmacy Closures and Anti-convulsant Medication Prescription Fills in Colorado." *JAMA*. Nov 2024.
10. Mattingly II TJ, **Sahu M**, Anderson KE. "Community pharmacy churn: putting pharmacy openings and closures in context." *JAMA Health Forum*. July 2025. (Accepted)

5.5 Alignment between pharmaceutical products and disease burden globally

In 2023–2024, I worked part-time with IHME's Integrated Analytics team, led by Dr. Stephen Lim, on two studies that expanded upon methods I developed in my first dissertation aim. One manuscript is currently under review and the other is in preparation. My contributions included strategic planning, data investigation, a coverage analysis using MarketScan pharmaceutical claims, and contributing to manuscript writing and revision.

11. Ozten Y, Bliss E, Silke F, Smith A, Morris A, Everdyke C, Paone K, Sandar L, **Sahu M**, Oleynikova M, Cortopossi N, Angelino O, Lim S. "Measuring pharmaceutical product research and development alignment with disease burden globally." (Under review)

12. Silke F, Ozten Y, Angelino O, Cortopassi N, Morris A, Paone K, Bliss E, Oleynikova M, **Sahu M**, Smith A, Lim S. “Disease burden averted by pharmaceutical products from 20 pharmaceutical companies: A systematic analysis of the Global Burden of Disease Study.” (In preparation)

5.6 Improving treatment delivery for HIV/AIDS medicines in South Africa

Building on my background in community field trials and economic evaluations, I worked with Dr. Ruanne Barnabas (formerly at the University of Washington, now Chief of Infectious Diseases at Massachusetts General Hospital) on studies evaluating the cost, performance, and population-level impact of differentiated delivery models for antiretroviral therapy (ART) in South Africa, Uganda, and Kenya. These efforts aimed to expand equitable access to HIV treatment by testing patient-centered delivery strategies with potential to improve clinical outcomes. I conducted secondary analyses from two field trials, including traditional epidemiologic studies and economic evaluations: (1) the DO ART Study, which evaluated community-based ART delivery and follow-up,¹⁷⁰ and (2) the Deliver Health Study, which evaluated home-based ART delivery and follow-up care.¹⁷¹ This work contributes to the broader evidence base on expanding access to essential medicines in low-resource settings through more patient-centered models of care.

13. **Sahu M**, Szpiro A, van Rooyen H, Asiimwe S, Shahmanesh M, Roberts DA, Krows M, Sausi K, Sithole N, Schaafsma T, Shapiro AE, van Heerden A, Barnabas RV. “Non-communicable disease (NCD) risk among people living with HIV in KwaZulu-Natal, South Africa: evidence from a randomized trial of community-based differentiated service delivery.” *JIAS. Supplement on ‘Differentiated service delivery – beyond HIV treatment for integration and other health needs’*. July 2025. (Accepted)

Role: This was a secondary analysis using DO ART trial data. I conceptualized the study as part of a class project (GH 536), conducted all analyses, generated tables and figures, wrote the paper, and led it through multiple revisions for a JIAS supplement on integrated care.

14. **Sahu M**, Schaafsma T, Szpiro A, van Rooyen H, Asiimwe S, Shahmanesh M, Krows M, Sithole N, van Heerden A, Barnabas RV. “Performance of patient-collected dried blood spot specimens for HIV-1 viral load testing in South Africa.” *AIDS*. Sep 2024.

Role: This was a sub-study embedded in the DO ART trial. I drafted the manuscript, interpreted results, and contributed to the methodology for this diagnostic accuracy study evaluating patient-collected specimens to support home-based HIV viral load monitoring.

15. **Sahu M***, Bayer C*, Roberts DA, van Rooyen H, van Heerden A, Shahmanesh M, Asiimwe S, Sausi K, Sithole N, Ying R, Rao D, Krows M, Shapiro AE, Baeten J, Celum C, Revill P, Barnabas RV. “Population health impact, cost-effectiveness, and affordability of community-

based HIV treatment and monitoring in South Africa: a health economics modelling study.” PLOS Glob Public Health. Sep 2023.

Role: For this follow-on mathematical modeling study of DO ART trial results, I led strategy development, conducted health economics and cost-effectiveness analyses, generated tables and figures, and drafted the manuscript.

16. Tseng A, Barnabas RV, van Heerden A, Ntinga X, **Sahu M**. “Costs of home-delivered antiretroviral therapy refills for persons living with HIV: evidence from a pilot randomized controlled trial in KwaZulu-Natal, South Africa.” *PLOS Glob Public Health. Dec 2024.*

Role: This was a costing study embedded in the Deliver Health Study. I advised a PhD student on costing methods and contributed to manuscript development and revisions.

17. Klein S, Hathaway C, Umutesi G, Sharma M, **Sahu M**, Tran J, Mugo N, Barnabas R, Heitner J. Modeling the impact of PEPFAR funding cuts on HIV and cervical cancer in Kenya. (In preparation)

Role: For this mathematical modeling study developed in response to recent funding cuts for HIV programs, I advised on methods development and contributed to writing and framing.

5.7 Mass Drug Administration (MDA) of albendazole in India, Benin, & Malawi

As a continuation of work I began at the Swiss Tropical and Public Health Institute in collaboration with Drs. Arianna Rubin Means and Judd Walson from the University of Washington Department of Global Health, I contributed to the economic evaluation of the DeWorm3 field trial — a large-scale study assessing the feasibility of eliminating soil-transmitted helminths through community-wide mass drug administration (MDA) with donated albendazole.¹⁷² The drugs were provided through a pharmaceutical donation program supported by the World Health Organization and GlaxoSmithKline. My roles included advising field teams on cost data collection and quality assurance, developing an early version of the cost model, and supporting manuscript writing and revision to produce estimates of community MDA costs.

18. Galactionova K, **Sahu M**, Gideon SP, Kaliappan SP, Morozoff C, Ajjampur SSR, Walson J, Means AR, Tediosi F. “Costing interventions in the field: preliminary cost estimates and lessons from an evaluation of community-wide Mass Drug Administration for elimination of Soil-Transmitted Helminths in the DeWorm3 trial.” *BMJ Open. Jul 2021.*
19. Morozoff C, Avokpaho E, Kaliappan SP, Simwanza J, Gideon SP, Lungu W, Hounnagnon P, Galactionova K, **Sahu M**, Kalua K, Luty A, Ibikounle M, Bailey R, Pullan R, Ajjampur SSR, Walson J, Means AR. “Costs of community-wide mass drug administration and school-based deworming for soil-transmitted helminths: evidence from Benin, India, and Malawi.” *BMJ Open. Jul 2022.*

5.8 Health insurance claims data in Sub-Saharan Africa

I contributed to a study analyzing Tanzanian national health insurance claims, a project initiated during my previous role at the Swiss Tropical and Public Health Institute and completed during my PhD. Building on that work, I co-authored a commentary with Dr. Angela Apeagyei (IHME) advocating for greater use of health insurance claims data to support Universal Health Coverage in sub-Saharan Africa. My contributions included literature review, data analysis, figure development, and manuscript drafting.

20. Durizzo K, Harrtgen K, Tediosi F, **Sahu M**, Joachim A, Salari P, Guenther I. “Towards mandatory health insurance in low-income countries? An analysis of claims data in Tanzania.” *Health Econ.* Aug 2022.
21. Apeagyei A, **Sahu M**. “Claims data from health insurance programs in sub-Saharan Africa – an untapped resource to promote Universal Health Coverage.” *BMJ Glob Health.* Jul 2024.

5.9 Large collaborative efforts

Finally, I contributed to eight large-scale collaborative studies led by members of my dissertation committee and senior colleagues at IHME, as well as a smaller study on tuberculosis (TB) spending in collaboration with the World Health Organization TB team. My contributions included literature review, figure and table generation, critical manuscript feedback, and writing support as needed.

5.9.1 Global health financing

Led by Drs. Joseph Dieleman and Angela Apeagyei and IHME’s Resource Tracking Team, these studies examined global health spending patterns, development assistance, and progress toward Sustainable Development Goal (SDG) targets. I supported figure development, literature synthesis, and manuscript revisions across multiple papers.

22. Global Burden of Disease Health Financing Collaborators [including **Sahu M**]. “Health spending and spending on HIV/AIDS, tuberculosis, and malaria, and development assistance for health: progress towards Sustainable Development Goal 3.” *Lancet.* Apr 2020.
23. Global Burden of Disease Health Financing Collaborators [including **Sahu M**]. “Tracking development assistance for health and for the COVID-19 pandemic: a review of development assistance, government, out-of-pocket, and other private spending for 204 countries, 1990–2050.” *Lancet.* Sep 2021.
24. Global Burden of Disease 2021 Health Financing Collaborators [including **Sahu M**]. “Global investments in pandemic preparedness and COVID-19: development assistance and domestic spending on health between 1990 and 2026.” *Lancet Glob Health.* Jan 2023.
25. Su Y, Garcia Baena I, Harle AC, Crosby SW, Micah AE, Siroka A, **Sahu M**, Tsakalos G, Murray CJL, Floyd K, Dieleman JL. “Tracking total spending on tuberculosis by source in low- and middle-income countries, 2000–2017: an economic modelling study.” *Lancet Infect Dis.* Apr 2020.

5.7.2 Global health systems

I contributed to two global systems analyses evaluating structural drivers of access to care: one examining trends in health workforce availability and its relationship to effective Universal Health Coverage, and another assessing access to quality health care. These analyses generated cross-country estimates of health system capacity, including pharmaceutical personnel density, and highlighted persistent disparities in care quality and workforce distribution between low- and high-income settings. Both projects were led by Dr. Annie Haakenstad and reflect broader efforts to monitor progress toward Universal Health Coverage.

26. *Global Burden of Disease 2019 Healthcare Access and Quality Collaborators [including Sahu M]. “Assessing performance of the Healthcare Access and Quality Index, overall and by select age groups, for 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019.” Lancet Glob Health. Oct 2022.*
27. *Global Burden of Disease 2019 Human Resources for Health Collaborators [including Sahu M]. “Measuring the availability of human resources for health and its relationship to universal health coverage: estimates for 204 countries and territories from 1990 to 2019.” Lancet. May 2022.*

5.7.3 Global burden of infectious diseases

Finally, I contributed to three collaborative studies estimating the burden of HIV/AIDS and antimicrobial resistance. These included high-resolution mapping of HIV prevalence, incidence, and mortality across sub-Saharan Africa, led by Dr. Laura Dwyer-Lindgren, and a global analysis of antimicrobial resistance with senior involvement from Dr. Andy Stergachis. My contributions focused on manuscript and figure review during internal review stages.

28. *Global Burden of Disease 2021 Antimicrobial Resistance Collaborators [including Sahu M]. “Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050.” Lancet. Sep 2024.*
29. *Local Burden of Disease HIV Prevalence Collaborators [including Sahu M]. “Mapping age- and sex-specific HIV prevalence in adults in sub-Saharan Africa, 2000–2018.” BMC Medicine. Dec 2022.*
30. *Local Burden of Disease HIV Collaborators [including Sahu M]. “Subnational mapping of HIV incidence and mortality among adults aged 15-49 in sub-Saharan Africa, 2000-18.” Lancet HIV. Jun 2021.*

Appendix A: Supplementary Materials for Chapter 2

I. Review of existing classification systems and maps for drugs and diagnoses

The healthcare market has developed a number of systems to describe and group drugs and diagnoses, and various linkage methods exist ranging from simple maps to complex machine learning processes – though there is no comprehensive probabilistic map for NDC to health condition. This section provides a brief overview of relevant systems and mappings between them. A first step to creating our model to map drugs to diagnoses (i.e. a map of NDC to DEX cause) was to review existing systems to understand how they can be leveraged.

(a) Classification systems: A number of systems and hierarchies exist to group both drugs and diagnoses, and are detailed in **Appendix Table A.1**. Drugs can be classified according to the active ingredient or other pharmacologic properties, or the therapeutic intent. Key classification systems are described below.

Drug classifications:

- National Drug Code (NDC): A unique 11-digit, 3-segment number administered by the FDA which is included in pharmaceutical claims data. The three segments of the code identify the manufacturer, product formulation (dosage/strength), package size, and route of administration.¹⁷³ There are more than 400,000 NDCs and this number continues to grow.
- Generic drug name: A name for a medicine which is formulated in the same way as the brand name medicine (same active ingredient, strength, type of product, route of administration, indications, etc).¹⁷⁴ There are a number of maps between NDC and the generic name (REDBOOK, Multum, Medi-Span, etc), but the names of generic names are not always standardized between them.
- Anatomical Therapeutic Chemical (ATC) Classification: A five-level hierarchical naming convention to classify drugs according to the organ/system they act on as well its therapeutic properties.⁸⁸ Generic drug names can be mapped to ATC using RxNorm (Table 1).⁹⁰

Disease classifications:

- International Classification of Diseases, Clinical Modification (ICD-9-CM / ICD-10-CM): The standard diagnosis coding system used in the US, which is included in outpatient claims data. The ICD-9-CM system consists of ~13,000 codes and the ICD-10-CM expanded to ~68,000 codes.¹⁷⁵
- DEX cause list: Our team established and iterated on a list of 148 health conditions, which map to ICD-9-CM and ICD-10-CM codes and can be applied directly to outpatient claims data containing these codes.

(b) Maps: We identified a few databases which could be leveraged to assist with mapping from NDC to generic name to DEX cause. These are summarized in **Appendix Table A.1**.

(c) Healthcare episode groupers: Several opaque, proprietary software analytic tools called “healthcare episode groupers” (also know simply as “groupers”) are available for purchase alongside health care insurance data, e.g. MarketScan has its own Medical Episode Grouper for their data, as do the Centers for Medicare & Medicaid Services (CMS).⁹² These allow for bundling medical claims into clinically meaningful “episodes” and can be useful for cost-of-illness studies. However, these groupers are less appropriate for specifically grouping pharmaceutical claims data (and not the associated inpatient and outpatient data), and tend to have inflexible lists of health conditions which do not easily map to the list of 143 health conditions for the DEX project, e.g. the MarketScan Medical Episode Grouper has a strict list of approximately 500 conditions which could not be used for our purposes.

(d) Prior research: There have been a few attempts to link drugs to their indications using maps or modelling.

- AHRQ: For a small subset of drugs, the Agency for Healthcare Research and Quality (AHRQ) has cross-walked drugs and indications using the Medical Expenditure Panel Survey (MEPS) Prescribed Medicines (PMED) files, which include nationally representative information on outpatient prescribed medicines received by the US civilian non-institutionalized population.⁹¹ To create this map, PMED files were linked to the MEPS Household Component (MEPS-HC) files, which include the self-reported condition(s) for which each drug was prescribed and the respondent’s sociodemographic information (age, sex, race, insurance status, region, etc.). However, this map only exists for a small subset of drugs and is not comprehensive.
- Dexhelpp: In Austria, the *dexhelpp* project created a few iterations of a map from ATC to ICD-9 and ICD-10 codes. The first map comprised of weighted translation tables which could be more or less directly applied to the ATC.^{176,177} However, the performance of this method was variable; in one analysis, 751/1071 assignments from ATC to ICD deviated from expert opinion.¹⁷⁸ This method was later was developed into a complex model using machine learning and natural language processing methodology which performed better; however, this latter method is highly specific to the Austrian reimbursement system.¹⁷⁹ Since our project covers only the US and interested in predicting DEX causes rather than ICD codes, these tables and model have limited utility.
- DEX: Prior work from our team estimated health spending, including pharmaceutical spending, at the national level.⁸⁵ This work directly used the MEPS PMED data (and associated diagnosis information). Values were then scaled to the National Health Expenditure Accounts (NHEA) data, which gives national estimates of total US personal health spending on prescribed pharmaceuticals over time.²³ In addition, for 136 high-cost

low-volume medications which may not have been adequately captured in the MEPS data (i.e. “specialty drugs”), drugs were mapped to a single health condition for which it was most commonly used. However, given that MEPS is a relatively small nationally representative survey, it is less appropriate for county-level estimation, and this method does not leverage additional information available in the data, including patient data (age, sex, region) and drug details (dosage, route of administration, etc).

- **Cody Horst modelling:** A 2017 master’s thesis by Cody Horst, a former member of the IHME DEX team, used MarketScan claims data to develop a machine learning model to predict diagnoses from drugs.⁸⁶ This project tested a logistic regression model for binary classification as well as a character-based recurrent neural network (RNN) model. This work compared models using three metrics: precision, recall, and f1-score (which is the harmonic mean of precision and recall). While the logistic regression performed better, both models had difficulty distinguishing between comorbidities due to the noise in the data (e.g., insulin predicts hypertension because diabetes and hypertension often co-occur in the same patients). The proposed research, which is already underway, builds on this work to reduce the noise in the model.

Table A1. Existing databases to assist with linking drug codes and diagnoses.

Database	Source	Description	What is mapped?	Utility
RED BOOK ¹⁸⁰	Micromedex / IBM Watson Health / MarketScan	This supplement to the MarketScan claims data contains the generic drug name, as well as its Therapeutic Group (~20) and Therapeutic Class (~400)	NDC → generic	Reduces the dimensions needed for mapping drugs to diagnoses
openFDA ¹⁸¹	US Food & Drug Administration (FDA)	Contains the on-label (i.e. FDA approved) usage for each drug, in a long unwieldy text column called “indications and usage” which includes information on how and when to use drugs, beyond just a list of the indications	NDC → on-label indication	Not useful because it does not include the off-label uses, and the “indications and usage” column is unwieldy
Micromedex ⁹⁴	Merative (formerly IBM Watson Health)	Contains both the on- and off-label uses for each generic drug, and the strength of evidence behind each use	Generic → on- and off-label indication	Useful for our purposes, though uses are in plain language
RxNorm ⁹⁰	National Library of Medicine	Standardizes generic and branded drug names across naming conventions, e.g. ATC, Micromedex, Multum, etc.	Generic → ATC; Generic → Generic	Useful for linking claims and survey data using other naming schemes

II. Complete list of 143 Health Conditions and 24 Condition Categories included in this study

The DEX study includes 148 health conditions, of which 143 are associated with retail pharmaceutical spending.⁸² The full list, along with their corresponding 24 aggregated condition categories, is provided below.

#	Category Name (Level 2)	Cause Name	Cause Abbreviation	Notes
1	Cardiovascular diseases	Atrial fibrillation and flutter	cvd_afib	
2	Cardiovascular diseases	Aortic aneurysm	cvd_aortic	
3	Cardiovascular diseases	Cardiomyopathy and myocarditis	cvd_cmp	
4	Cardiovascular diseases	Endocarditis	cvd_endo	
5	Cardiovascular diseases	Heart Failure	cvd_hf	
6	Cardiovascular diseases	Hypertensive heart disease	cvd_htn	
7	Cardiovascular diseases	Ischemic heart disease	cvd_ihd	
8	Cardiovascular diseases	Other cardiovascular and circulatory diseases	cvd_other	
9	Cardiovascular diseases	Pulmonary Arterial Hypertension	cvd_pah	
10	Cardiovascular diseases	Lower extremity peripheral arterial disease	cvd_pvd	
11	Cardiovascular diseases	Rheumatic heart disease	cvd_rhd	
12	Cardiovascular diseases	Stroke	cvd_stroke	
13	Cardiovascular diseases	Non-rheumatic valvular heart disease	cvd_valvu	
14	Chronic respiratory diseases	Asthma	resp_asthma	
15	Chronic respiratory diseases	Chronic obstructive pulmonary disease	resp_copd	
16	Chronic respiratory diseases	Interstitial lung disease and pulmonary sarcoidosis	resp_interstitial	
17	Chronic respiratory diseases	Other chronic respiratory diseases	resp_other	
18	Chronic respiratory diseases	Pneumoconiosis	resp_pneum	
19	Other non-communicable diseases	Congenital birth defects	cong	
20	Diabetes and kidney diseases	Diabetes mellitus type 1	diabetes_typ1	
21	Diabetes and kidney diseases	Diabetes mellitus type 2	diabetes_typ2	
22	Digestive diseases	Cirrhosis and other chronic liver diseases	cirrhosis	
23	Digestive diseases	Appendicitis	digest_appendicitis	
24	Digestive diseases	Gallbladder and biliary diseases	digest_bile	

25	Digestive diseases	Inguinal, femoral, and abdominal hernia	digest_hernia	
26	Digestive diseases	Inflammatory bowel disease	digest_ibd	
27	Digestive diseases	Paralytic ileus and intestinal obstruction	digest_ileus	
28	Digestive diseases	Other digestive diseases	digest_other	
29	Digestive diseases	Pancreatitis	digest_pancreatitis	
30	Digestive diseases	Upper digestive system diseases	digest_upper	
31	Digestive diseases	Vascular intestinal disorders	digest_vascular	
32	Other non-communicable diseases	Endocrine, metabolic, blood, and immune disorders	endo	
33	Enteric infections	Enteric infections	_enteric_all	
34	Other non-communicable diseases	Gynecological diseases	gyne	
35	HIV/AIDS and sexually transmitted infections	HIV/AIDS	hiv	
36	HIV/AIDS and sexually transmitted infections	Sexually transmitted infections excluding HIV	std	
37	Other non-communicable diseases	Hemoglobinopathies and hemolytic anemias	hemog	
38	Self-harm and interpersonal violence	Interpersonal violence	_intent_agg	
39	Self-harm and interpersonal violence	Self-harm	inj_suicide	
40	Transport injuries	Transport injuries	inj_trans	
41	Unintentional injuries	Other unintentional injuries	_unintent_agg	
42	Unintentional injuries	Falls	inj_falls	
43	Unintentional injuries	Exposure to mechanical forces	inj_mech	
44	Diabetes and kidney diseases	Chronic kidney disease	ckd	
45	Diabetes and kidney diseases	Acute glomerulonephritis	glomerulo	
46	Diabetes and kidney diseases	Acute renal failure	renal_failure	
47	Maternal and neonatal disorders	Ectopic pregnancy	maternal_abort_ectopic	
48	Maternal and neonatal disorders	Maternal abortion and miscarriage	maternal_abort_mis	
49	Maternal and neonatal disorders	Maternal hemorrhage	maternal_hem	
50	Maternal and neonatal disorders	Maternal deaths aggravated by HIV/AIDS	maternal_hiv	
51	Maternal and neonatal disorders	Maternal hypertensive disorders	maternal_htn	
52	Maternal and neonatal disorders	Indirect maternal deaths	maternal_indirect	
53	Maternal and neonatal disorders	Late maternal deaths	maternal_late	

54	Maternal and neonatal disorders	Maternal obstructed labor and uterine rupture	maternal_obstruct	
55	Maternal and neonatal disorders	Other direct maternal disorders	maternal_other	
56	Maternal and neonatal disorders	Maternal sepsis and other maternal infections	maternal_sepsis	
57	Well care	Pregnancy and postpartum care	exp_well_pregnancy	
58	Mental disorders	Attention-deficit/hyperactivity disorder	mental_adhd	
59	Mental disorders	Anxiety disorders	mental_anxiety	
60	Mental disorders	Bipolar disorder	mental_bipolar	
61	Mental disorders	Conduct disorder	mental_conduct	Not included for pharmaceuticals
62	Mental disorders	Eating disorders	mental_eating	
63	Mental disorders	Idiopathic developmental intellectual disability	mental_id	
64	Mental disorders	Other mental disorders	mental_other	
65	Mental disorders	Autism spectrum disorders	mental_pdd	
66	Mental disorders	Schizophrenia	mental_schizo	
67	Mental disorders	Depressive disorders	mental_unipolar	
68	Musculoskeletal disorders	Gout	msk_gout	
69	Musculoskeletal disorders	Osteoarthritis	msk_osteoarthritis	
70	Musculoskeletal disorders	Other musculoskeletal disorders	msk_other	
71	Musculoskeletal disorders	Low back pain	msk_pain_lowback	
72	Musculoskeletal disorders	Neck pain	msk_pain_neck	
73	Musculoskeletal disorders	Rheumatoid arthritis	msk_rheumarthritis	
74	Neglected tropical diseases and malaria	Neglected tropical diseases and malaria	_ntd	
75	Maternal and neonatal disorders	Neonatal encephalopathy due to birth asphyxia and trauma	neonatal_enceph	
76	Maternal and neonatal disorders	Hemolytic disease and other neonatal jaundice	neonatal_hemolytic	
77	Maternal and neonatal disorders	Other neonatal disorders	neonatal_other	
78	Maternal and neonatal disorders	Neonatal preterm birth	neonatal_preterm	

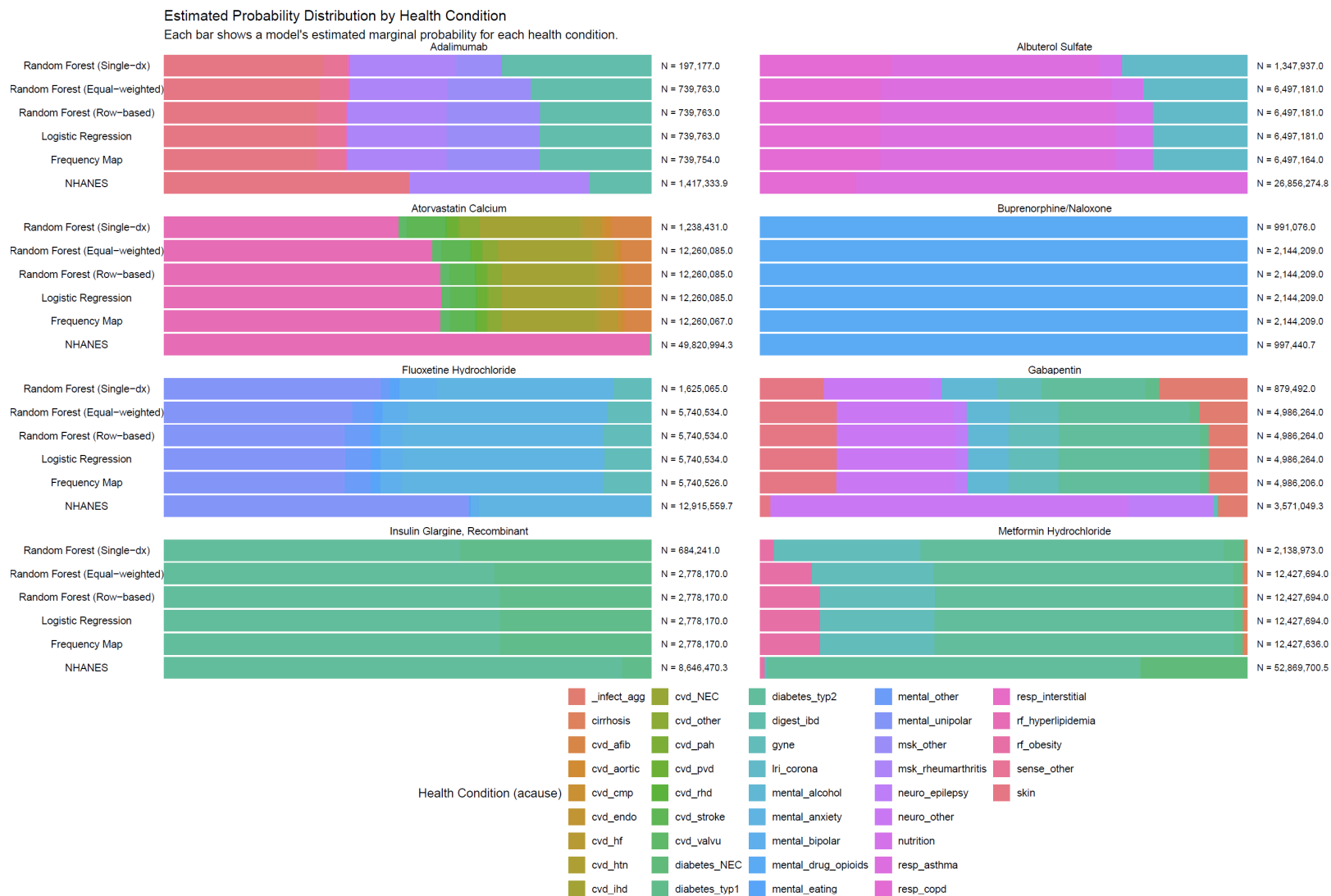
79	Maternal and neonatal disorders	Neonatal sepsis and other neonatal infections	neonatal_sepsis	
80	Neoplasms	Bladder cancer	neo_bladder	
81	Neoplasms	Malignant neoplasm of bone and articular cartilage	neo_bone	
82	Neoplasms	Brain and central nervous system cancer	neo_brain	
83	Neoplasms	Breast cancer	neo_breast	
84	Neoplasms	Cervical cancer	neo_cervical	
85	Neoplasms	Colon and rectum cancer	neo_colorectal	
86	Neoplasms	Esophageal cancer	neo_esophageal	
87	Neoplasms	Eye cancer	neo_eye	
88	Neoplasms	Gallbladder and biliary tract cancer	neo_gallbladder	
89	Neoplasms	Hodgkin lymphoma	neo_hodgkins	
90	Neoplasms	Kidney cancer	neo_kidney	
91	Neoplasms	Larynx cancer	neo_larynx	
92	Neoplasms	Leukemia	neo_leukemia	
93	Neoplasms	Liver cancer	neo_liver	
94	Neoplasms	Tracheal, bronchus, and lung cancer	neo_lung	
95	Neoplasms	Non-Hodgkin lymphoma	neo_lymphoma	
96	Neoplasms	Malignant skin melanoma	neo_melanoma	
97	Neoplasms	Mesothelioma	neo_meso	
98	Neoplasms	Lip and oral cavity cancer	neo_mouth	
99	Neoplasms	Multiple myeloma	neo_myeloma	
100	Neoplasms	Nasopharynx cancer	neo_nasopharynx	
101	Neoplasms	Neuroblastoma and other peripheral nervous cell tumors	neo_neuro	
102	Neoplasms	Non-melanoma skin cancer	neo_nmssc	
103	Neoplasms	Other neoplasms	neo_other_benign	
104	Neoplasms	Other malignant neoplasms	neo_other_cancer	
105	Neoplasms	Other pharynx cancer	neo_otherpharynx	
106	Neoplasms	Ovarian cancer	neo_ovarian	
107	Neoplasms	Pancreatic cancer	neo_pancreas	
108	Neoplasms	Prostate cancer	neo_prostate	

109	Neoplasms	Stomach cancer	neo_stomach	
110	Neoplasms	Testicular cancer	neo_testicular	
111	Neoplasms	Thyroid cancer	neo_thyroid	
112	Neoplasms	Soft tissue and other extraosseous sarcomas	neo_tissue_sarcoma	
113	Neoplasms	Uterine cancer	neo_uterine	
114	Neurological disorders	Alzheimer's disease and other dementias	neuro_dementia	
115	Neurological disorders	Idiopathic epilepsy	neuro_epilepsy	
116	Neurological disorders	Headache disorders	neuro_headach	
117	Neurological disorders	Multiple sclerosis	neuro_ms	
118	Neurological disorders	Motor neuron disease	neuro_neurone	
119	Neurological disorders	Other neurological disorders	neuro_other	
120	Neurological disorders	Parkinson's disease	neuro_parkinsons	
121	Nutritional deficiencies	Nutritional deficiencies	nutrition	
122	Other non-communicable diseases	Oral disorders	_oral	
123	Well care	Well dental	exp_well_dental	Not included for pharmaceuticals
124	Other infectious diseases	Other infectious diseases	_infect_agg	
125	Other infectious diseases	Acute hepatitis C	hepatitis_c	
126	Other infectious diseases	Meningitis	meningitis	
127	Other infectious diseases	Septicemia	septicemia	
128	Respiratory infections and tuberculosis	Lower respiratory infections	lri	
129	Respiratory infections and tuberculosis	COVID-19	lri_corona	
130	Respiratory infections and tuberculosis	Otitis media	otitis	
131	Respiratory infections and tuberculosis	Tuberculosis	tb	
132	Respiratory infections and tuberculosis	Upper respiratory infections	uri	
133	Risk factors	Treatment of hyperlipidemia	rf_hyperlipidemia	
134	Risk factors	Treatment of hypertension	rf_hypertension	

135	Risk factors	Treatment of obesity	rf_obesity	
136	Risk factors	Tobacco intervention	rf_tobacco	
137	Sense organ diseases	Age-related and other hearing loss	sense_hearing	
138	Sense organ diseases	Other sense organ diseases	sense_other	
139	Sense organ diseases	Blindness and vision loss	sense_visionlos	
140	Skin and subcutaneous diseases	Skin and subcutaneous diseases	skin	
141	Substance use disorders	Alcohol use disorders	mental_alcohol	
142	Substance use disorders	Other substance use disorders	mental_drug_agg	
143	Substance use disorders	Opioid use disorders	mental_drug_opioids	
144	Other non-communicable diseases	Urinary diseases and male infertility	urinary	
145	Well care	Donor	exp_donor	
146	Well care	Family planning	exp_family_planning	
147	Well care	Social services	exp_social_services	Not included for pharmaceuticals
148	Well care	Well person	exp_well_person	Not included for pharmaceuticals

III. Additional Results

Figure A.1: Estimated marginal probabilities for tracer drugs across models (at the DEX health condition level)



Appendix B: Supplementary Materials for Chapter 3

B1 DATA DESCRIPTION

This section describes the underlying data for retail pharmaceutical spending and utilization, race and ethnicity, prevalence and incidence, and population denominators used for this study. As detailed in the parent study by Dieleman et al (2025),¹⁸² datasets were aggregated, adjusted for comparability and representativeness, modeled to smooth and fill data gaps, and then scaled to state-level totals; accordingly, that study generated estimates of all types of care – including retail pharmaceutical spending – by age, sex, payer, health condition, and US county for 2010 to 2019. A follow-on study further disaggregated these estimates by race and ethnicity.¹⁰⁵ The present study is a deeper dive on just the retail pharmaceutical estimates, disaggregated by race and ethnicity and linked to disease prevalence.

B1.1 Retail Pharmaceutical Data

Estimates of retail pharmaceutical spending and utilization relied on a total of **27.3 billion retail prescription drug claims and encounters** from national insurance data and surveys covering all payers (private insurance, Medicare, Medicaid, and out-of-pocket payments) from 2010 to 2019. While this manuscript presents estimates for 2019 only, the estimates draw on data from the full 2010-2019 period.

Table B1.1.1: Retail pharmaceutical data used to inform our study estimates (2010-2019).

	Dataset	Source	Data Type	Payers covered	Years included	Purpose for this study	# Pharmaceutical Observations*
1	State Health Expenditure Accounts ¹⁰⁴	CMS	Modelled total estimates	All	2010-19	Envelopes for state-level estimates of retail pharmaceutical spending.	N/A
2	Market Scan Commercial Claims and Encounters (CCAE); Medicare Supplement (MS) ⁹³	IBM	Claims	Private insurance and Medicare Advantage	2010-19	Development of ‘drugs to diagnosis’ model; Estimation of total private utilization and spending [not race-specific].	3.4 billion (2.72 billion CCAE + 641 million MS)
3	Medical Expenditure Panel Survey (MEPS) ¹⁸³	AHRQ	Nationally representative survey	All	2010-19	Estimation of total and out-of-pocket utilization and spending by race/ethnicity.	3.2 million (<i>prescribed medicines files</i>)
4	Medicare Part D Event (PDE) File ¹⁸⁴	CMS	Claims	Medicare Part D (FFS + MA)	2010, 2014-16, 2019	Estimation of Medicare utilization and spending by race/ethnicity.	6.1 billion
6	Medicaid T-MSIS Analytic Files (TAF) ¹⁸⁵	CMS	Claims	Medicaid	2016, 2019	Estimation of Medicaid utilization and spending by race/ethnicity.	1.7 billion
7	Kythera Payer “Complete” Dataset ¹⁸⁶	Kythera Labs	Claims	Private insurance	2015-19	Estimation of private utilization and spending by race/ethnicity.	13.8 billion
8	Health Care Cost Institute (HCCI) 2.0 commercial claims ¹⁸⁷	HCCI	Claims	Private insurance	2012-19	Estimation of total private utilization and spending [not race-specific].	2.3 billion

*See Dieleman et al. (2025) Supplement S2 for year-by-year availability of prescribed medicines data.¹⁸²

B1.2 Race and Ethnicity Data and Missingness

We used the five minimum, mutually exclusive groups defined by the 1977 US Office of Management and Budget (OMB) guidelines: (1) American Indian/Alaskan Native non-Hispanic, (2) Asian/Pacific Islander non-Hispanic, (3) Black non-Hispanic, (4) Hispanic, and (5) White non-Hispanic.¹⁸⁸ While these categorizations are limited given that recent US census data included nearly 1,500 racial and ethnic groups including over 1,200 American Indian and Alaska Native tribes and villages,¹⁸⁹ we were constrained by what could be reasonably captured and imputed from our datasets.

Values were recoded as follows:

- **Ethnicity:** Entries containing the term ‘Hispanic’ (e.g., ‘Other-Hispanic,’ but not ‘non-Hispanic’) were recoded as Hispanic.
- **Race:** After ethnicity recoding, all race entries not falling into one of the five OMB categories listed above—including ‘unknown’, ‘other’ and ‘multi-race’—were recoded as ‘unknown.’

This section describes the available race and ethnicity data in our four data sources which included race and ethnicity information (Medicare, Medicaid, MEPS, and Kythera) and the extent of ‘unknown’ values. Race and ethnicity were more completely captured in public insurance claims data (Medicare, Medicaid) and national surveys covering all payers (MEPS), but less so in our private insurance claims database (Kythera).

As described in **Appendix B2**, we imputed missing race/ethnicity information as appropriate and used these datasets to inform the race and ethnicity small area model which was then scaled to all-population estimates produced by Dieleman et al¹⁸² which included additional non-race specific private insurance datasets (MarketScan and HCCI).

Table B1.2.1: Pharmaceutical observations and corresponding race and ethnicity information

Data Source	Payer	Race Captured?	Pharmaceutical observations [2019]	Beneficiaries [2019]	# Beneficiaries with known race/ethnicity* [2019]	% Beneficiaries with known race/ethnicity [2019]
MEPS	All	Yes	293K	28,512	27,680	97%
Medicare Part D	Medicare	Yes	609M	66,333,096	64,553,435	97%
Medicaid	Medicaid	Yes	830M	76,576,338	62,449,452	82%, but variable by state
Kythera	Private	Yes	2,887M	230,255,536	20,213,000	9%

**We define ‘known’ race/ethnicity as a race/ethnicity group which could be mapped to our 1977 OMB categories, not ‘unknown’, ‘other’, or ‘multi-race’—all of which were mapped to ‘unknown’ and considered missing. Note, these are for beneficiaries with pharmaceutical claims/data only – so coverage estimates differ slightly from those described in our imputation methods (S2.1).*

A. Medical Expenditure Panel Survey (MEPS)

The Medical Expenditure Panel Survey (MEPS) is a nationally representative household survey, covering both insured and uninsured populations.¹⁹⁰ MEPS collects self-reported data on prescription drug utilization and expenditures through its prescribed medicines files. During the interview, participants are asked to self-identify their race and ethnicity using 12 race categories and 2 ethnicity categories. Except for the 'Multiple Race' category, these align directly with the 1977 OMB classification system.¹⁸⁸ When race or ethnicity is not reported, MEPS fills in gaps using data from prior interviews or from immediate family members, resulting in high coverage of this information. For our study, individuals identifying as 'Multiple Race' were mapped to 'Other' and considered missing.

Below, we describe the distribution of demographic characteristics (age and sex) among individuals with and without known race and ethnicity data, reporting column-wise percentages.

Table B1.2.2: Race and ethnicity data coverage in the Medical Expenditure Panel Survey (MEPS) by age group and sex (2019)

Category	With 'known' race/ethnicity	Missing ('Unknown' or 'Other') race/ethnicity
Sex		
Female	14332 (52%)	417 (50%)
Male	13111 (48%)	409 (50%)
Age		
<20	6939 (25%)	360 (44%)
20-44	8057 (29%)	247 (30%)
45-64	7103 (26%)	132 (16%)
65+	5344 (19%)	87 (11%)

Notes: (1) Percents are column-wise. (2) “‘Usable’ race/ethnicity” includes respondents with race and ethnicity data coded into defined categories. “Missing” includes those marked as “Unknown” and “Other,” including ‘Multiple Race’ individuals.

B. Medicare Part D

Medicare Part D provides prescription drug coverage for Medicare and Medicare Advantage beneficiaries and includes detailed information on drug utilization and spending through Prescription Drug Event (PDE) files. Race and ethnicity data for Part D beneficiaries are not collected directly by the program but are instead sourced from the Medicare enrollment database, which relies on self-reported data originally collected by the Social Security Administration (SSA). Prior to 1980, the SSA captured only limited race information—typically White, Black, or Other. In 1980, the SSA revised its race and ethnicity categories to align with the 1977 OMB standards, introducing six standardized categories.¹⁸⁸ This change is relevant for our purposes because most Medicare beneficiaries were born before 1980, meaning their race and ethnicity information may still reflect older, less granular classifications.

In recent years, efforts to improve the accuracy of these data—such as the Research Triangle Institute (RTI) race code—have used imputation and name-based algorithms to enhance classification. We used the RTI variable in our study. Several studies have evaluated its validity.^{191–193} Prior research suggests that the RTI race code performs well for non-Hispanic Black ($\kappa = 0.96$), non-Hispanic White ($\kappa = 0.90$), and Hispanic individuals ($\kappa = 0.87$), but shows lower agreement for Asian/Pacific Islander ($\kappa = 0.77$) and American Indian/Alaska Native ($\kappa = 0.44$) groups.¹⁹¹

Below, we summarize missingness in Medicare beneficiary race and ethnicity information (RTI variable) by age, sex, and US region.

Table B1.2.3: Race and ethnicity data coverage for Medicare beneficiaries by age, sex and region (2019)

Category	With ‘usable’ race/ethnicity	Missing (‘Unknown’ or ‘Other’) race/ethnicity
Sex		
Female	35,250,023 (55%)	735,928 (41%)
Male	29,303,412 (45%)	1,043,733 (59%)
Age		
<20	7928 (0%)	519 (0%)
20-44	1,889,907 (3%)	88,326 (5%)
45-64	8,807,073 (14%)	145,131 (8%)
65+	53,848,527 (83%)	1,545,684 (87%)
US Region		
South	24,745,295 (38%)	493,795 (28%)
West	13,986,055 (22%)	488,650 (27%)
Northeast	11,702,884 (18%)	428,840 (24%)
Midwest	14,119,201 (22%)	368,376 (21%)

Notes: (1) Percents are column-wise. (2) “‘Usable’ race/ethnicity” includes respondents with race and ethnicity data coded into defined categories. “Missing” includes those marked as “Unknown” or “Other.”

C. Medicaid

Medicaid is administered jointly by states and the federal government, and each state uses its own application form to collect race and ethnicity data. While federal guidelines require adherence to 1977 OMB minimum standards—states differ in how these questions are formatted, presented, enforced, and validated. In some cases, states treat these questions as effectively optional, resulting in substantial variation in data completeness, which can range from 50% to 100%.¹⁹⁴

Some states also collect or validate race and ethnicity data through contracts with managed care organizations or by linking to patient encounter or survey data, though the methods and intensity of validation vary widely. A recent brief by the Medicaid and CHIP Payment and Access Commission (MACPAC) noted that while all state Medicaid programs report race and ethnicity data, the quality varies, with only 30 states meeting the minimum data quality standards

necessary in the 2019 data.¹⁹⁴ These inconsistencies affect both the quality and comparability of race and ethnicity data in Medicaid.

In our analysis, we use the Transformed Medicaid Statistical Information System Analytic Files (TAF), which reflect these underlying state-level differences. Since we did not perform any corrections to the underlying data, our estimates are constrained by the data quality of self-reported race and ethnicities.

In the following table, we summarize missingness in Medicaid beneficiary race and ethnicity information by age, sex, and US state.

Table B1.2.4: Race and ethnicity data coverage for Medicaid beneficiaries by age, sex, and US state (2019)

Category	With 'usable' race/ethnicity	Missing ('Unknown' or 'Other') race/ethnicity
Sex		
Female	34,835,217 (56%)	7,561,099 (54%)
Male	27,614,235 (44%)	6,565,788 (46%)
Age		
<20	27,564,418 (44%)	7,138,115 (51%)
20-44	18,226,044 (29%)	3,713,115 (26%)
45-64	10,513,543 (17%)	2,103,048 (15%)
65+	6,145,446 (10%)	1,172,608 (8%)
US State		
Alabama	383,590 (1%)	663,659 (5%)
Alaska	198,976 (0%)	27,213 (0%)
Arizona	1,302,407 (2%)	641,521 (5%)
Arkansas	663,010 (1%)	237,099 (2%)
California	12,678,100 (20%)	1,086,813 (8%)
Colorado	747,937 (1%)	483,939 (3%)
Connecticut	609,606 (1%)	352,154 (2%)
Delaware	235,574 (0%)	467 (0%)
District of Columbia	194,472 (0%)	70,858 (1%)
Florida	3,770,273 (6%)	663,448 (5%)
Georgia	1,771,118 (3%)	271,885 (2%)
Hawaii	261,392 (0%)	81,477 (1%)
Idaho	266,079 (0%)	1,255 (0%)
Illinois	2,642,154 (4%)	136,490 (1%)
Indiana	1,282,942 (2%)	152,688 (1%)
Iowa	454,419 (1%)	179,833 (1%)
Kansas	72,768 (0%)	304,390 (2%)
Kentucky	1,294,414 (2%)	250,263 (2%)
Louisiana	1,048,738 (2%)	578,020 (4%)
Maine	265,151 (0%)	26,010 (0%)
Maryland	1,134,363 (2%)	286,533 (2%)
Massachusetts	845,257 (1%)	855,420 (6%)
Michigan	2,278,587 (4%)	191,960 (1%)

Minnesota	974,553 (2%)	83,760 (1%)
Mississippi	580,452 (1%)	87,500 (1%)
Missouri	706,337 (1%)	221,113 (2%)
Montana	222,265 (0%)	30,576 (0%)
Nebraska	235,151 (0%)	19,363 (0%)
Nevada	619,594 (1%)	19,823 (0%)
New Hampshire	172,472 (0%)	23,690 (0%)
New Jersey	1,434,619 (2%)	166,405 (1%)
New Mexico	822,126 (1%)	13,741 (0%)
New York	4,383,822 (7%)	1,706,781 (12%)
North Carolina	2,084,024 (3%)	90,186 (1%)
North Dakota	86,286 (0%)	6,159 (0%)
Ohio	2,583,952 (4%)	204,052 (1%)
Oklahoma	643,939 (1%)	28,870 (0%)
Oregon	858,592 (1%)	159,494 (1%)
Pennsylvania	2,516,053 (4%)	141,412 (1%)
Rhode Island	1,327 (0%)	206,548 (1%)
South Carolina	823,518 (1%)	457,714 (3%)
South Dakota	115,843 (0%)	556 (0%)
Tennessee	13,141 (0%)	1,513,255 (11%)
Texas	3,600,919 (6%)	689,932 (5%)
Utah	175,588 (0%)	120,021 (1%)
Vermont	140,737 (0%)	28,680 (0%)
Virginia	1,347,105 (2%)	50,656 (0%)
Washington	1,586,872 (3%)	163,882 (1%)
West Virginia	359,660 (1%)	160,696 (1%)
Wisconsin	909,400 (1%)	175,866 (1%)
Wyoming	49,781 (0%)	12,760 (0%)

Notes: (1) Percents are column-wise. (2) “‘Usable’ race/ethnicity” includes respondents with race and ethnicity data coded into defined categories. “Missing” includes those marked as “Unknown” or “Other.”

D. Kythera

Kythera is a commercial claims dataset that includes race and ethnicity information based on a combination of self-reported, provider-reported (from electronic health record data), and imputed data. Race and ethnicity are imputed by a third-party vendor using patient demographics and consumer information, and may also incorporate survey data. As typical of many commercial administrative claims databases,¹²⁶ the validation methods are not published or well described.

Race and ethnicity data were available for only 12% of enrollees; we excluded the remaining observations from race- and ethnicity-stratified analyses assuming the remaining 88% were missing at random, i.e., assuming that missingness was unrelated to individuals’ actual race or ethnicity. Despite low coverage, the dataset still included 20.2 million beneficiaries with usable race and ethnicity information. The large sample size, combined with nationally representative survey data (MEPS), supported generally robust estimation for the privately insured population.

Table B1.2.5: Race and ethnicity data coverage in for private insurance beneficiaries in Kythera data by age, sex, and region (2019)

Category	With 'usable' race/ethnicity	Missing ('Unknown' or 'Other') race/ethnicity
Sex		
Female	11,917,642 (59%)	115,856,652 (55%)
Male	8,295,358 (41%)	94,185,884 (45%)
Age		
<20	2,360,588 (12%)	50,610,536 (24%)
20-44	4,829,304 (24%)	58,808,479 (28%)
45-64	6318573 (31%)	54,991,639 (26%)
65+	6,704,535 (33%)	45,631,882 (22%)
US Region		
South	8,759,606 (43%)	80,317,389 (38%)
West	4,121,188 (20%)	43,880,251 (21%)
Northeast	3,359,561 (17%)	41,199,194 (20%)
Midwest	3,915,305 (19%)	44,178,802 (21%)

Notes: (1) Percents are column-wise. (2) “‘Usable’ race/ethnicity” includes respondents with race and ethnicity data coded into defined categories. “Missing” includes those marked as “Unknown” or “Other.”

B1.3 Race-Specific Prevalence and Incidence

As described in Beauchamp and Thomson,¹⁰⁵ we used estimates from the GBD 2021 study¹⁹⁵ of condition-specific incidence (for cancers and injuries) or prevalence (for all other health conditions) by state, age, and sex for 52 causes. These were combined with race- and ethnicity-specific mortality estimates from IHME⁸⁴ to generate incidence and prevalence by race and ethnicity, using estimated incidence-to-mortality or prevalence-to-mortality ratios (by state, age, sex, and race/ethnicity). We calculated spending per incident or prevalent case, referring to both as “spending per prevalent case” for simplicity.

B1.4 Population Denominators

Finally, we estimated coverage of each payer by race and ethnicity in order to generate estimates of pharmaceutical spending and utilization by payer and disaggregated by our other dimensions (age, sex, race/ethnicity, and US state). These coverage rates were specific to 6 insurance types: 1) private insurance only, 2) Medicaid only, 3) Medicare only, 4) uninsured, 5) Medicare and Medicaid, and 6) Medicare and private. For Medicare, we distinguished between Fee-For-Service coverage and managed care (Medicare Advantage) coverage. Further detail can be found in the Supplement to Beauchamp and Thomson,¹⁰⁵ but in brief, several sources were used to determine the fraction of each race/ethnicity that comprised the all-population payer denominators (the methods of which can be found in pages 81-84 of *Dieleman et al, 2025*¹⁸²).

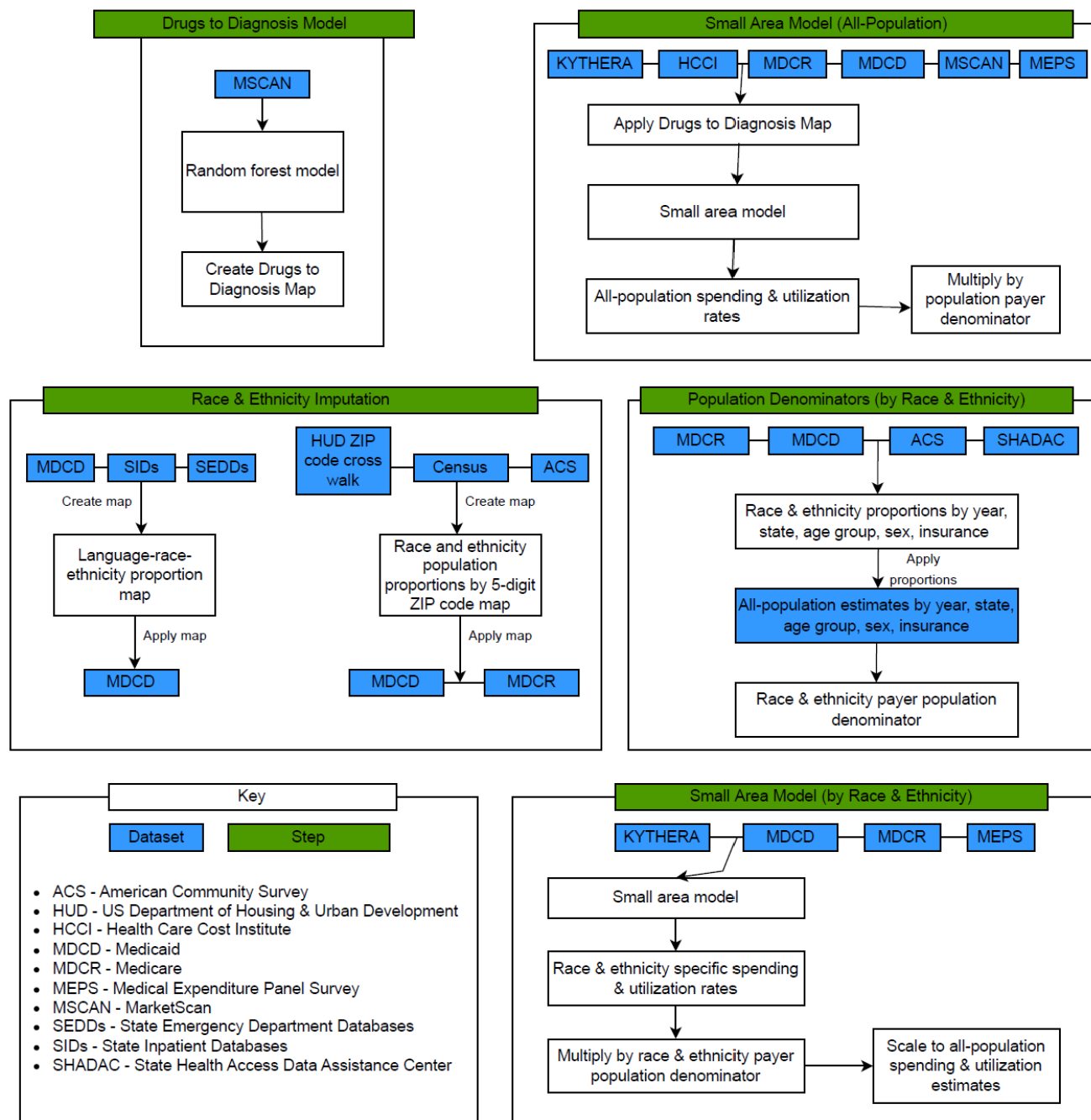
To generate the race fractions for the uninsured, privately insured, and those with Medicare and private insurance, we extracted data for years 2009 to 2019 from the American Community Survey (ACS)¹⁹⁶ and data for years 2008 to 2019 from the State Health Access Data Assistance Center (SHADAC).¹⁹⁷ Data were proportionally split into the 19 age bins and 2 sexes in our final

estimates using the all-race population denominators as a reference. Smooth time trends between 2000 and 2019 were ensured using loess and linear regression. To generate race fractions for Medicaid and Medicare, we used the enrollment data to create a set of race proportions by age group and sex. We fit a loess regression for years 2009 to 2019 to create a complete time trend and then a linear regression to predict estimates back to 2000. All race fractions were then multiplied by all-population payer-specific totals.

B2 IMPUTATION AND MODELLING METHODS

The steps involved in the estimation of pharmaceutical utilization and spending are described in detail in ‘Tracking US County Health Spending by Health condition, 2010-2019,’¹⁸² with key elements summarized here. State-level estimates of prescription drug utilization and spending were then disaggregated by race and ethnicity as described in further detail here and in a forthcoming paper.¹⁰⁵

Figure B2.0: Overview of datasets and steps used to generate estimates of pharmaceutical spending and utilization by race and ethnicity



B2.1 Race and Ethnicity Imputation

Where race and ethnicity were missing, we applied imputation methods consistent with recommendations from the Institute of Medicine (US) Subcommittee on Standardized Collection of Race/Ethnicity,¹⁹⁸ as briefly outlined in **Appendix B2 Imputation and Modelling Methods** and discussed in greater detail in a forthcoming paper by Beauchamp and Thomson.¹⁰⁵

Missingness of race and ethnicity information across our datasets is described in Section S.1.2. Overall, 12.7% of records were missing race/ethnicity information across MEPS, Medicaid, and Medicare Part D (RTI);¹⁰⁵ the proportion of missing data was lower for retail pharmaceutical data only (**Table B1.2.1**). Race and ethnicity were imputed for Medicare and Medicaid only. For Kythera, which had limited race/ethnicity coverage (~12%; see S1.2D), we used available data to inform distributions, scaling them to payer-specific totals estimated from Kythera, MarketScan, and HCCI (**Section B2.4**).

Table B2.1.1: Dataset and whether imputation was performed

Dataset	Race/Ethnicity Imputed?	Notes
MEPS	No	High completeness; no imputation performed.
Medicare (RTI)	Yes	Race/ethnicity imputed.
Medicaid	Yes	Race/ethnicity imputed.
Kythera	No	Poor coverage (~12% had race/ethnicity). Assumed missingness at random and dropped all data from beneficiaries with unknown race/ethnicity.

To impute missing race and ethnicity data for Medicare and Medicaid, we used a two-step modeling process. These methods are described in more detail in the forthcoming manuscript ‘Tracking US State Health Spending by Race and Ethnicity, 2010-2019,’¹⁰⁵ and are summarized here. The imputation approach drew on not only pharmaceutical data but also the State Inpatient Databases (SIDs) and State Emergency Department Databases (SEDDs), which include race and ethnicity information.

A. Language to race and ethnicity mapping

Using Medicaid as well as SIDs and SEDDs, which record both the patient’s primary language and their race and ethnicity, we generated a map from language to race and ethnicity. All native languages present in our data (Achinese, Aleut, Algonquian languages, Apache languages, Central American Indian languages, Choctaw, Dakota, Guarani Native American Language, Haida, Inuktitut, Inupiaq, Mi’kmaq, North American Indian languages, Navajo, Quechua, Salishan languages, Siouan languages, Zuni) were assigned to American Indian/Alaska Native. In addition, a primary language of Spanish was always assigned to Hispanic ethnicity. Then, all remaining data containing valid language values other than English and valid race values were used to generate a set of proportions of each race and ethnicity within a language. This map was validated using Medicaid, SIDS and SEDDS data where both language and race and ethnicity were known.

Table B2.1.2: Dataset and whether primary language was included

Dataset	Primary Language Included?	% of total observations with imputed race based on primary language	% observations with imputed matched self-report race/ethnicity
Medicaid	Yes	2.13%	99.2%
Medicare	No	N/A	N/A
SEDDs	Yes	0.03%	93.5%
SIDs	Yes	0.03%	88.5%

B. Probabilistic zip code level imputation

The Decennial Census and American Community Survey (ACS) report population data by race and ethnicity at the census tract level.^{199,200} We applied the U.S. Department of Housing & Urban Development (HUD) zip code crosswalk files³ to the census and ACS data to map race population proportions from census tracts to 5-digit zip code geographic areas. Medicaid and Medicare beneficiaries with outstanding missing race information after language imputation but known 5-digit ZIP code were probabilistically assigned a race based on the racial composition of the population in that ZIP code, i.e. each individual was randomly assigned a race according to the ZIP code's population distribution. This map was validated using Medicaid, Medicare, SIDs, and SEDDs data where both zip code and race and ethnicity were known.

Table B2.1.3: Dataset and whether zip code was available

Dataset	Zip code included?	% of total observations with imputed race based on zip code	% observations with imputed matched self-report race/ethnicity
Medicaid	Yes	11.4%	57.2%
Medicare	Yes	1.6%	76.3%
SEDDs	Yes	8.3%	60.7%
SIDs	Yes	11.0%	68.6%

B2.2 Drugs to Diagnosis Model

The methods for linking drugs to their corresponding indications are published as part of the methods appendix in a recent study, ‘Tracking US County Health Spending by Health Condition, 2010-2019,’ in **Section B3.1H National Drug Codes (NDCs) for pharmaceuticals** – the excerpt from this appendix is provided below with minimal edits.¹ A forthcoming paper from Sahu and Weil (focused on off-label prescribing) further describes the internal and external validation as well as the model section process.

A. National Drug Codes (NDCs) for pharmaceuticals

Prescription drug claims data contain 11-digit FDA-assigned National Drug Codes (NDCs) identifying pharmaceuticals by the manufacturer, product formulation, and package details, but not the diagnoses for which they are prescribed. We probabilistically mapped NDCs to the list of health conditions outlined in S3.1E in a multi-step modeling process as follows: (1) using MarketScan retail pharmaceutical and outpatient visit data, we linked prescription drug claims to the most recent outpatient visit occurring up to 30 days before the pharmaceutical claim, (2) we restricted the data to on- and off-label uses from the Merative Micromedex pharmaceutical compendium, which we also manually mapped to our 143 conditions, and (3) we used a random forest classification model to predict one of 143 on- and off-label uses given the generic drug name, for 8 age/sex groups and leveraging drug details including dosage, package size, and route of administration, and (4) we probabilistically applied the model to all our pharmaceutical data containing NDCs including Medicare, Medicaid, private, and out-of-pocket spending.

B. MarketScan data processing

To develop our training dataset, we first linked MarketScan pharmaceutical claims and diagnoses from outpatient visits as follows. The pharmaceutical claims included NDCs, which we linked to their corresponding generic drug names using Market Scan’s Redbook. We restricted the data to only original prescriptions (not refills), to capture claims likely to be associated with a recent provider visit. Using the beneficiary ID, prescription drug claims were then linked to the most recent outpatient visit occurring within 30 days prior to the date of the pharmaceutical claim; this 30-day window aligns with the duration often allotted to pharmacies for reimbursement by insurers. The outpatient visit data contained up to 4 diagnoses, recorded using ICD codes, per visit. These ICD codes were then comprehensively mapped to the 143 health conditions using a process described in detail in Dieleman et al.¹ This resulted in a dataset with pharmaceutical claims linked to up to 4 diagnoses (i.e., up to 4 rows per pharmaceutical claim), but since diagnoses were not necessarily the condition for which the drug was prescribed, this dataset was insufficient for our purposes because it included a large number of commonly occurring co-morbidities (e.g., insulin being linked to both Type 2 Diabetes and hypertension).

Table B2.2.1: MarketScan data used to inform classification model

Pharmaceutical observations [2019]	Beneficiaries [2019]
240 million (196 Commercial Claims & Encounters + 44 Medicare Advantage)	20,159,606

C. Restriction using the Merative Micromedex pharmaceutical compendium

To restrict our training dataset to only plausible drug-cause pairs, we used the Merative Micromedex database which is a pharmaceutical compendium that describes both the on- and off-label indications for each drug's generic name. Because the indications are given in plain English language rather than any standardized coding (such as ICDs), we had to manually map indications listed in Micromedex to the set of 148 health conditions in our analysis. We conducted this mapping for 755 drugs that covered >95% of total spending, utilization, and out-of-pocket spending in the MarketScan data. Drug-cause pairs were then further restricted to plausible demographics by age and sex. The resulting dataset included 160,329,831 rows of plausible drug-cause pairs on which we trained our model. Figure 2 gives a full description of the data processing and restriction steps to generate the final training and test datasets for the model.

D. Random forest model

To predict the probability of each of our 143 health conditions for each drug, we trained a random forest model with 1000 estimators. We ran a separate model for each of our 755 drugs and 8 age/sex groups, and included as covariates the package size, dose strength, the route of administration, and the total spending. For the remainder of the ~8500 drugs covering <5% of spending, utilization and out-of-pocket spending, we applied the same model to the dataset that included restrictions by age and sex but not by indication. After modelling, the output probabilities from the random forest models were used to estimate the likelihood of a NDC being prescribed to a condition. Probabilities for a condition were computed as the average prediction probability for that condition across all trees, where the predicted probability for a single tree is the fraction of leaves that result in the condition. Full details of the model selection and validation to choose this model are described in a forthcoming paper from Sahu and Weil.

E. Model Application

Finally, these probabilities (which were specific to generic drug name and age/sex group) were then used to assign health conditions across our prescription claims data from Medicare, Medicaid, HCCI, Kythera, and MEPS; specifically, each pharmaceutical claim with an NDC was probabilistically assigned to a single health condition from the set of possible conditions associated with that code.

B2.3 Small Area Modelling

Our all-population pharmaceutical spending and utilization estimates (and associated uncertainty) were generated using a small area model as described in the detailed methods appendix of the published study ‘Tracking US County Health Spending, 2010-2019.’¹⁸² The full model specification is available on pages 69-70 of the published appendix. The model was implemented using Template Model Builder (TMB), an open-source R package designed for fitting nonlinear random effects models.²⁰¹ Race-specific estimates were generated using a similar modeling approach, with an additional level of disaggregation, as detailed in the

forthcoming study ‘Tracking US State Health Spending by Race and Ethnicity, 2010-2019.’¹⁰⁵ The full model specification, which is similar to that for the all-population estimates but includes an additional term for race/ethnicity, is described in the methods appendix of Beauchamp and Thomson et al.¹⁰⁵

B2.4 Post-Model Steps and Scaling

Two steps were required to ensure that our race/ethnicity results summed to all-population estimates, which are scaled to the total expenditures reported by the government State Health Expenditure Account (SHEA). First, we needed to ensure that for every all-population estimate combination,¹⁸² (geography-type of care-primary payer-payer-sex- health condition) we had at least one race-specific estimate, such that all race-specific estimates summed to the total population. This step was referred to as post model imputation, and more detail is available in the Supplement to *Tracking US State Health Spending by Race and Ethnicity, 2010-2019*.¹⁰⁵

The second step, used for generating all-population estimates, ensured that all estimates summed to the total expenditures reported by SHEA (state, year, payer, and category—“type of care”—specific). SHEA data was downloaded from the official Centers for Medicare & Medicaid Services website²⁰² and converted to 2019 currency. Our state-level retail pharmaceutical expenditure envelopes were the sum of two categories in the SHEA data, ‘Prescription Drug Expenditures’ and ‘Other Non-Durable Medical Products Expenditures.’ The SHEA rebate adjustment uses Medical Loss Ratio data to adjust retail pharmaceutical sales data.¹¹⁰

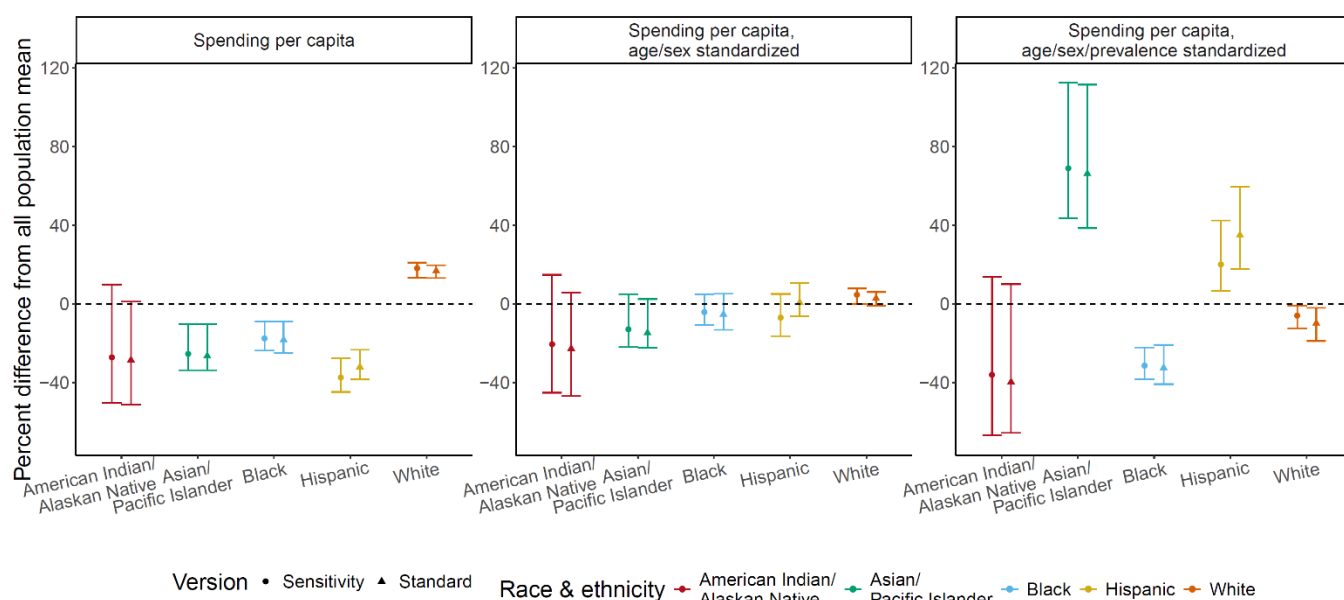
B2.5 Validation

Given the substantial missingness in race and ethnicity data and our reliance on imputation in some cases, we conducted sensitivity analyses to test the robustness of our findings: (1) we re-estimated results without imputing race or ethnicity for any individuals (note: imputation was used only for Medicare and Medicaid data); (2) to assess whether our private insurance estimates were biased due to poor data quality in the Kythera dataset, we compared our national-level estimates to those from MEPS, a nationally representative survey (note: both Kythera and MEPS data were used to generate our private insurance estimates).

A. Imputation sensitivity analysis

We conducted a sensitivity analysis of our imputation methods using language and zip information by creating estimates with and without imputation of missing race and ethnicity information. We found a negligible difference between the two.

Figure B.5.1: Sensitivity analysis national pharmaceutical spending per capita variation by race and ethnicity, 2019



The circle represents the sensitivity analysis point estimate while the triangle represents the standard analysis. The facets from left to right represent the unstandardized, age/sex standardized, and age/sex/prevalence standardized spending. The far-right panel includes spending for only the 52 causes for which prevalence information was available while the left and middle panels include spending for all 143 causes. The dashed line indicates the all-population weighted mean, where each data point is the percent difference from that value. Error bars indicate 95% uncertainty intervals (UIs). UIs were based on completing the analysis on 50 independent samples of the underlying data.

B. Private insurance benchmarking

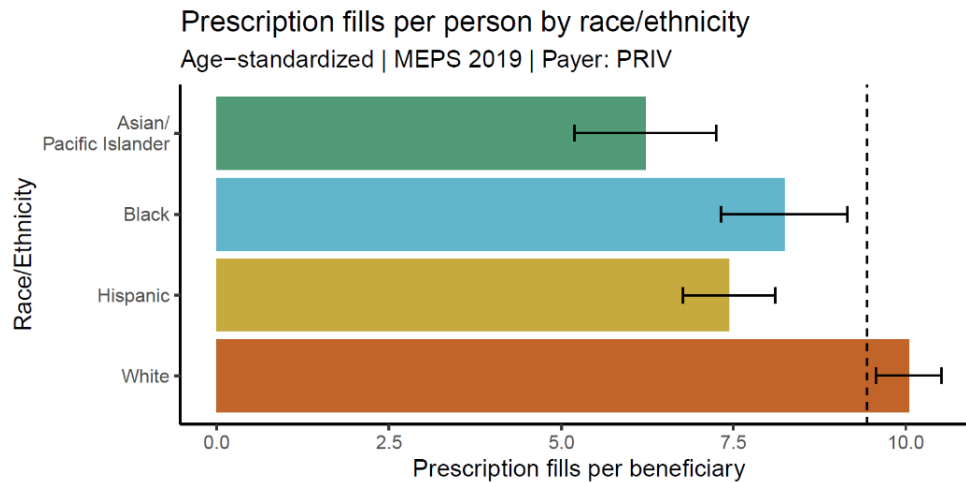
Finally, to assess the reasonableness of our private insurance estimates—which draw from Kythera data with substantial missingness as well as the nationally representative Medical Expenditure Panel Survey (MEPS)—we benchmarked our national-level estimates against age- and sex-standardized MEPS data for privately insured individuals for 2019. MEPS is nationally representative across race/ethnicity groups, and allows for national disaggregation of pharmaceutical utilization and spending by both payer type and race/ethnicity. Prior literature using MEPS, suggests that even among privately insured individuals, White beneficiaries use health care at a higher rate than other groups.¹²⁷

As shown in **Figure B.5.2**, our modeled estimates align with MEPS-derived estimates for age- and sex-standardized prescription fills per privately insured beneficiary across racial and ethnic groups. In **Panel A**, we present nationally representative, age- and sex-standardized estimates of 2019 prescription fills and pharmaceutical spending per beneficiary using MEPS, applying survey weights and accounting for the complex survey design. **Panel B** is from the main paper Figure 2, using our study estimates (DEX). Both demonstrate a similar pattern: White individuals had the highest fill rates, Asian/Pacific Islander individuals the lowest, and Hispanic and Black individuals had somewhere in between. Although absolute values and exact rank ordering differ

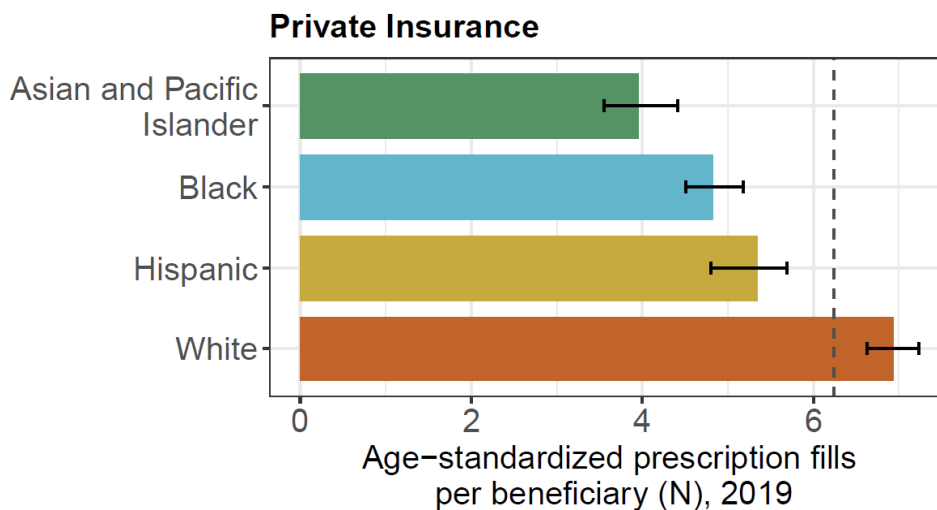
slightly, the general alignment supports the validity of our model-based estimates despite limitations in the underlying Kythera dataset.

Figure B.5.2: Benchmarking prescription drug utilization per beneficiary for private insurance using the Medical Expenditure Panel Survey (MEPS), 2019

A) MEPS-derived estimates, age-standardized



B) DEX estimates, age-standardized (main paper Figure 2)



B3 NATIONAL-LEVEL SUPPLEMENTAL RESULTS

A. Table B3.1: *Pharmaceutical spending and utilization per capita by race and ethnicity (age-standardized), for 143 health conditions*

Here we present spending and utilization per capita by race/ethnicity, for each of 143 health conditions included in the US Disease Expenditure (DEX) project. Health conditions are ranked by total retail pharmaceutical spending. All groups other than Hispanic are non-Hispanic ethnicity. The mean values across 100 draws are presented in the top row, with the 95% Uncertainty Interval below.

Rank (total drug spending)	Health Condition	Spending per capita (US\$)				Prescription fills per capita (n)			
		Asian and Pacific Islander	Black	Hispanic	White	Asian and Pacific Islander	Black	Hispanic	White
1	Diabetes mellitus type 2	149.1 (130.2, 177.2)	193.3 (173.6, 221.3)	213.7 (194.2, 236.5)	167.8 (160.6, 176)	0.97 (0.9, 1.1)	1.13 (1, 1.3)	1.2 (1.1, 1.3)	0.71 (0.7, 0.7)
2	Skin and subcutaneous diseases	59.1 (46.5, 79.5)	45.6 (37, 54.6)	63.7 (51.4, 76.2)	83.1 (78.4, 86.8)	0.42 (0.4, 0.5)	0.35 (0.3, 0.4)	0.48 (0.4, 0.6)	0.43 (0.4, 0.4)
3	Endocrine, metabolic, blood, and immune disorders	38.8 (28.8, 52.4)	27.7 (24.8, 31.3)	45.8 (39.9, 51.6)	69.6 (67, 72.4)	0.38 (0.3, 0.5)	0.26 (0.2, 0.3)	0.45 (0.4, 0.5)	0.47 (0.5, 0.5)
4	Treatment of hypertension	42.1 (36.8, 48.7)	74.3 (67, 80)	37.6 (34.2, 42.5)	40.5 (39.1, 41.7)	1.07 (0.9, 1.2)	1.41 (1.3, 1.5)	0.9 (0.8, 1)	0.8 (0.8, 0.8)
5	Other musculoskeletal disorders	40.7 (32.3, 53.8)	31.7 (25.8, 38.7)	32.6 (27.2, 39.4)	47.3 (45.2, 49.5)	0.23 (0.2, 0.3)	0.22 (0.2, 0.2)	0.21 (0.2, 0.2)	0.22 (0.2, 0.2)
6	HIV/AIDS	18.3 (7.2, 45.3)	60.5 (42, 94.6)	76.3 (38, 124.4)	26.1 (12, 38.6)	0.01 (0, 0)	0.03 (0, 0)	0.03 (0, 0)	0.02 (0, 0)
7	Multiple sclerosis	8.9 (4.1, 17.2)	29.3 (19.2, 47)	15.7 (8.1, 27.6)	43.7 (38.6, 48.8)	0 (0, 0)	0.01 (0, 0)	0 (0, 0)	0.01 (0, 0)
8	Atrial fibrillation and flutter	58.4 (39.9, 79.7)	24.9 (21.2, 31)	41.4 (32.8, 55.2)	47.4 (43.4, 51.8)	0.27 (0.2, 0.4)	0.14 (0.1, 0.2)	0.22 (0.2, 0.3)	0.2 (0.2, 0.2)
9	Asthma	16.2 (13.4, 20.1)	20.5 (18.8, 22.5)	25.8 (22.8, 30.3)	30.5 (29.6, 31.6)	0.12 (0.1, 0.1)	0.17 (0.2, 0.2)	0.19 (0.2, 0.2)	0.2 (0.2, 0.2)
10	Rheumatoid arthritis	9.3 (4.9, 18.6)	14.3 (11.1, 20.4)	31.8 (21.8, 38.7)	28.3 (26.2, 31.1)	0.02 (0, 0)	0.03 (0, 0)	0.05 (0, 0.1)	0.05 (0, 0)

11	Treatment of hyperlipidemia	25	18.8	24.1	22.9	0.43	0.34	0.41	0.35
		(21, 28.8)	(16.8, 21.8)	(21.8, 27.6)	(21.7, 23.9)	(0.4, 0.5)	(0.3, 0.4)	(0.4, 0.4)	(0.3, 0.4)
12	Chronic obstructive pulmonary disease	25.8	18.3	22	28.9	0.13	0.12	0.14	0.12
		(17.7, 36)	(16.1, 21.1)	(17.3, 28.5)	(27.9, 30.1)	(0.1, 0.2)	(0.1, 0.1)	(0.1, 0.2)	(0.1, 0.1)
13	Inflammatory bowel disease	17.9	12.2	15.8	24.3	0.03	0.03	0.04	0.05
		(7.5, 41.3)	(7, 18.6)	(9.9, 24.6)	(20.8, 27)	(0, 0)	(0, 0)	(0, 0)	(0, 0.1)
14	Leukemia	31.5	28	25.7	15.5	0.01	0.01	0.01	0
		(16, 62.6)	(18.5, 40.6)	(15.2, 39)	(12.9, 18.4)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
15	Multiple myeloma	45	38.3	36.7	19.5	0.01	0.01	0.01	0
		(18.6, 85.7)	(24.9, 56.2)	(22.2, 55.9)	(16, 23.5)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
16	Urinary diseases and male infertility	13.9	14.4	17.7	19.3	0.21	0.21	0.26	0.22
		(10.7, 18)	(12.7, 16.1)	(15, 20.2)	(18.8, 19.8)	(0.2, 0.3)	(0.2, 0.2)	(0.2, 0.3)	(0.2, 0.2)
17	Attention-deficit/hyperactivity disorder	3.5	6.1	7.2	26.3	0.02	0.04	0.05	0.16
		(2.5, 5.4)	(5.3, 7.2)	(6.1, 8.4)	(24.7, 27.8)	(0, 0)	(0, 0)	(0, 0.1)	(0.2, 0.2)
18	Breast cancer	16.4	24.8	23.8	20.7	0.02	0.04	0.04	0.04
		(9, 29.3)	(16.5, 35.2)	(14.8, 37.4)	(19.3, 22.8)	(0, 0)	(0, 0.1)	(0, 0.1)	(0, 0)
19	Ischemic heart disease	14.2	13.4	13.8	21	0.29	0.3	0.32	0.32
		(10.1, 19.4)	(11.4, 15.8)	(11.6, 17.2)	(20.2, 22.2)	(0.2, 0.4)	(0.3, 0.3)	(0.3, 0.4)	(0.3, 0.3)
20	Depressive disorders	4.5	7.2	9.8	19.9	0.11	0.17	0.25	0.38
		(3.4, 6.4)	(6.3, 8)	(8.1, 11.1)	(18.8, 21.5)	(0.1, 0.2)	(0.1, 0.2)	(0.2, 0.3)	(0.4, 0.4)
21	Prostate cancer	31.7	37.8	46.1	14	0.02	0.02	0.02	0.01
		(15.8, 51.4)	(27.8, 54.7)	(28.7, 68.9)	(11.8, 15.9)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
22	Diabetes mellitus type 1	26.7	16.1	18.1	12.4	0.05	0.03	0.03	0.03
		(14.4, 45.6)	(10.2, 26)	(12.2, 25.1)	(10, 15)	(0, 0.1)	(0, 0)	(0, 0)	(0, 0)
23	Other chronic respiratory diseases	11.1	10.7	11.3	16.8	0.16	0.17	0.2	0.23
		(8.3, 14.4)	(9.3, 12.6)	(9.5, 13.4)	(15.9, 17.5)	(0.1, 0.2)	(0.1, 0.2)	(0.2, 0.2)	(0.2, 0.2)
24	Other neurological disorders	7.6	12	9.1	16	0.09	0.16	0.12	0.19
		(5.2, 10.5)	(10.3, 13.9)	(6.9, 12.1)	(15.4, 16.5)	(0.1, 0.1)	(0.1, 0.2)	(0.1, 0.1)	(0.2, 0.2)
25	Other neoplasms	16.2	17.9	16.6	12.6	0.03	0.04	0.04	0.04
		(9.9, 22.5)	(12.8, 24.2)	(11.7, 21.1)	(11.4, 13.5)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
26	Other sense organ diseases	15.5	12.9	13.6	12.5	0.1	0.09	0.09	0.07
		(11.4, 21)	(10.2, 15.8)	(9.9, 17.1)	(11.7, 13.2)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)
27	Upper digestive system diseases	7.1	9.3	11.7	12.9	0.19	0.24	0.28	0.27
		(5.5, 8.7)	(8.2, 10.5)	(10.1, 14.1)	(12.1, 13.8)	(0.2, 0.2)	(0.2, 0.3)	(0.2, 0.3)	(0.3, 0.3)

28	Anxiety disorders	3.4	5.3	5.4	15.5	0.09	0.17	0.22	0.48
		(2.7, 4.4)	(4.8, 6.2)	(4.8, 6.5)	(14.7, 16.7)	(0.1, 0.1)	(0.2, 0.2)	(0.2, 0.2)	(0.5, 0.5)
29	Tracheal, bronchus, and lung cancer	42.1	13.4	32.1	9.7	0.01	0.01	0.01	0
		(24.2, 70.4)	(8.1, 22.8)	(17.6, 53.8)	(7.4, 12.2)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
30	Other digestive diseases	8.1	7.8	13.1	10.6	0.06	0.05	0.08	0.07
		(5.3, 12.3)	(6.1, 14.2)	(9.8, 17.3)	(9.7, 11.5)	(0, 0.1)	(0, 0.1)	(0.1, 0.1)	(0.1, 0.1)
31	Other infectious diseases	8.1	7.5	6	10.5	0.08	0.1	0.09	0.13
		(5.9, 11.7)	(5.9, 9.9)	(5.2, 7.6)	(10, 11.1)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)
32	Bipolar disorder	3.7	9.3	5	12.7	0.02	0.06	0.04	0.09
		(2.5, 6.5)	(8.1, 11.4)	(4.3, 5.7)	(11.9, 13.6)	(0, 0)	(0.1, 0.1)	(0, 0)	(0.1, 0.1)
33	Gynecological diseases	4.3	5.7	4.9	13.2	0.07	0.1	0.09	0.16
		(3.4, 5.5)	(4.6, 7.1)	(3.9, 6.5)	(12.4, 13.9)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)	(0.2, 0.2)
34	Acute hepatitis C	12.1	14.6	9.9	7.3	0	0	0	0
		(6.5, 19.3)	(10.5, 20.8)	(6, 13.9)	(6.4, 8.5)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
35	Idiopathic epilepsy	4.9	11.4	6.4	8.3	0.02	0.05	0.03	0.05
		(2.8, 8.3)	(9.5, 13.3)	(4.9, 7.8)	(7.6, 8.8)	(0, 0)	(0, 0.1)	(0, 0)	(0, 0)
36	Other mental disorders	2.7	4.7	5.4	9.7	0.06	0.08	0.14	0.15
		(1.6, 4.8)	(3.6, 6.2)	(4.1, 6.5)	(9, 10.2)	(0, 0.1)	(0.1, 0.1)	(0.1, 0.2)	(0.1, 0.2)
37	Other cardiovascular and circulatory diseases	5	7.5	3.7	8.2	0.05	0.06	0.04	0.05
		(2, 9.1)	(5.1, 10.4)	(2, 7.6)	(7.6, 8.7)	(0, 0.1)	(0, 0.1)	(0, 0.1)	(0, 0.1)
38	Family planning	2.5	3.4	3.5	10.2	0.04	0.04	0.05	0.15
		(1.8, 3.9)	(2.6, 4.4)	(2.7, 4.4)	(9.2, 11.3)	(0, 0)	(0, 0.1)	(0, 0.1)	(0.1, 0.2)
39	Treatment of obesity	5.7	5.8	4	7.9	0.05	0.04	0.03	0.05
		(3.5, 8.6)	(4.5, 7.1)	(2.6, 5.8)	(7.4, 8.3)	(0, 0.1)	(0, 0.1)	(0, 0)	(0, 0)
40	Non-Hodgkin lymphoma	6.7	10.9	8.8	5.5	0	0	0	0
		(4.3, 10.3)	(8.5, 13)	(5.9, 12.5)	(4.9, 6.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
41	Kidney cancer	7.5	5.8	10.1	5.8	0	0	0	0
		(3.9, 14.2)	(3.9, 8.4)	(7.3, 13.7)	(5.2, 6.3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
42	Blindness and vision loss	6.4	11	5.5	5.7	0.07	0.14	0.07	0.06
		(5.1, 8.3)	(9.1, 12.8)	(4.6, 6.4)	(5.4, 5.9)	(0.1, 0.1)	(0.1, 0.2)	(0.1, 0.1)	(0.1, 0.1)
43	Schizophrenia	3.7	12.8	4.4	6.6	0.01	0.03	0.01	0.02
		(3.2, 4.3)	(11.7, 14)	(4, 4.8)	(6.1, 7)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
44	Headache disorders	1.4	2.7	3.4	8.6	0.04	0.07	0.07	0.14
		(0.9, 2.3)	(2.3, 3.1)	(2.6, 4.1)	(8.2, 9.1)	(0, 0)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)

45	Lower respiratory infections	2.6	4.1	5.3	6.2	0.07	0.07	0.11	0.11
		(2.1, 3.4)	(3.5, 4.8)	(4.7, 6.1)	(5.9, 6.4)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)
46	Opioid use disorders	2.5	5.4	4.4	7.5	0.02	0.03	0.02	0.04
		(1.3, 4.1)	(4.3, 7.2)	(3.2, 5.5)	(6.9, 8)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
47	Upper respiratory infections	2.7	3.4	3.7	5.7	0.12	0.13	0.16	0.2
		(2, 3.4)	(2.9, 3.9)	(3.1, 4.4)	(5.4, 6)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.2)	(0.2, 0.2)
48	Heart Failure	4.3	5.9	5.2	4.4	0.07	0.12	0.1	0.08
		(2.1, 8)	(3.8, 9)	(3, 8.9)	(3.9, 4.8)	(0, 0.1)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)
49	Low back pain	3.3	4.1	3.1	5.3	0.03	0.05	0.04	0.05
		(1.8, 5)	(3.4, 5)	(2.3, 4.2)	(4.8, 5.7)	(0, 0)	(0, 0.1)	(0, 0.1)	(0.1, 0.1)
50	Osteoarthritis	4.4	8.4	5.3	7.4	0.09	0.24	0.13	0.14
		(2.9, 6.2)	(6.9, 9.9)	(4.1, 6.5)	(6.7, 8.3)	(0.1, 0.1)	(0.2, 0.3)	(0.1, 0.2)	(0.1, 0.1)
51	Interstitial lung disease and pulmonary sarcoidosis	4.8	4.8	6.9	3.8	0	0	0	0
		(1.6, 11.3)	(2.6, 7.1)	(3.1, 12.2)	(3, 4.4)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
52	Stroke	5.4	4.5	4.8	4	0.04	0.05	0.05	0.03
		(2.4, 10.2)	(3.4, 5.8)	(3, 7.2)	(3.7, 4.4)	(0, 0.1)	(0, 0.1)	(0, 0.1)	(0, 0)
53	Nutritional deficiencies	3.9	5.8	4.2	3.7	0.05	0.06	0.05	0.04
		(3.2, 4.9)	(4.6, 6.8)	(3.3, 5.1)	(3.4, 3.9)	(0, 0.1)	(0, 0.1)	(0, 0.1)	(0, 0)
54	Hemoglobinopathies and hemolytic anemias	3.1	5.4	3.3	3	0.01	0.01	0.01	0.01
		(2.1, 4.8)	(4.6, 6.3)	(2.3, 4.7)	(2.8, 3.3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
55	Pancreatitis	7.6	3.1	4.1	3	0.01	0	0	0
		(4.8, 13.4)	(1.9, 5.7)	(2.4, 6.9)	(2.4, 3.5)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
56	Chronic kidney disease	4.4	6.2	5.1	2	0.04	0.05	0.04	0.02
		(2.8, 6.8)	(5, 7.6)	(3.9, 6.7)	(1.9, 2.3)	(0, 0.1)	(0, 0.1)	(0, 0.1)	(0, 0)
57	Colon and rectum cancer	6	4.9	3.8	3.2	0.01	0.01	0.01	0.01
		(4.1, 8.7)	(3.7, 6.3)	(2.7, 5.4)	(2.9, 3.6)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
58	Gout	7.9	5.7	1.8	3.1	0.07	0.06	0.02	0.04
		(4.9, 11.6)	(3.9, 7.9)	(1.3, 3)	(2.7, 3.5)	(0.1, 0.1)	(0, 0.1)	(0, 0)	(0, 0)
59	Congenital birth defects	2.2	2.1	2.1	3.1	0.01	0.01	0.02	0.01
		(1.4, 4)	(1.6, 2.7)	(1.7, 2.6)	(2.8, 3.4)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
60	Alzheimer's disease and other dementias	7.8	6	14	4.7	0.08	0.12	0.18	0.08
		(3.3, 16.5)	(4.2, 8.5)	(7.3, 21.6)	(3.7, 5.9)	(0, 0.1)	(0.1, 0.2)	(0.1, 0.3)	(0.1, 0.1)
61	Tobacco intervention	1.6	2.6	2.2	3.1	0.01	0.01	0.01	0.02
		(1, 2.2)	(2.1, 3.1)	(1.6, 2.8)	(2.9, 3.3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)

62	Cirrhosis and other chronic liver diseases	4 (2.5, 6.6)	2.8 (2, 3.4)	4.4 (3.4, 6.3)	2.3 (2.1, 2.5)	0.03 (0, 0)	0.03 (0, 0)	0.03 (0, 0)	0.02 (0, 0)
63	Liver cancer	7.1 (4, 10.5)	2.8 (1.9, 3.7)	5 (3.8, 6.6)	2 (1.8, 2.2)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
64	Non-rheumatic valvular heart disease	5.6 (2.5, 11.1)	4.2 (2.7, 6.7)	3.9 (2, 6.7)	2.7 (2.2, 3.2)	0.04 (0, 0.1)	0.04 (0, 0.1)	0.03 (0, 0.1)	0.02 (0, 0)
65	Ovarian cancer	5.1 (1.9, 11.4)	3.3 (1.6, 6.3)	4.9 (2.5, 8.9)	2.4 (1.6, 3)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
66	Acute renal failure	2.4 (1.9, 2.9)	4.4 (3.7, 5.1)	3.5 (2.8, 4.1)	1.7 (1.5, 1.8)	0.01 (0, 0)	0.02 (0, 0)	0.01 (0, 0)	0.01 (0, 0)
67	Parkinson's disease	3.4 (1.6, 6.6)	1.1 (0.6, 1.6)	1.7 (1, 2.6)	3.2 (3, 3.4)	0.02 (0, 0)	0.01 (0, 0)	0.02 (0, 0)	0.02 (0, 0)
68	Otitis media	1.4 (0.9, 2)	1.3 (1.1, 1.6)	1.7 (1.4, 2)	2.5 (2.3, 2.6)	0.05 (0, 0.1)	0.04 (0, 0)	0.06 (0.1, 0.1)	0.08 (0.1, 0.1)
69	Other unintentional injuries	0.9 (0.7, 1.2)	1.4 (1.2, 1.5)	1.5 (1.3, 1.6)	2.3 (2.2, 2.4)	0.01 (0, 0)	0.01 (0, 0)	0.02 (0, 0)	0.02 (0, 0)
70	Autism spectrum disorders	1.4 (0.3, 4.1)	1.3 (0.9, 2)	1.1 (0.7, 1.9)	2.5 (2, 2.9)	0.01 (0, 0)	0.01 (0, 0)	0.01 (0, 0)	0.01 (0, 0)
71	Neck pain	1.2 (0.7, 1.7)	1.5 (1.1, 1.8)	1.6 (1.2, 2.3)	2.4 (2.2, 2.6)	0.01 (0, 0)	0.01 (0, 0)	0.01 (0, 0)	0.01 (0, 0)
72	Lower extremity peripheral arterial disease	4.1 (2.4, 6.9)	5.6 (3.5, 8.3)	4.8 (2.7, 6.7)	3.3 (2.8, 3.8)	0.06 (0, 0.1)	0.09 (0.1, 0.1)	0.07 (0, 0.1)	0.04 (0, 0)
73	Malignant neoplasm of bone and articular cartilage	3.6 (2.3, 5.6)	1.4 (0.8, 2)	1.6 (0.9, 2.3)	1.3 (1.1, 1.4)	0.01 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
74	Cardiomyopathy and myocarditis	1.5 (0.5, 4.2)	1.9 (1.2, 2.9)	1.7 (0.9, 2.9)	1.2 (1, 1.4)	0.02 (0, 0)	0.02 (0, 0)	0.02 (0, 0)	0.01 (0, 0)
75	Non-melanoma skin cancer	1.9 (0.8, 4.4)	0.8 (0.5, 1.4)	1.1 (0.6, 2.1)	2 (1.8, 2.1)	0.01 (0, 0)	0.01 (0, 0)	0.01 (0, 0)	0.01 (0, 0)
76	Malignant skin melanoma	0.9 (0.5, 1.4)	1 (0.6, 1.5)	0.9 (0.6, 1.3)	1.8 (1.6, 2)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
77	Gallbladder and biliary diseases	1.2 (0.7, 2)	0.8 (0.5, 1.2)	1.5 (0.7, 2.2)	1.2 (1.1, 1.3)	0 (0, 0)	0 (0, 0)	0.01 (0, 0)	0 (0, 0)
78	Pregnancy and postpartum care	1.2 (0.8, 2.1)	2.1 (1.5, 2.7)	1.5 (1.1, 2.4)	2.1 (1.8, 2.5)	0.01 (0, 0)	0.01 (0, 0)	0.01 (0, 0)	0.01 (0, 0)

79	Enteric infections	0.9	1	1	1.1	0.01	0.01	0.01	0.01
		(0.5, 1.4)	(0.7, 1.3)	(0.8, 1.3)	(1, 1.2)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
80	Hypertensive heart disease	1.9	1.2	1.6	1	0.02	0.02	0.02	0.01
		(1, 3.6)	(0.7, 1.9)	(1, 2.6)	(0.9, 1.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
81	Brain and central nervous system cancer	0.4	0.5	0.4	0.9	0	0	0	0
		(0.2, 1)	(0.3, 0.7)	(0.3, 0.6)	(0.9, 1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
82	Thyroid cancer	0.7	0.6	0.8	0.8	0.01	0.01	0.01	0.01
		(0.3, 1.5)	(0.4, 0.9)	(0.4, 1.1)	(0.7, 0.9)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
83	Sexually transmitted infections excluding HIV	0.8	1.1	0.7	0.5	0.01	0.01	0.01	0.01
		(0.3, 1.3)	(0.7, 1.5)	(0.4, 1)	(0.4, 0.7)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
84	Inguinal, femoral, and abdominal hernia	1.1	0.8	1	0.5	0.01	0.01	0.01	0.01
		(0.4, 2)	(0.3, 1.4)	(0.4, 1.5)	(0.4, 0.6)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
85	Oral disorders	0.6	0.7	0.6	0.6	0.03	0.03	0.03	0.02
		(0.5, 0.8)	(0.6, 0.9)	(0.5, 0.8)	(0.5, 0.6)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
86	Alcohol use disorders	0.2	0.5	0.3	0.7	0.01	0.01	0.01	0.01
		(0.1, 0.3)	(0.4, 0.6)	(0.3, 0.4)	(0.7, 0.7)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
87	Pulmonary Arterial Hypertension	1.1	0.7	0.9	0.5	0.01	0.01	0.01	0
		(0.5, 2.3)	(0.4, 1.1)	(0.4, 1.6)	(0.4, 0.6)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
88	Soft tissue and other extrasosseous sarcomas	0.2	0.3	0.3	0.6	0	0	0	0
		(0.1, 0.3)	(0.2, 0.4)	(0.2, 0.4)	(0.6, 0.7)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
89	Other malignant neoplasms	0.2	0.3	0.3	0.6	0	0	0	0
		(0.1, 0.2)	(0.2, 0.3)	(0.2, 0.3)	(0.6, 0.7)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
90	Aortic aneurysm	1.6	0.8	1	0.4	0.01	0.01	0.01	0
		(0.6, 2.9)	(0.3, 1.6)	(0.6, 1.5)	(0.4, 0.6)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
91	Pancreatic cancer	0.6	0.5	0.9	0.4	0	0	0	0
		(0.3, 1)	(0.3, 0.9)	(0.5, 1.5)	(0.4, 0.6)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
92	Other direct maternal disorders	0.4	0.8	0.4	0.8	0	0	0	0
		(0.4, 0.5)	(0.7, 1)	(0.3, 0.4)	(0.7, 0.9)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
93	Neglected tropical diseases and malaria	0.2	0.2	0.3	0.4	0	0	0	0
		(0.2, 0.5)	(0.2, 0.4)	(0.2, 0.5)	(0.4, 0.5)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
94	Rheumatic heart disease	0.6	0.6	0.5	0.3	0.01	0.01	0.01	0
		(0.3, 1.1)	(0.3, 0.8)	(0.3, 0.7)	(0.3, 0.3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
95	Other substance use disorders	0.1	0.3	0.2	0.4	0	0	0	0
		(0.1, 0.3)	(0.2, 0.4)	(0.1, 0.4)	(0.3, 0.4)	(0, 0)	(0, 0)	(0, 0)	(0, 0)

96	Septicemia	0.3	0.3	0.3	0.3	0	0	0	0
		(0.1, 0.6)	(0.2, 0.5)	(0.2, 0.4)	(0.3, 0.3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
97	Eating disorders	0.5	0.5	0.2	0.6	0	0	0	0
		(0.2, 1.1)	(0.2, 0.8)	(0.1, 0.4)	(0.5, 0.7)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
98	Age-related and other hearing loss	0.4	0.2	0.3	0.2	0	0	0	0
		(0.2, 0.7)	(0.2, 0.4)	(0.2, 0.5)	(0.2, 0.3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
99	Uterine cancer	0.3	0.4	0.3	0.3	0	0	0	0
		(0.1, 0.6)	(0.2, 0.6)	(0.2, 0.5)	(0.2, 0.4)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
100	Falls	0.1	0.2	0.2	0.2	0	0	0	0
		(0.1, 0.2)	(0.1, 0.3)	(0.1, 0.3)	(0.2, 0.2)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
101	Idiopathic developmental intellectual disability	0.1	0.1	0.1	0.2	0	0	0	0
		(0, 0.3)	(0.1, 0.1)	(0.1, 0.1)	(0.2, 0.3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
102	Cervical cancer	0.1	0.2	0.2	0.2	0	0	0	0
		(0.1, 0.2)	(0.1, 0.3)	(0.1, 0.3)	(0.2, 0.3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
103	Motor neuron disease	0.1	0.1	0.1	0.2	0	0	0	0
		(0.1, 0.2)	(0.1, 0.1)	(0.1, 0.1)	(0.2, 0.2)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
104	Bladder cancer	0.4	0.2	0.2	0.2	0	0	0	0
		(0.2, 0.6)	(0.1, 0.5)	(0.1, 0.4)	(0.2, 0.2)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
105	Paralytic ileus and intestinal obstruction	0.1	0.1	0.1	0.1	0	0	0	0
		(0.1, 0.2)	(0.1, 0.2)	(0.1, 0.2)	(0.1, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
106	Tuberculosis	0.1	0.1	0.1	0.1	0	0	0	0
		(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)	(0.1, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
107	Stomach cancer	0.1	0.1	0	0.1	0	0	0	0
		(0, 0.1)	(0, 0.1)	(0, 0.1)	(0.1, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
108	Endocarditis	0	0	0.1	0.1	0	0	0	0
		(0, 0.1)	(0, 0.1)	(0, 0.1)	(0.1, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
109	Esophageal cancer	0.1	0.1	0.1	0.1	0	0	0	0
		(0, 0.1)	(0.1, 0.1)	(0, 0.1)	(0.1, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
110	Lip and oral cavity cancer	0.1	0.1	0	0.1	0	0	0	0
		(0, 0.1)	(0, 0.1)	(0, 0.1)	(0.1, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
111	Indirect maternal deaths	0.1	0.1	0.1	0.1	0	0	0	0
		(0, 0.1)	(0.1, 0.1)	(0, 0.1)	(0, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
112	Hodgkin lymphoma	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)

113	Other neonatal disorders	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
114	Other pharynx cancer	0	0	0	0	0	0	0	0
		(0, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
115	Maternal abortion and miscarriage	0.1	0	0	0	0	0	0	0
		(0, 0.1)	(0, 0.1)	(0, 0.1)	(0, 0.1)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
116	Transport injuries	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
117	Exposure to mechanical forces	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
118	Vascular intestinal disorders	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
119	Larynx cancer	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
120	Gallbladder and biliary tract cancer	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
121	Testicular cancer	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
122	Neuroblastoma and other peripheral nervous cell tumors	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
123	Maternal hemorrhage	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
124	Meningitis	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
125	Neonatal encephalopathy due to birth asphyxia and trauma	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
126	Appendicitis	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
127	Maternal hypertensive disorders	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
128	Mesothelioma	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
129	Maternal sepsis and other maternal infections	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)

130	Neonatal preterm birth	0.5	0.5	0.6	0.5	0.01	0.01	0.01	0.01
		(0.3, 1)	(0.3, 0.7)	(0.4, 0.7)	(0.4, 0.7)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
131	Pneumoconiosis	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
132	Interpersonal violence	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
133	Eye cancer	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
134	Nasopharynx cancer	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
135	Maternal obstructed labor and uterine rupture	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
136	Self-harm	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
137	Late maternal deaths	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
138	Donor	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
139	Hemolytic disease and other neonatal jaundice	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
140	Ectopic pregnancy	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
141	Neonatal sepsis and other neonatal infections	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
142	Acute glomerulonephritis	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
143	Maternal deaths aggravated by HIV/AIDS	0	0	0	0	0	0	0	0
		(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)

B. Table B3.2: Pharmaceutical spending and utilization per prevalent case by race and ethnicity (age-standardized), for 52 health conditions

Here we present spending and utilization per prevalent case by race/ethnicity, for each of 52 health conditions with available prevalence/incidence estimates. Health conditions are ranked by total retail pharmaceutical spending. All groups other than Hispanic are non-Hispanic ethnicity. The mean values across 100 draws are presented in the top row, with the 95% Uncertainty Interval below.

Rank (Total drug spending)	Health Condition	Spending per prevalent case (US\$)				Prescription fills per prevalent case (n)			
		Asian and Pacific Islander	Black	Hispanic	White	Asian and Pacific Islander	Black	Hispanic	White
1	Diabetes mellitus type 2	1005 (904.3, 1126.6)	553.7 (507.5, 612.7)	918.8 (846.7, 983)	1187.7 (1120.6, 1241.9)	7.2 (6.6, 7.6)	3.3 (3, 3.6)	5.2 (4.8, 5.6)	4.6 (4.5, 4.8)
2	HIV/AIDS	18256.4 (6371.1, 51082.5)	6078.6 (4726.3, 8726.6)	13165.4 (7827.9, 20827.4)	10558.5 (4850.7, 15692.7)	9.5 (5.3, 16.1)	4.7 (3.9, 5.4)	7 (5.8, 8.5)	9.4 (8.5, 11)
3	Multiple sclerosis	57755.3 (23107.1, 125175.9)	13719.8 (9023.9, 21152.8)	27099.3 (12376.4, 53288.3)	26049.9 (22494.8, 29521)	13.5 (6.6, 25.9)	2.8 (2, 3.9)	6 (3.9, 10.5)	4.9 (4.5, 5.5)
4	Atrial fibrillation and flutter	1795 (1356, 2303)	860.5 (763.3, 969.7)	1708.9 (1353.3, 2091.1)	1724.8 (1569.7, 1893)	8.6 (6.9, 10.7)	4.1 (3.5, 4.8)	9.3 (8.1, 11.2)	6.6 (6.3, 7)
5	Rheumatoid arthritis	2678.6 (1083.4, 6859.9)	3301.6 (2109.9, 5338.9)	3982.2 (2802.5, 5209.7)	7379.2 (6506.7, 8324.8)	6 (3.2, 14)	9.1 (6.3, 12.9)	7.1 (6, 8.4)	15.8 (14, 17.1)
6	Chronic obstructive pulmonary disease	802.7 (438, 1302)	220.6 (185.4, 261.9)	319.6 (261.3, 374.5)	346.8 (319.3, 371.4)	3.6 (2.6, 4.8)	1.6 (1.4, 1.8)	2.3 (1.9, 2.7)	1.7 (1.5, 1.7)
7	Leukemia	304139.6 (115217.2, 839019.2)	137571.2 (67585.9, 263655)	198000.6 (106546.4, 344080.3)	133171.7 (91141.8, 184144)	84.9 (42.7, 143.8)	65.2 (40.5, 91.5)	46.1 (32.2, 65.8)	58.4 (47.5, 69.2)
8	Multiple myeloma	787512.9	246458.8	516394.8	474993.1	436.8	96.8	195.9	199.4

		(342180.9, 1581820.9)	(170763.1, 368745.4)	(329565.3, 800040.4)	(369643.9, 587760.4)	(155.6, 1049.1)	(46.3, 185.7)	(95.5, 312.9)	(133.4, 272.9)
9	Breast cancer	14643.4	10714.7	15854.5	27538.1	46.6	41.1	56.3	109.5
		(8346.9, 21594)	(7962, 15804.7)	(11509.9, 22453.9)	(24666.7, 30396.8)	(30.6, 76.8)	(30.4, 60)	(43.8, 77.8)	(98.6, 120.5)
10	Ischemic heart disease	796.7	571.4	664.7	1610.1	10.5	6	11.3	13.5
		(536, 1555.9)	(361.3, 920.3)	(493.4, 1013.3)	(1321, 1797)	(7.3, 14.4)	(5.4, 6.8)	(9.6, 13.4)	(12.5, 14.3)
11	Prostate cancer	19921.6	4734.3	12639.6	9947.4	217.6	49	78.5	103.5
		(8379, 48392)	(3139.7, 6411.5)	(7158.3, 19534.8)	(6392.9, 14423.4)	(41.3, 593.7)	(13.9, 93.8)	(26, 172.8)	(57.9, 165.4)
12	Tracheal, bronchus, and lung cancer	91102.7	23709.3	54980.9	35865.5	100.9	33.2	44.7	30.3
		(46217.5, 141585.7)	(14623, 35272)	(35935.3, 89790.5)	(28810, 43596)	(35, 201.2)	(17.8, 51)	(23.8, 79.9)	(22.5, 40.8)
13	Non-Hodgkin lymphoma	13958.2	17045.9	16133	11969.8	14.4	9.6	10.4	9.3
		(9891.6, 20353.4)	(12906.1, 20920.7)	(10915, 20973.7)	(10036.6, 13824.8)	(8.3, 23.4)	(6.7, 12.5)	(7.1, 13.1)	(7.5, 10.8)
14	Kidney cancer	31593.9	14544.6	27703.3	18855.4	21	6.7	12	9.9
		(18312.1, 57865.9)	(9960.7, 20311.3)	(21353.4, 34612.4)	(15959.5, 22093.3)	(11.4, 42.9)	(4.8, 8.8)	(9, 16.1)	(8, 12.2)
15	Lower respiratory infections	19624.8	16129.6	41130.1	52582.1	561.3	246.4	695.1	774
		(16433.5, 23851.6)	(13712.7, 18981.3)	(36811.1, 46363.8)	(49354.7, 55547.7)	(470.5, 682.3)	(225.4, 269.6)	(636.6, 764)	(742.1, 800.7)
16	Opioid use disorders	2801.9	643.7	1002.4	725.4	19	3.5	5.3	3.9
		(1173.3, 5211.7)	(443.6, 915.7)	(582.5, 1648.4)	(670.5, 782.3)	(10.2, 29.7)	(2.5, 5.5)	(3.3, 8.7)	(3.6, 4.3)
17	Stroke	140.8	66.3	118.1	159.2	1.2	0.8	1.2	1.3
		(76.3, 232.5)	(53.5, 79.8)	(84.9, 159.3)	(148.7, 168)	(0.6, 1.8)	(0.6, 0.9)	(1, 1.5)	(1.3, 1.4)
18	Nutritional deficiencies	132.8	74.3	99.7	65.6	1.7	0.7	1	0.7
		(102.6, 172.1)	(58.1, 88.9)	(79, 122.6)	(59.6, 69.4)	(1.3, 2.1)	(0.6, 0.8)	(0.9, 1.2)	(0.6, 0.7)
19	Colon and rectum cancer	7044.8	4159.6	5514.9	6286.2	25.9	19.8	22.4	28.7

		(4465.8, 12362.2)	(2577.5, 6073.6)	(3778.9, 7727.9)	(5452.7, 7226.8)	(16, 41.2)	(13.1, 27.3)	(14.5, 29.5)	(25, 31.9)
20	Liver cancer	46935.5	18790.6	22312.9	35627	52.1	44.9	38.7	51.7
		(20960.2, 77281.6)	(11411.5, 31118.9)	(17019.9, 30471.8)	(29594.2, 41928.7)	(23, 85.7)	(27.7, 66.8)	(25.6, 60.6)	(41.9, 62.6)
21	Ovarian cancer	24880.5	11399.1	18587.4	9018.5	11.6	6.4	7.8	5
		(10093.1, 50052.1)	(5189, 21677.1)	(9398.1, 31801.4)	(5320, 11391.1)	(5.8, 20.1)	(4.4, 9.9)	(4.8, 11.1)	(4, 6)
22	Malignant neoplasm of bone and articular cartilage	438166.4	81535.3	118127.6	110163.6	758	163.4	272.5	175.1
		(292980, 644293.4)	(50560.7, 117322.1)	(74250.1, 168777.5)	(98339, 121327.5)	(562.9, 923)	(113.4, 258.8)	(195.7, 347.8)	(159.1, 190.2)
23	Cardiomyopathy and myocarditis	1206.6	522.6	1171.4	798.8	13.7	5.2	13.8	7.3
		(394.7, 3068.5)	(304.7, 782.9)	(688.5, 1794.5)	(638.9, 933.1)	(5.8, 30.5)	(3.6, 7.4)	(8.7, 18.8)	(6.1, 8.6)
24	Malignant skin melanoma	10425.3	7628.8	5035	4508	22.8	15.7	8.8	3.9
		(5165.4, 17278.7)	(4875.8, 12679.1)	(2887.5, 7188.8)	(4000.9, 4988.9)	(10.2, 39.9)	(11.3, 26.3)	(5.3, 13.5)	(3.3, 4.6)
25	Hypertensive heart disease	700.6	96	374.1	276.3	10.3	1.7	4.8	4.2
		(354.6, 1438)	(57.6, 134.8)	(248.3, 562.4)	(236.8, 308.2)	(6, 15.6)	(1.1, 2.3)	(3.4, 6.3)	(3.8, 4.7)
26	Brain and central nervous system cancer	7664.4	5649.9	6393.5	11509.9	27	20.2	20.7	15.5
		(3803.4, 16961.1)	(3666.6, 8915.6)	(4902.1, 9805.2)	(10726.7, 12452.7)	(17.1, 43)	(14, 27.4)	(15.9, 26.2)	(14, 17.7)
27	Thyroid cancer	8621.1	6544.6	6525.4	11865.8	75.2	58.7	53.4	106.1
		(2678.9, 20121.3)	(4185.2, 11560.9)	(4511.8, 8393.9)	(9938.9, 14207.6)	(29.5, 180.9)	(44, 85.8)	(34, 71.9)	(97.2, 115.7)
28	Inguinal, femoral, and abdominal hernia	5992.7	1323	2080.6	1321.9	90.3	24	33.5	26.4
		(2347.7, 12221.3)	(797.1, 2156.2)	(1181.7, 3042.5)	(1130.7, 1534)	(45.1, 169.6)	(16.9, 36)	(22.3, 44.3)	(23, 29.8)
29	Alcohol use disorders	33.1	28.1	30.2	63.9	1.1	0.8	0.9	1.8
		(23.7, 46.3)	(24.2, 32.5)	(25.5, 34.5)	(55.3, 70.2)	(0.8, 1.4)	(0.7, 0.9)	(0.8, 1.1)	(1.7, 1.8)
30	Soft tissue and other extraosseous sarcomas	6963.7	7167.9	7078.5	14911.1	17.1	9.8	13.9	12.5
		(3212.1, 14233.5)	(4776.4, 11581.2)	(4804.7, 10068.3)	(12780.7, 17199.7)	(10.1, 24.5)	(7, 13.9)	(10.5, 18.5)	(10.8, 14.1)

31	Other malignant neoplasms	2861.1 (1532.7, 4989.2)	4752.2 (2461.9, 6830.3)	2376.5 (1666.7, 3460.6)	5784.2 (5134.2, 6381.2)	7.9 (6.8, 9.4)	7.6 (6.6, 9.3)	8 (7.2, 9.1)	9.6 (9.1, 10.1)
32	Pancreatic cancer	7608.6 (2736.4, 18264.8)	2796.9 (1831.9, 5202)	4560.6 (2486.9, 6234.5)	6102.7 (5082.7, 7156.3)	45.3 (21.2, 72.8)	15.6 (10, 27.1)	25.1 (14.8, 37.1)	23 (18.5, 27.9)
33	Uterine cancer	1421.3 (313.8, 3812.9)	450.2 (192.2, 834.4)	454.5 (225.3, 765.3)	1068.7 (622.1, 1459.6)	7.9 (2.2, 22.4)	3.7 (1.6, 6.8)	3.1 (1.8, 4.9)	4.5 (2.5, 6.9)
34	Falls	3.4 (2.3, 5.2)	3.8 (2.6, 5.8)	3.6 (2.6, 4.7)	3.8 (3.5, 4)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)
35	Cervical cancer	2783.7 (843.8, 7787.1)	990.3 (592.7, 1714.9)	1784.7 (1089, 2799.9)	2138.1 (1645.7, 2536.1)	24.5 (9.5, 56.9)	10.4 (6.4, 17.8)	20.5 (12.9, 32.2)	20.5 (15.3, 24.6)
36	Bladder cancer	1446.7 (708.1, 3083.4)	544.7 (311.6, 905.4)	770.5 (511.6, 1033.3)	586.1 (498, 708.2)	20.2 (11.6, 32.2)	9.8 (6, 13.8)	10.1 (7, 15.8)	7.1 (5.4, 9.1)
37	Tuberculosis	0.7 (0.5, 1)	0.9 (0.7, 1)	1.2 (1, 1.4)	1.6 (1.4, 1.8)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
38	Stomach cancer	1210.6 (194.7, 2697.3)	536.6 (303.5, 961.8)	406.7 (196.6, 640.9)	2318.2 (1868, 2686.1)	9.3 (3.5, 19.5)	6.2 (4, 10.8)	5.7 (3.3, 9.7)	14.7 (11.6, 17.1)
39	Endocarditis	692.5 (391.3, 1294.8)	322.1 (199.5, 450.2)	472 (292.8, 729.6)	752.8 (680.6, 813)	15.2 (9, 30.2)	5.8 (4, 8)	14.9 (10.5, 20.5)	13.3 (11.8, 14.9)
40	Esophageal cancer	1963.6 (732.6, 3699.9)	1810.3 (704.6, 3789.3)	1244.8 (770.3, 1874)	1548.1 (1098.9, 1975.5)	25.1 (13, 44.1)	15.6 (9.3, 24.4)	15.6 (11, 21.2)	13.5 (10.7, 16.4)
41	Lip and oral cavity cancer	672.1 (273, 1313.5)	514.6 (272, 1066.7)	489.3 (287.6, 785.8)	715.9 (598.3, 805.9)	8.9 (4.6, 15.7)	8.8 (5.4, 13.6)	8.4 (6.3, 12.1)	6.7 (5.7, 7.7)
42	Hodgkin lymphoma	1643.1 (603.3, 4087.8)	2328.7 (1478.9, 3268.9)	1598.3 (649.9, 2900.4)	4445.2 (3169.2, 8764.1)	33.4 (14.3, 69.7)	25.7 (13.9, 53.5)	19.7 (10.8, 49.3)	21.8 (16.9, 33.1)

43	Other pharynx cancer	4971.7	1250.8	2212.5	1517.7	48	21.1	31.7	20.2
		(766.7, 9916.8)	(516.9, 2620.1)	(617.6, 4899.2)	(997.8, 2311.4)	(14.8, 96)	(9.9, 38.6)	(13.5, 61.8)	(12, 31.5)
44	Transport injuries	2.7	2.6	2.4	2.2	0.1	0.1	0.2	0.1
		(1.6, 4)	(1.8, 3.3)	(1.7, 3.5)	(2, 2.5)	(0.1, 0.2)	(0.1, 0.1)	(0.1, 0.2)	(0.1, 0.1)
45	Exposure to mechanical forces	1.9	1.2	1.3	0.9	0.1	0	0	0
		(1.1, 3.3)	(0.9, 1.6)	(1, 1.7)	(0.8, 1)	(0, 0.1)	(0, 0)	(0, 0.1)	(0, 0)
46	Vascular intestinal disorders	206.4	82.2	115.2	117.3	7.4	2.6	4.3	2.7
		(113.7, 342.6)	(56.7, 110.8)	(76.8, 163)	(104.1, 139.7)	(5.1, 9.4)	(2.1, 3.1)	(3.3, 5.5)	(2.4, 3)
47	Larynx cancer	1153.4	546.6	456.9	473.6	27	11.3	12.5	10.5
		(516.2, 2174.3)	(246.1, 855.4)	(276.3, 772.5)	(377.4, 635.8)	(13.5, 51.8)	(5, 21.3)	(8.5, 18.2)	(7.9, 13.5)
48	Gallbladder and biliary tract cancer	553.7	536.5	389.9	925.9	10.6	7.4	7.8	11
		(233.4, 1424.4)	(223.1, 1089.2)	(209.2, 682.4)	(661, 1114.6)	(5.2, 19.8)	(5, 12.2)	(4.7, 12.1)	(8.1, 14.3)
49	Mesothelioma	1560.8	1123.4	779.1	906	46.6	55.9	49	29.5
		(578.3, 3489.3)	(675.8, 1736.8)	(400.2, 1376.1)	(707.4, 1236.5)	(21.3, 121.1)	(23.2, 151.9)	(16.1, 129.9)	(12.3, 61.4)
50	Neonatal preterm birth	3.9	1.8	4.6	5.2	0.1	0	0	0.1
		(2.3, 7.4)	(1.2, 2.7)	(2.9, 5.9)	(4.1, 6.7)	(0, 0.1)	(0, 0)	(0, 0.1)	(0, 0.1)
51	Nasopharynx cancer	260	303.7	535.6	828.4	3.9	6.1	9.9	10.1
		(98.9, 797.8)	(157.6, 539.4)	(228.8, 834)	(662.4, 1077.6)	(1.7, 11.5)	(3.5, 9.9)	(6.3, 15.8)	(7.4, 14.4)
52	Self-harm	5.7	5.7	4.8	1.9	0.2	0.4	0.2	0.1
		(2, 12.8)	(2.9, 10.5)	(2.7, 9.2)	(1.3, 2.4)	(0.1, 0.4)	(0.2, 0.7)	(0.1, 0.4)	(0.1, 0.1)

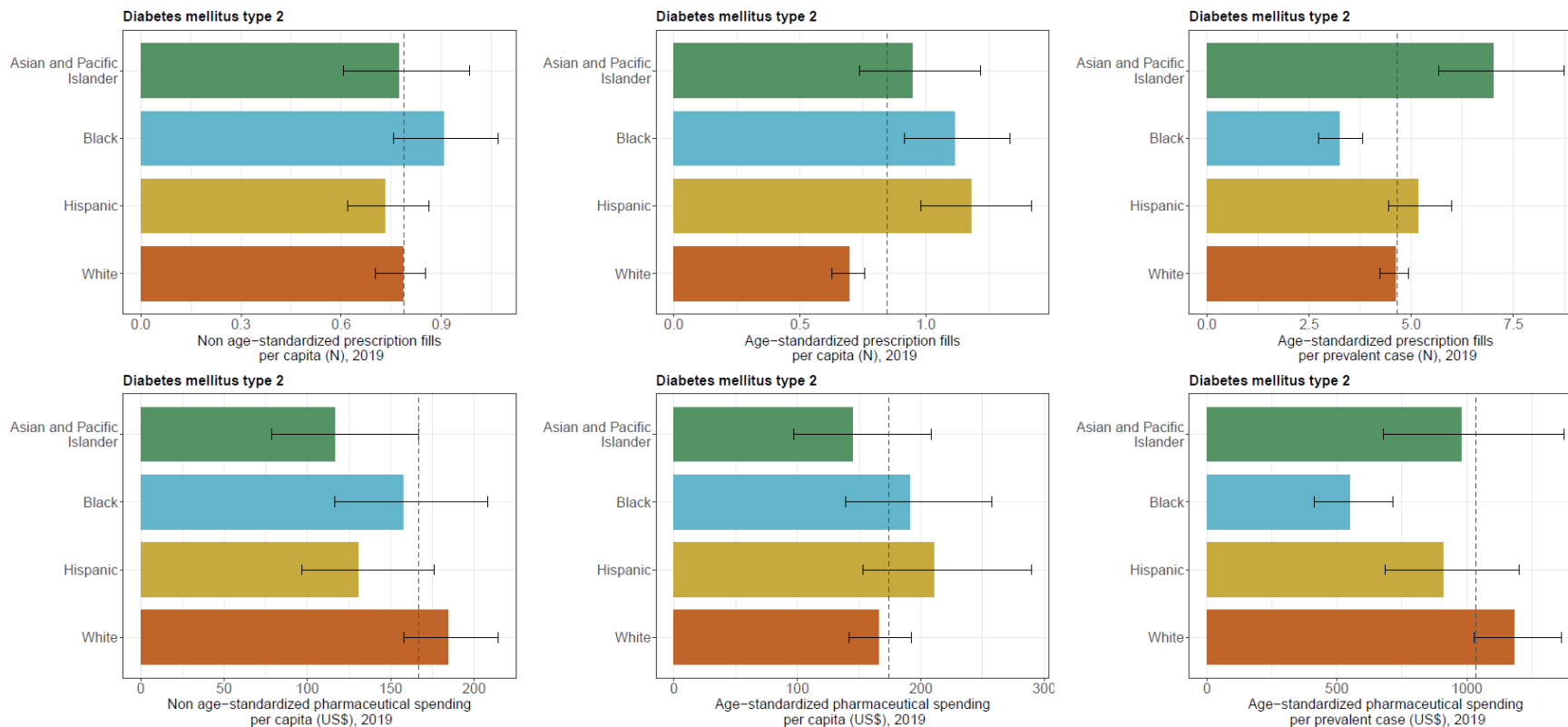
C. Figure B3.1: Age- and prevalence-standardized pharmaceutical spending and utilization by race and ethnicity (for 52 health conditions)

Here we present the full set of figures for all 52 health conditions with available prevalence data. Each figure is organized as follows:

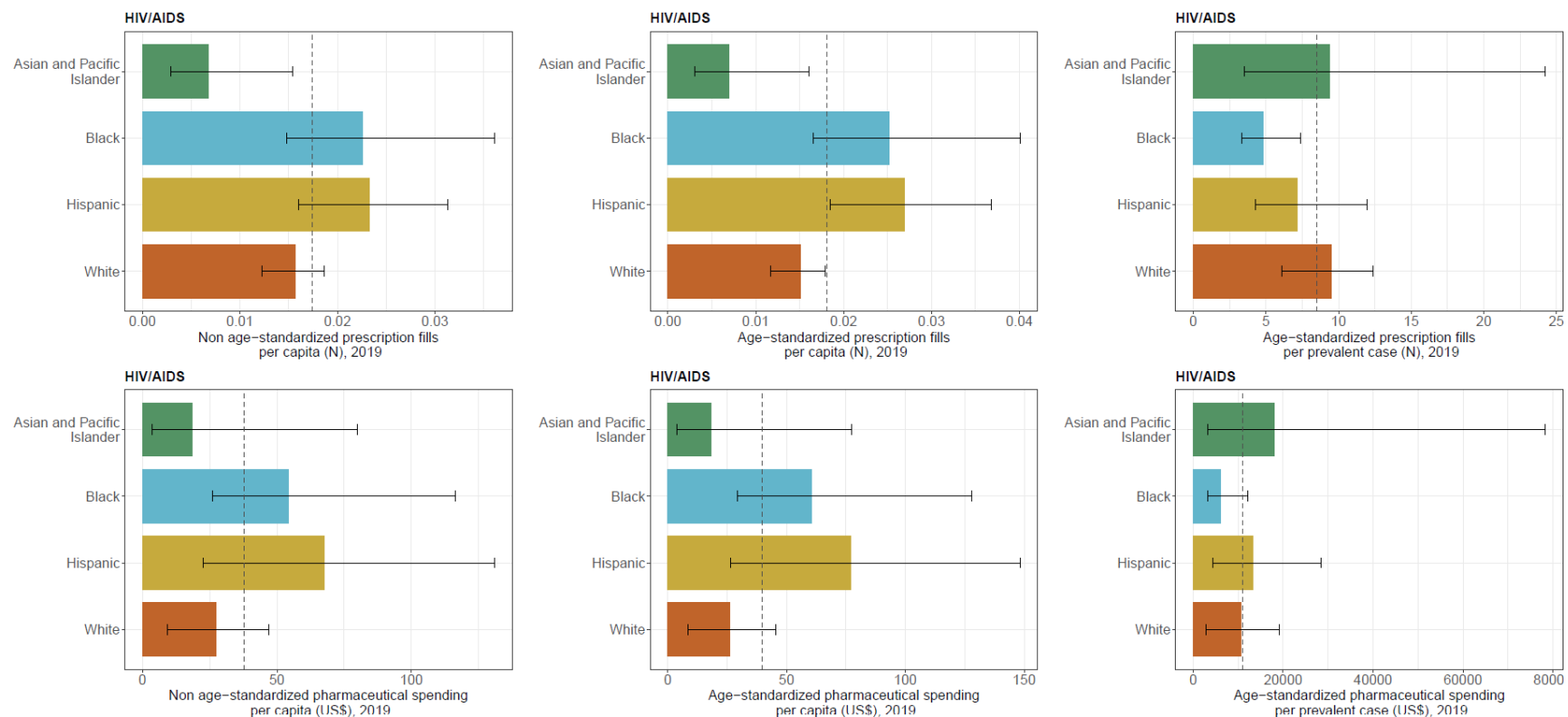
- (I) The top row represents utilization (prescription fills) and (II) the bottom row represents pharmaceutical spending.
- (i) The left column is non-age-standardized per capita estimate, (ii) the middle column is the age-standardized per capita estimate, and (iii) the right column is the age- and prevalence-standardized estimate (i.e. age-standardized fills/spending per prevalent case)

Health conditions are ranked by total retail pharmaceutical spending, as shown in **Table B3.2**.

(1) Diabetes mellitus type 2

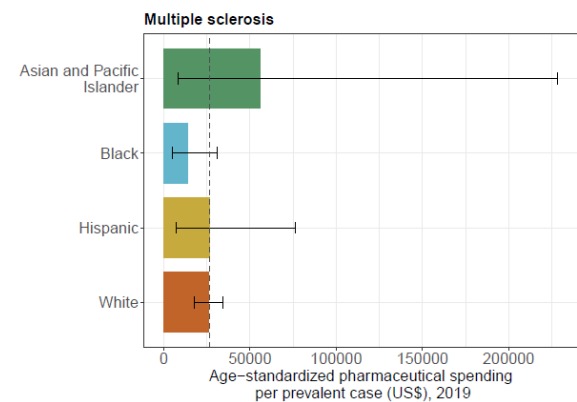
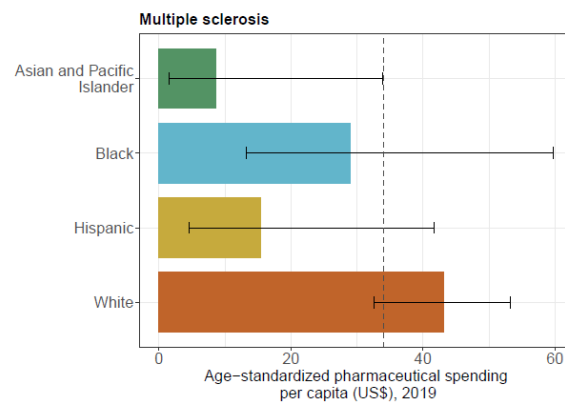
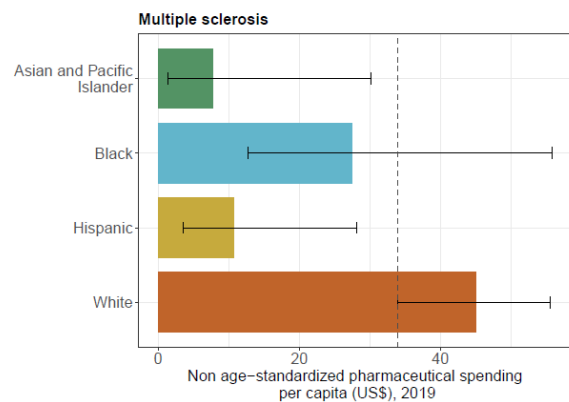
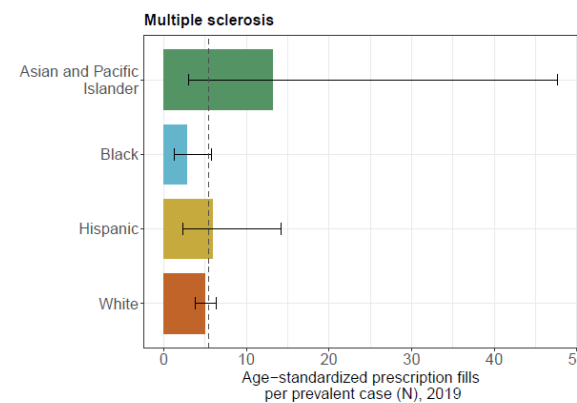
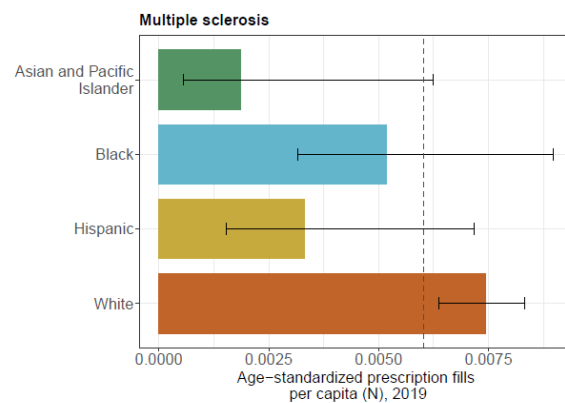
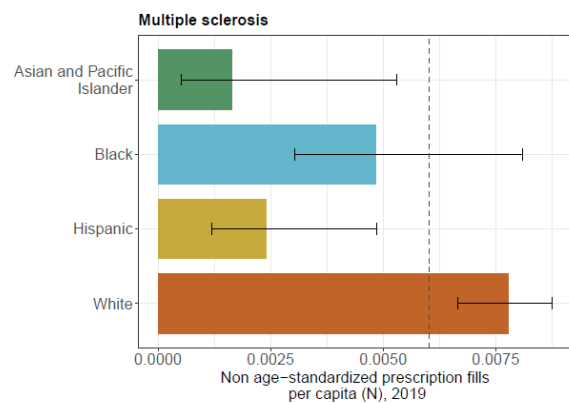


(2) HIV/AIDS

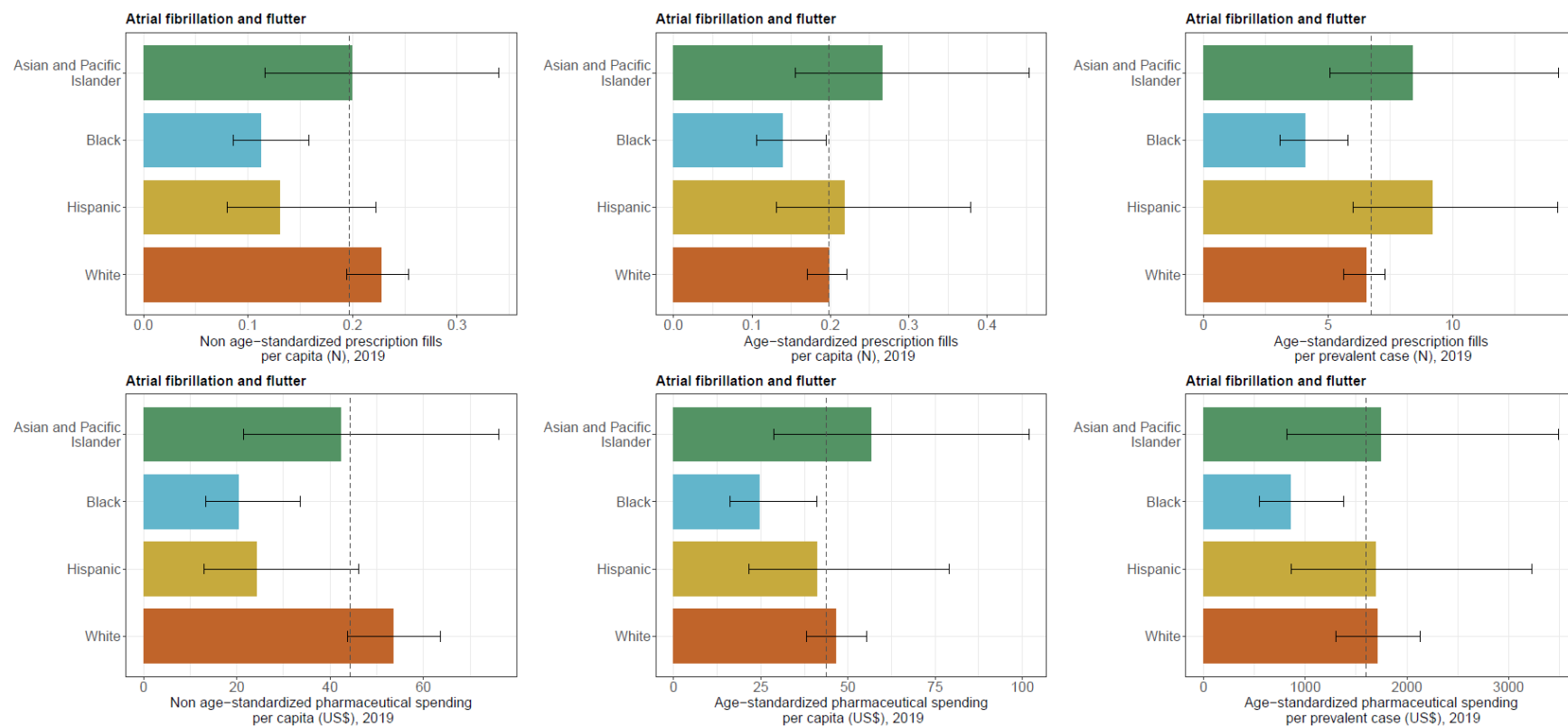


Note: Utilization/spending is inclusive of preventive therapy for HIV (PrEP), so ‘per prevalent case’ estimates should be interpreted with caution.

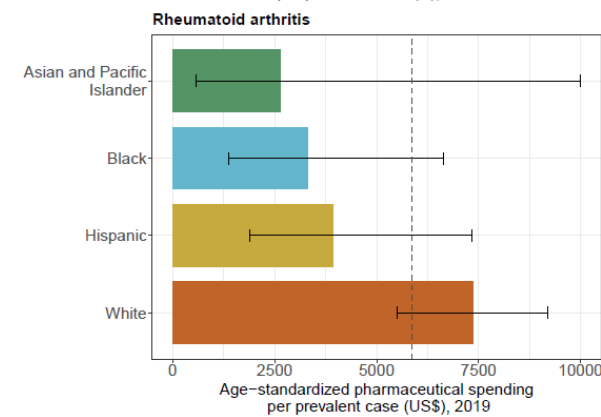
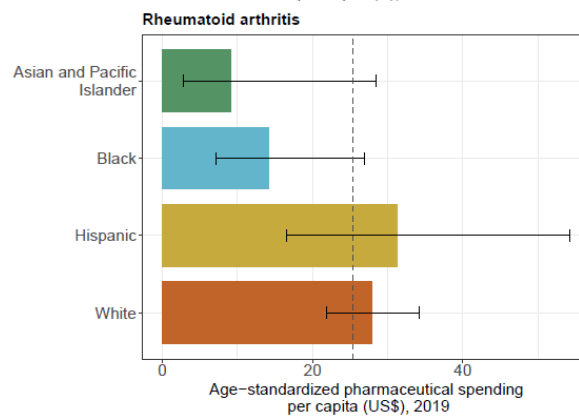
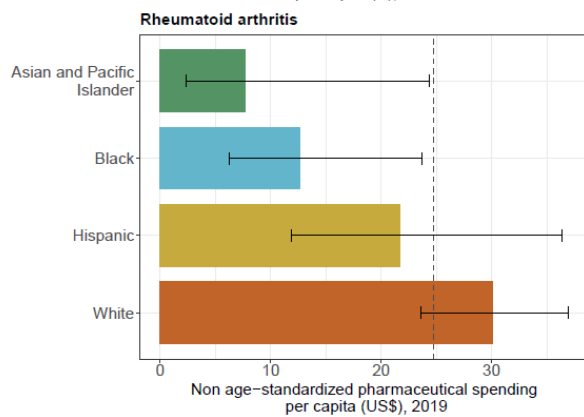
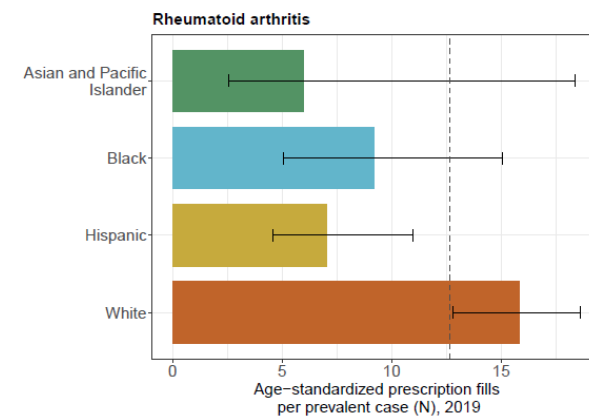
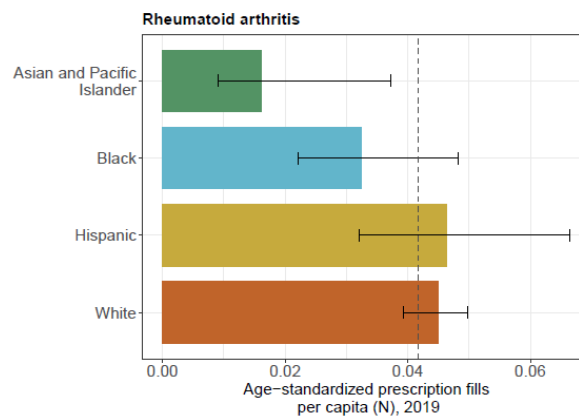
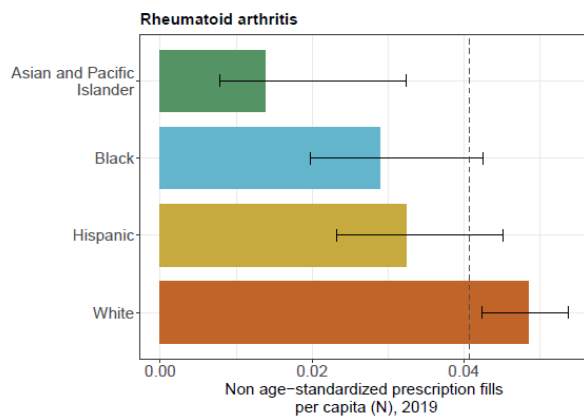
(3) Multiple sclerosis



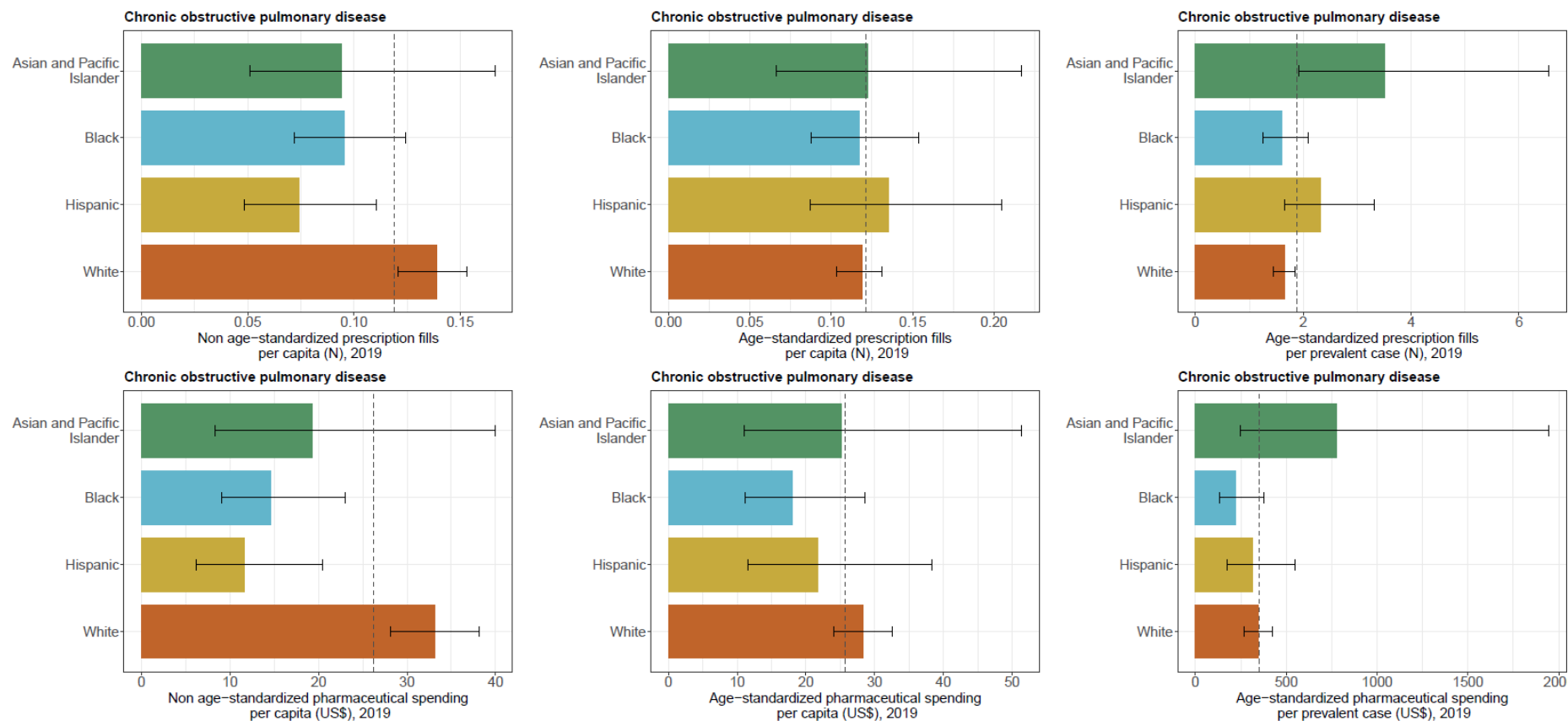
(4) Atrial fibrillation and flutter



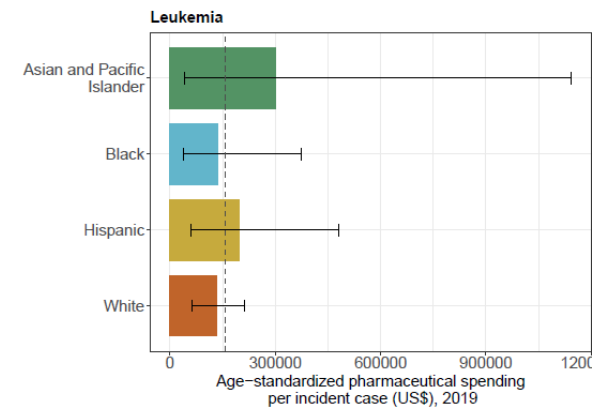
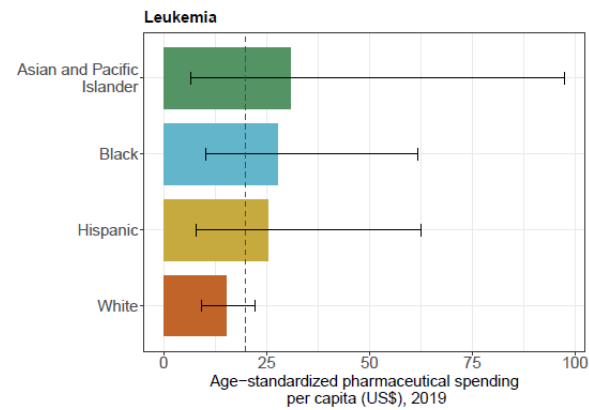
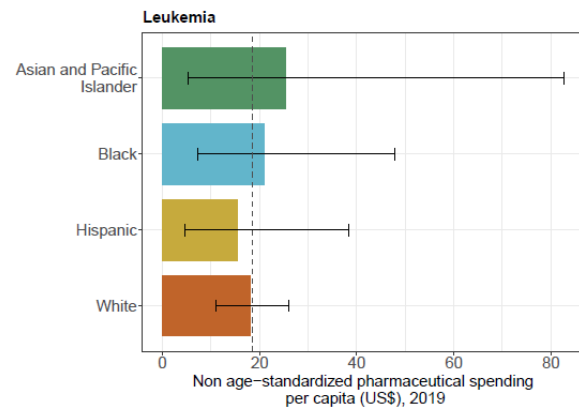
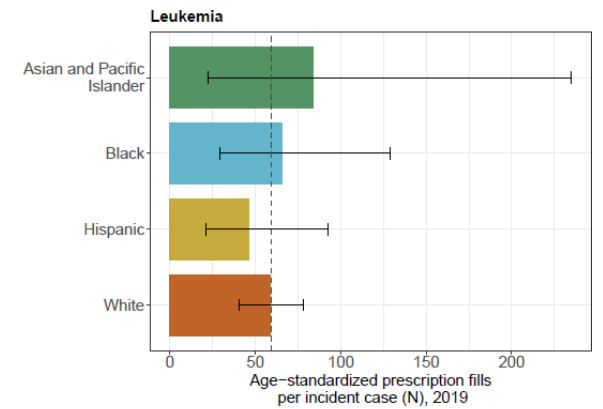
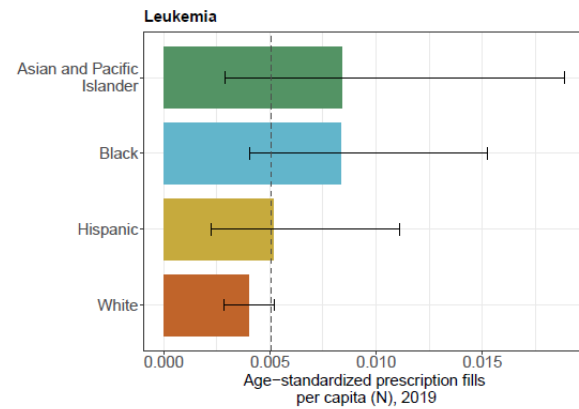
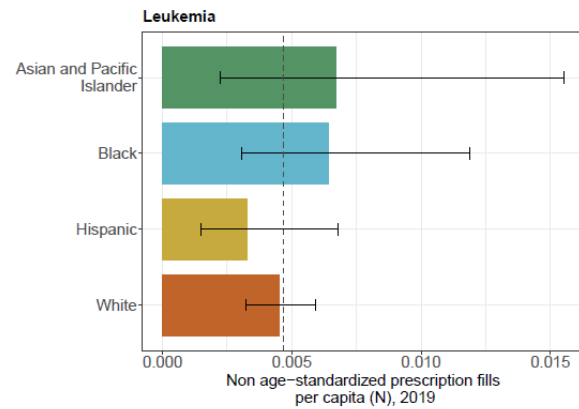
(5) Rheumatoid arthritis



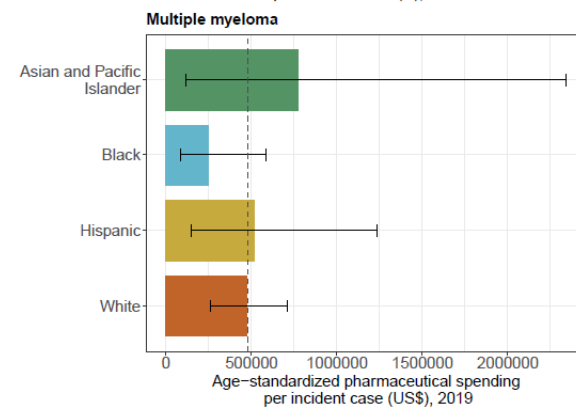
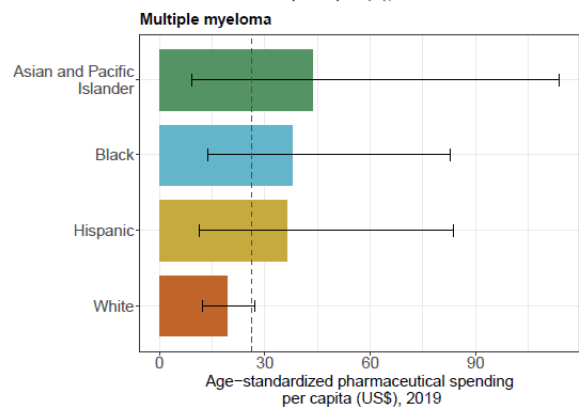
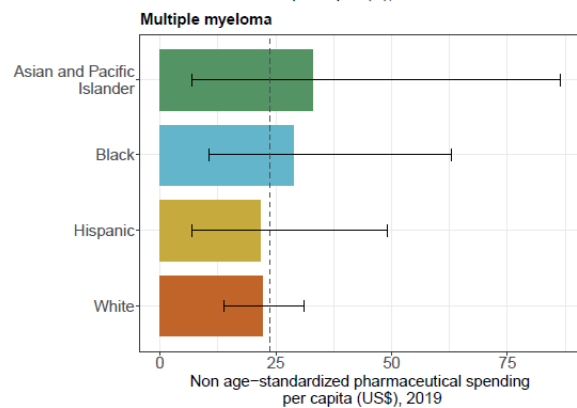
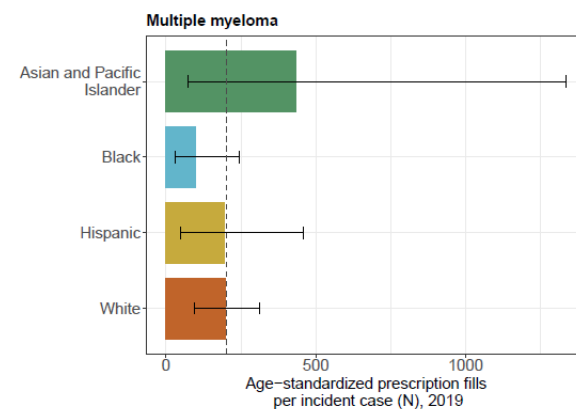
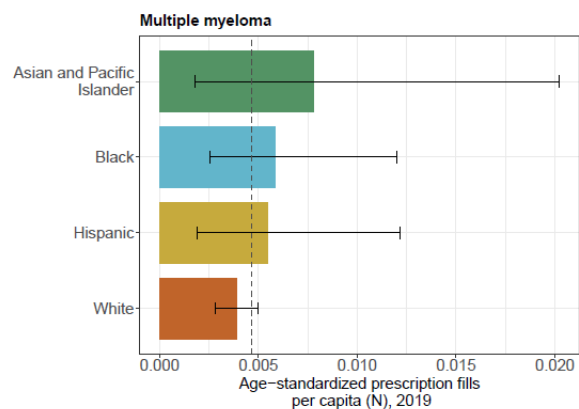
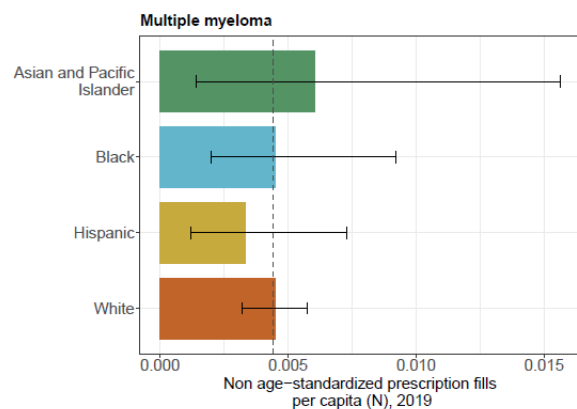
(6) Chronic obstructive pulmonary disease



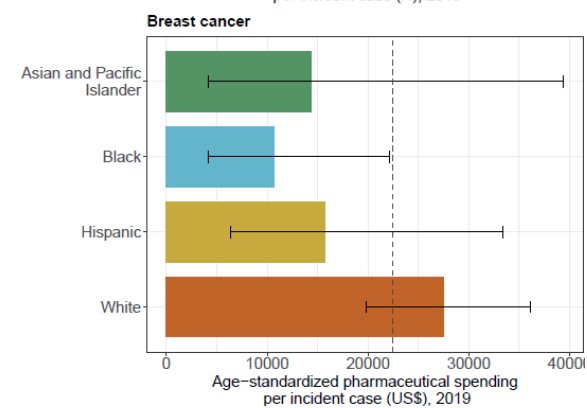
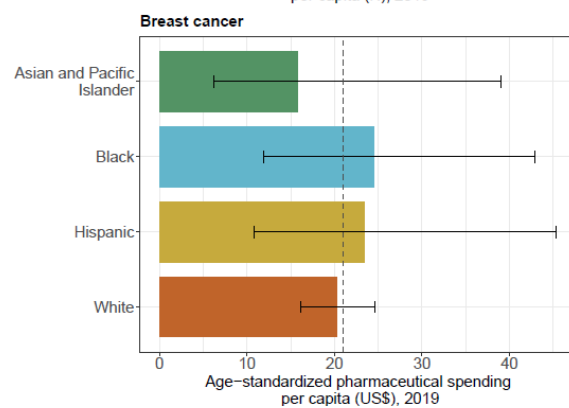
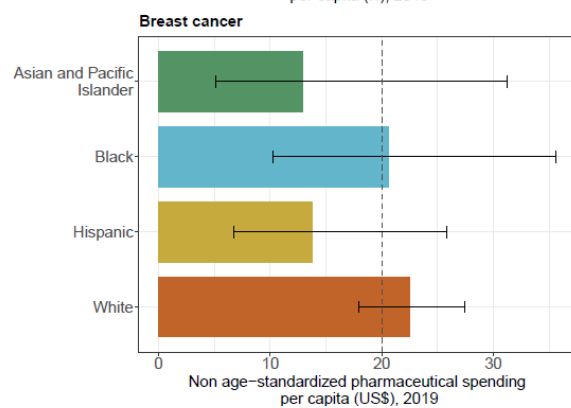
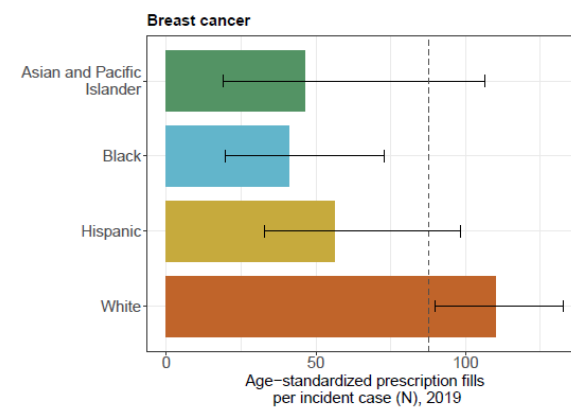
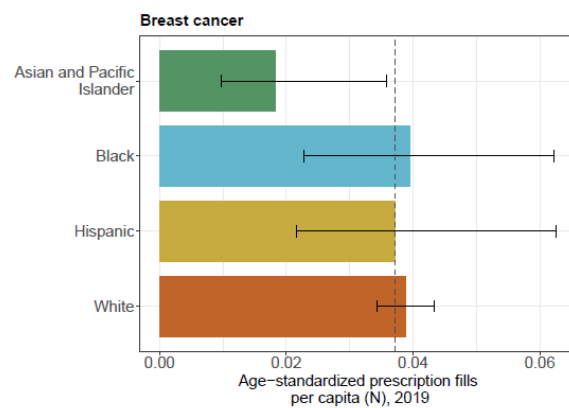
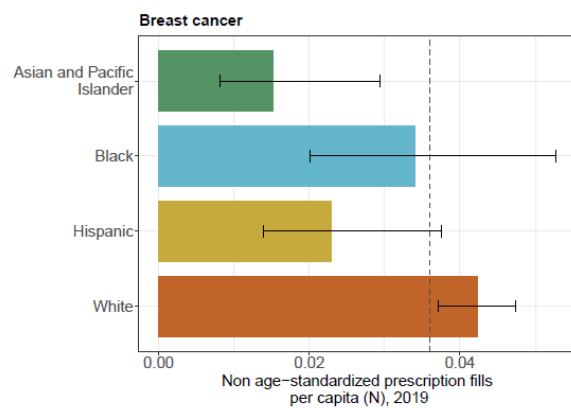
(7) Leukemia



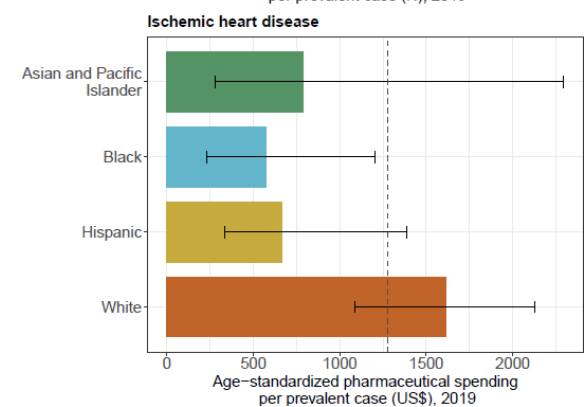
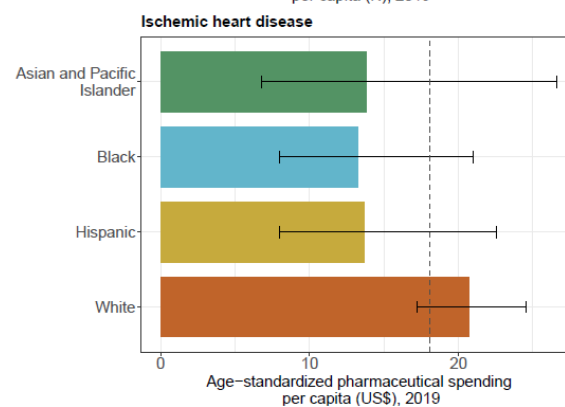
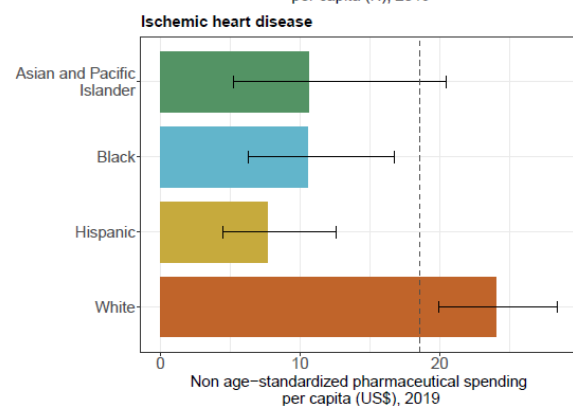
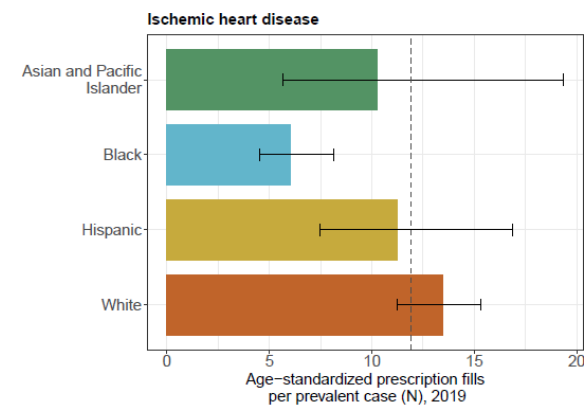
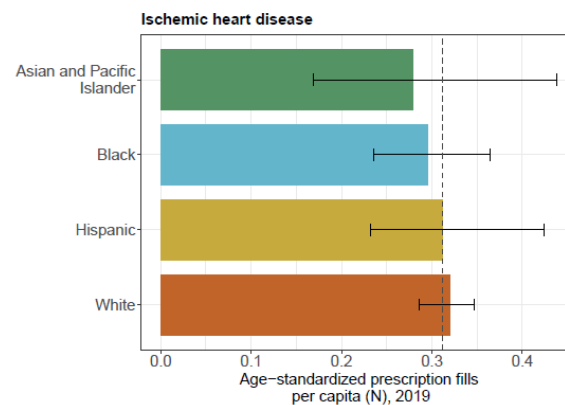
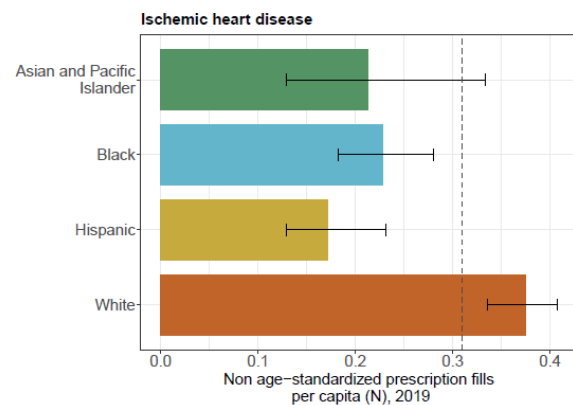
(8) Multiple myeloma



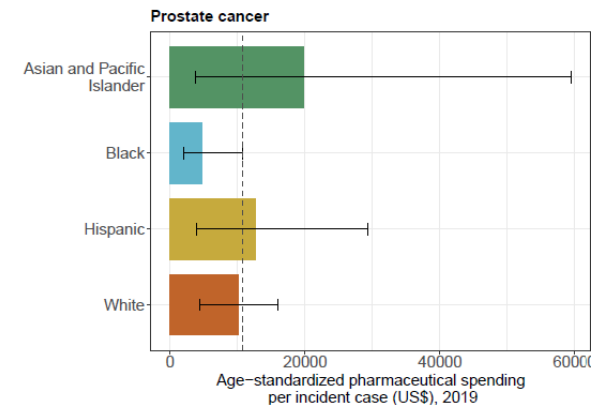
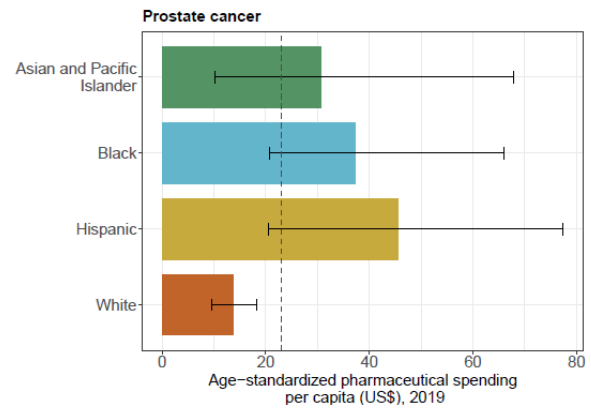
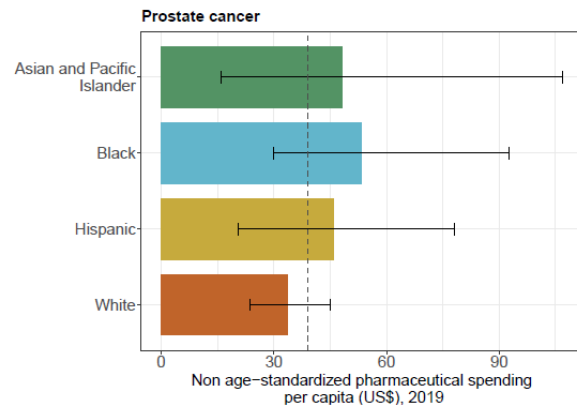
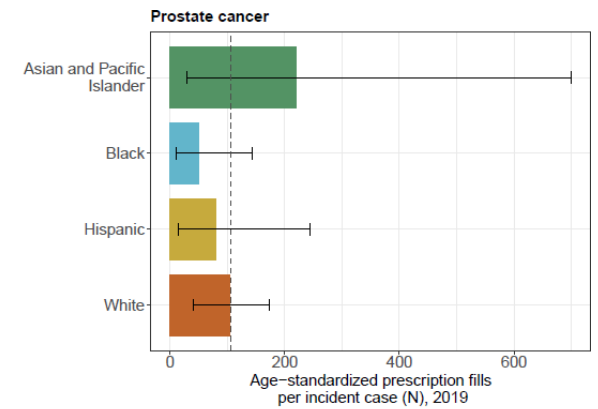
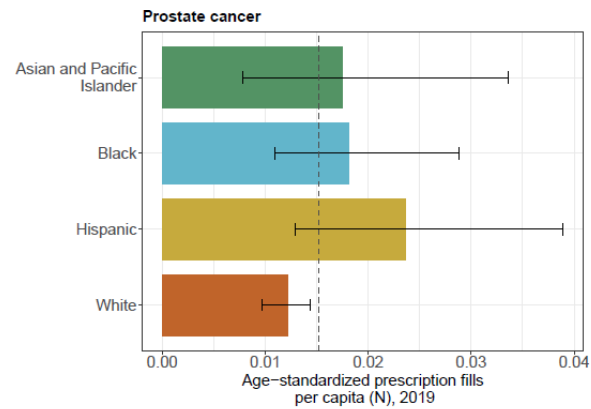
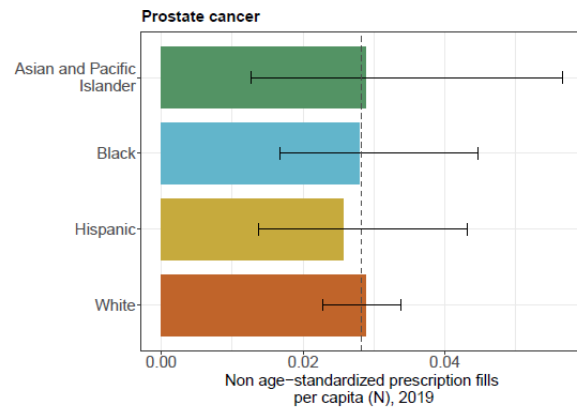
(9) Breast cancer



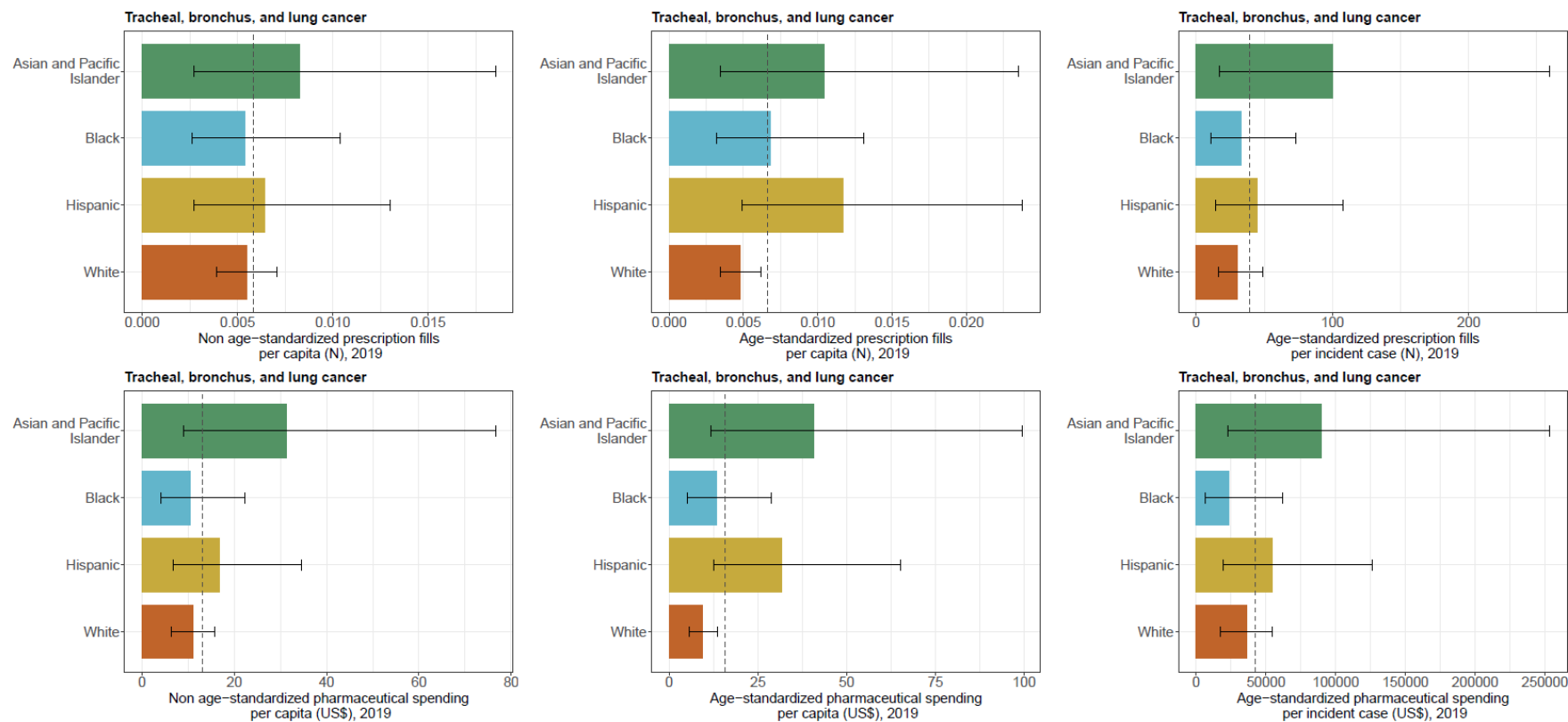
(10) Ischemic heart disease



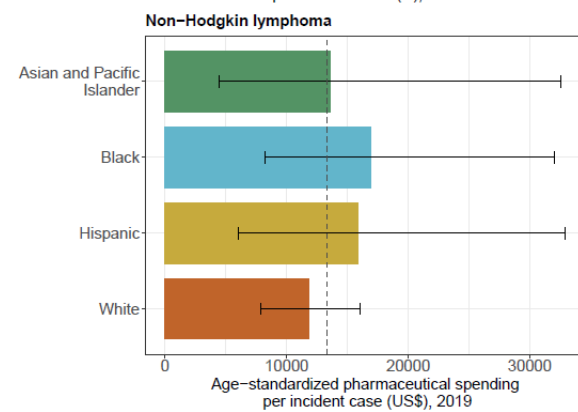
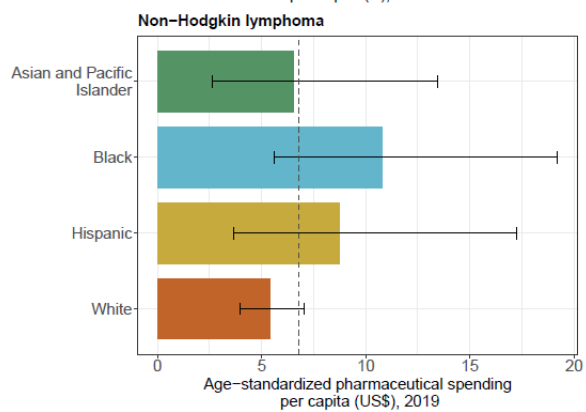
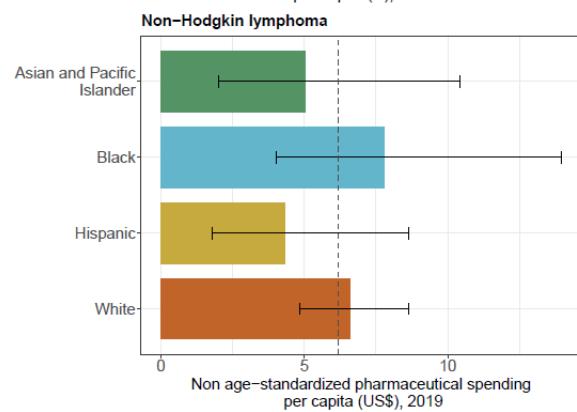
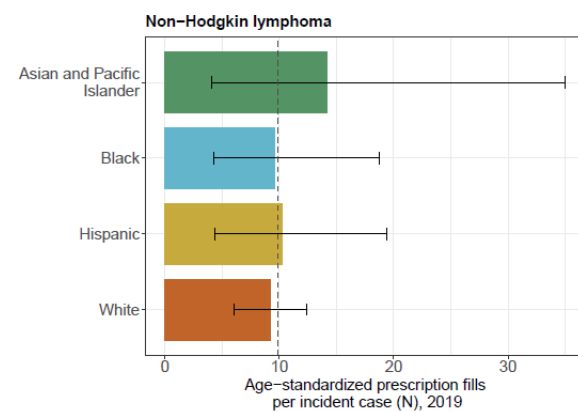
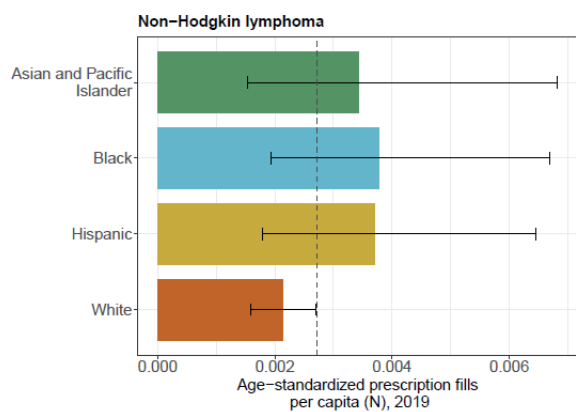
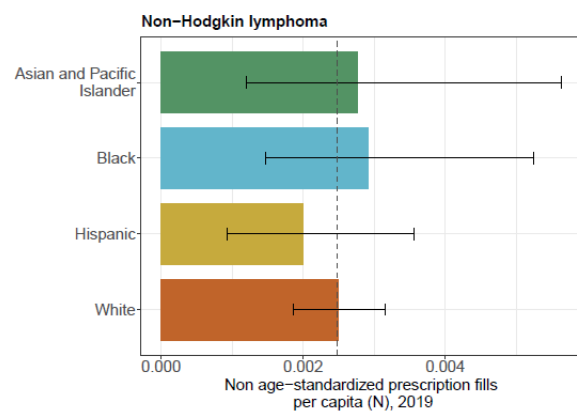
(11) Prostate cancer



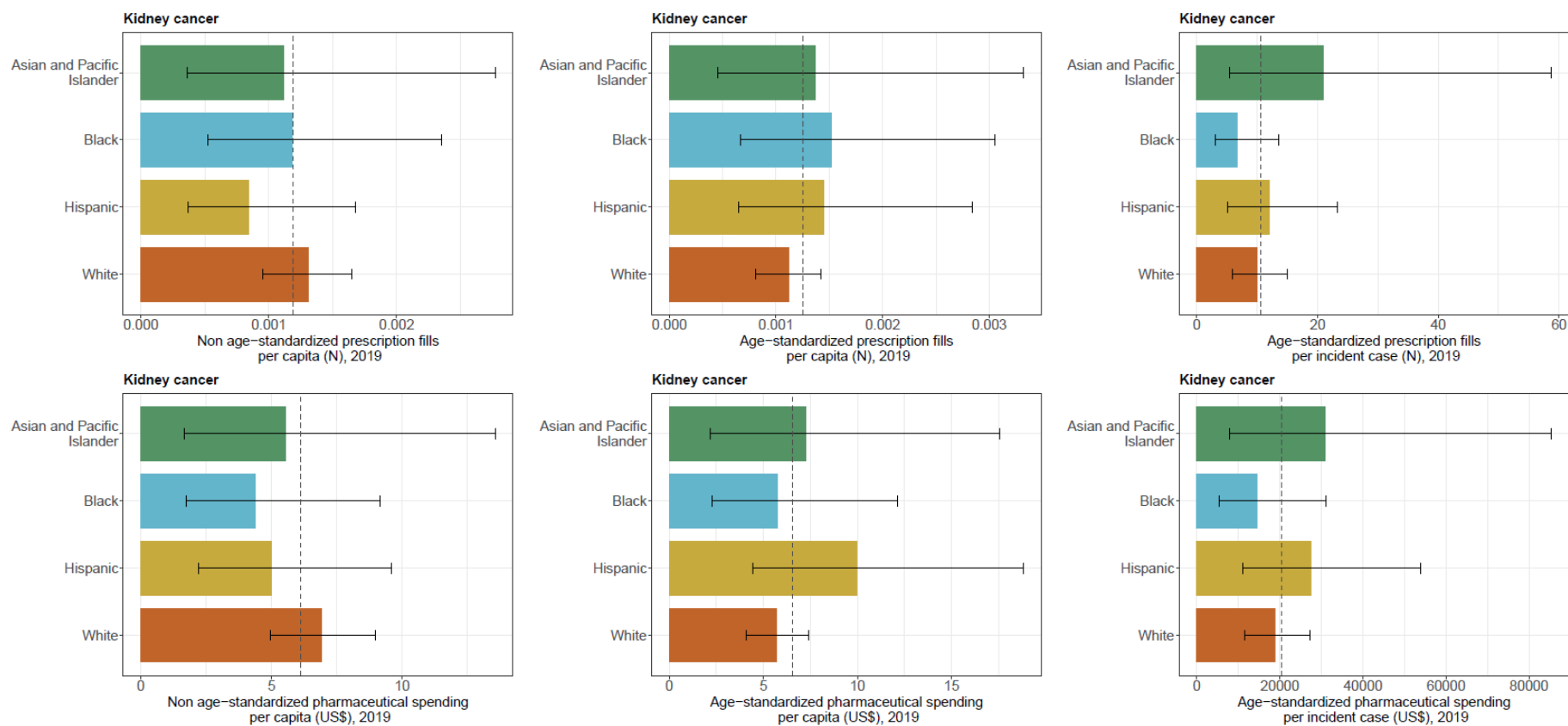
(12) Tracheal, bronchus, and lung cancer



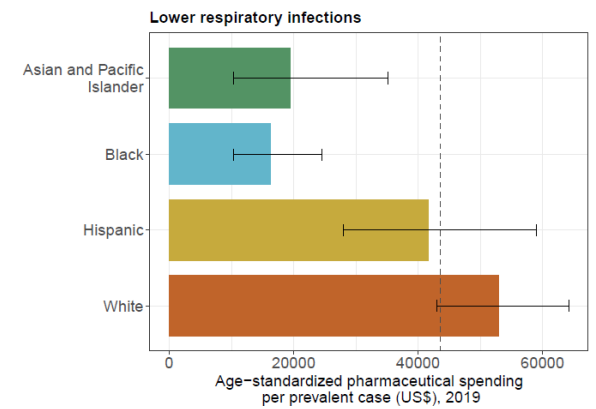
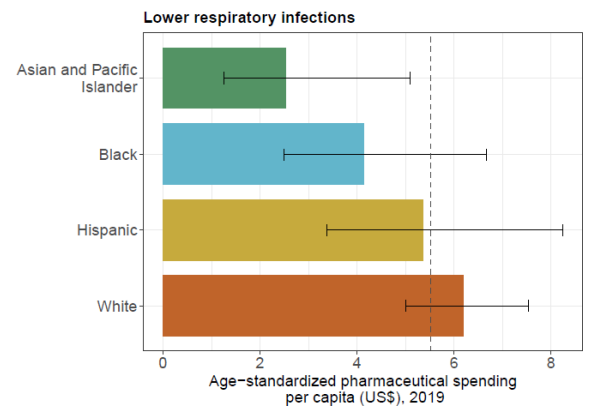
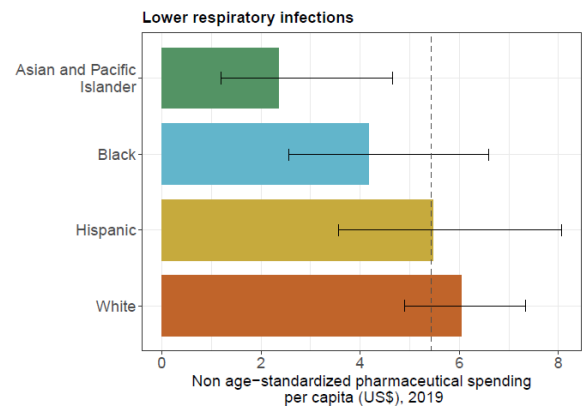
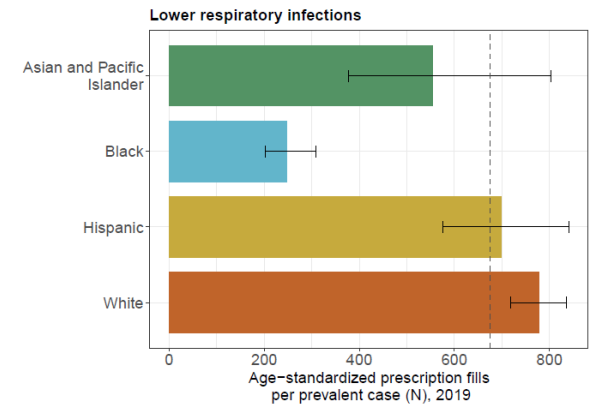
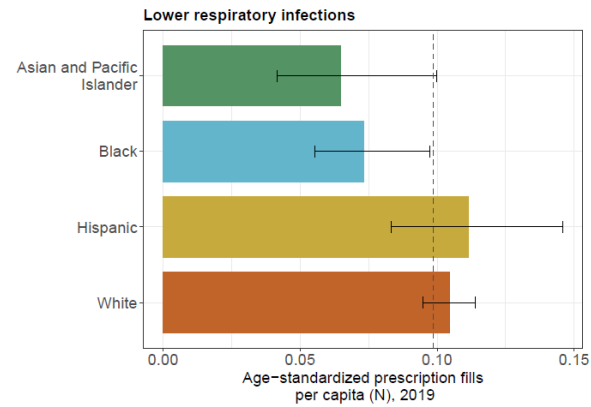
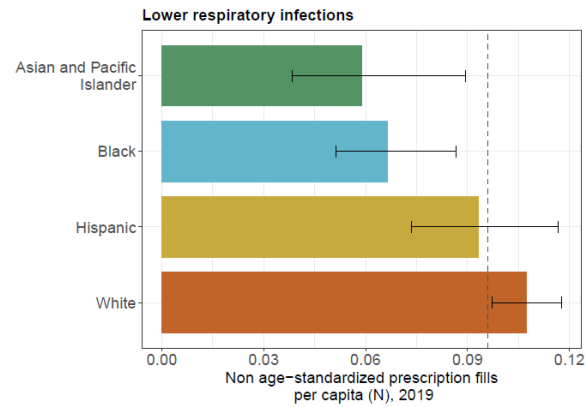
(13) Non-Hodgkin lymphoma



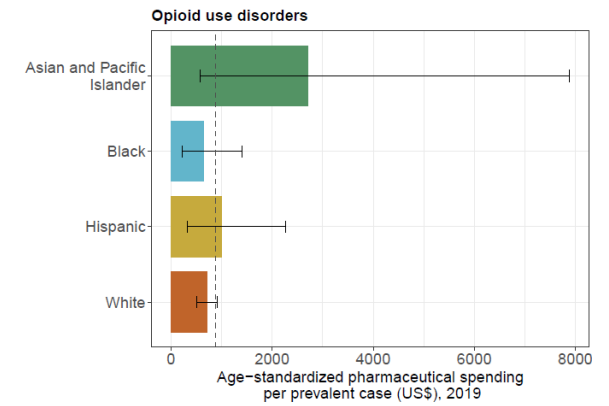
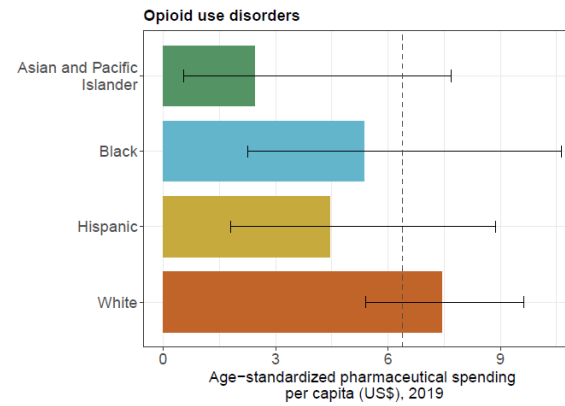
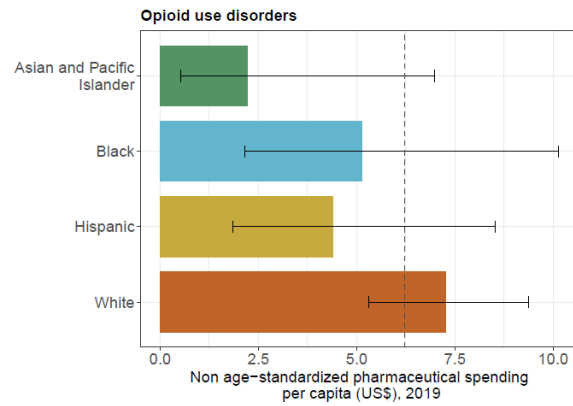
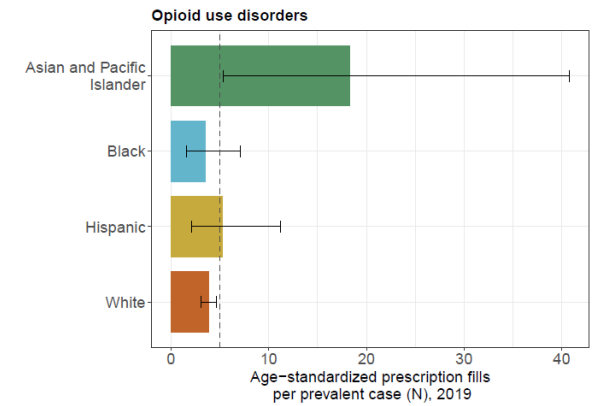
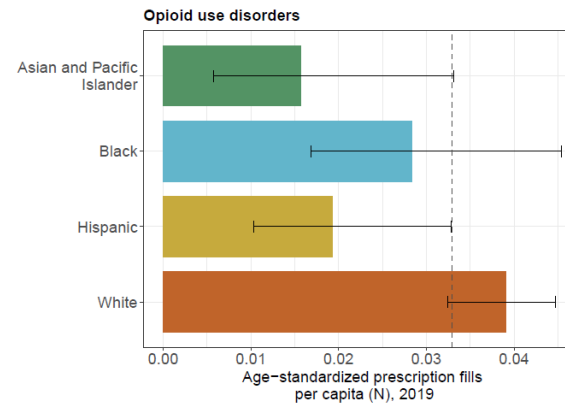
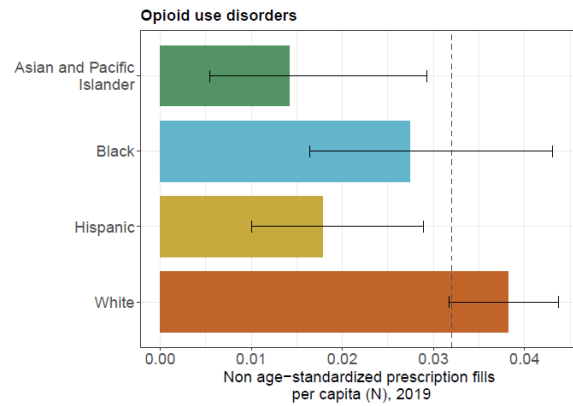
(14) Kidney cancer



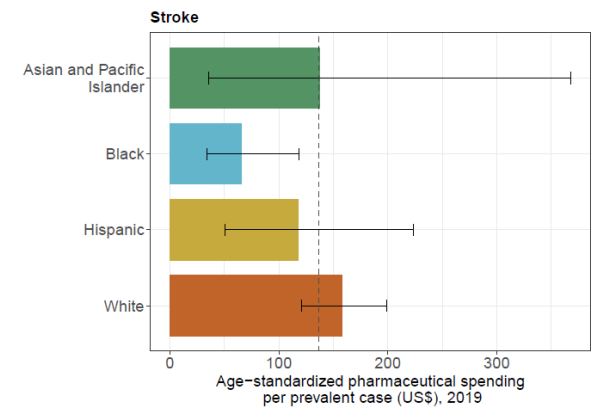
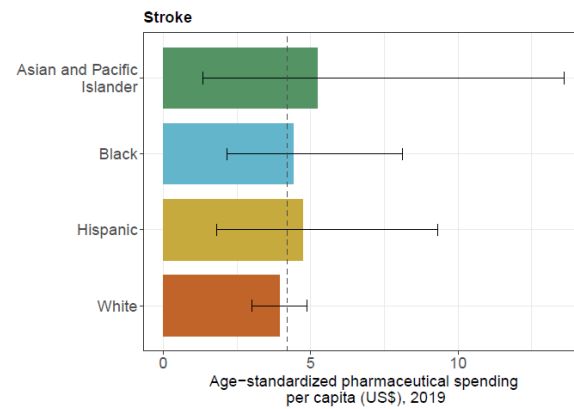
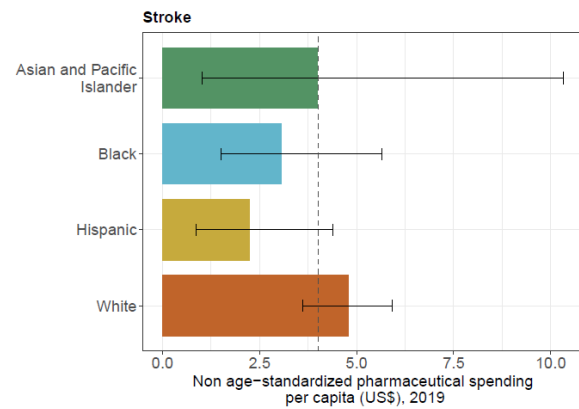
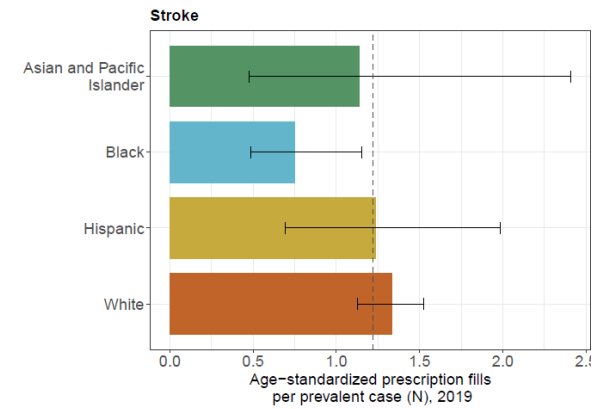
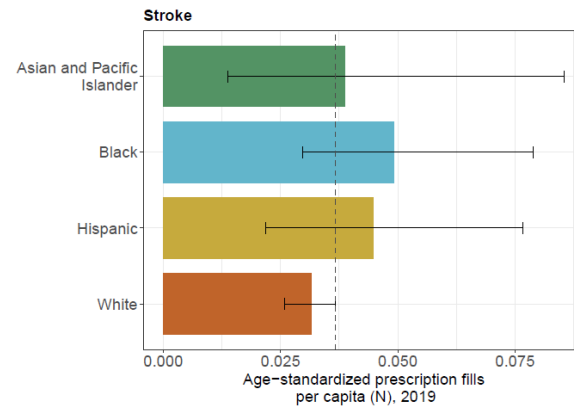
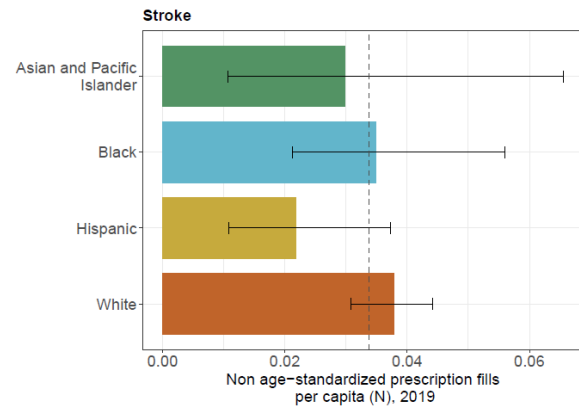
(15) Lower respiratory infections



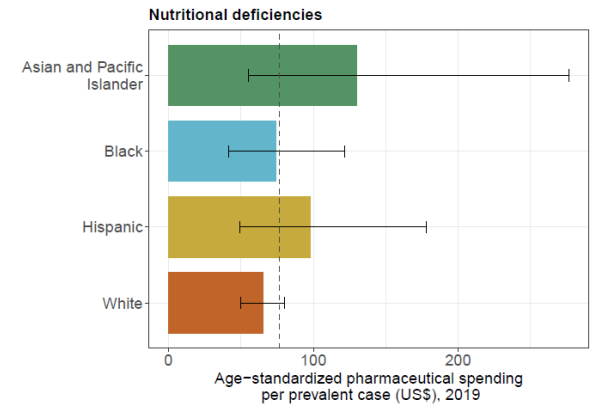
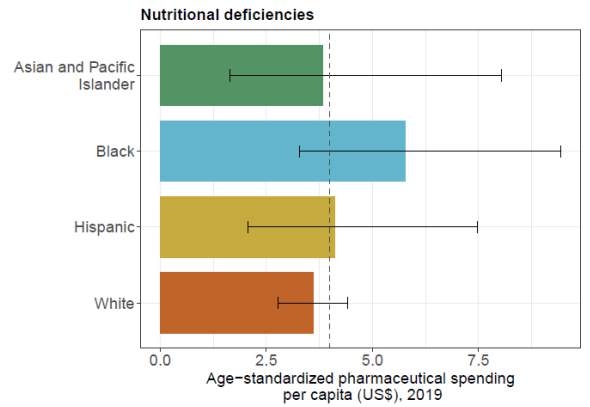
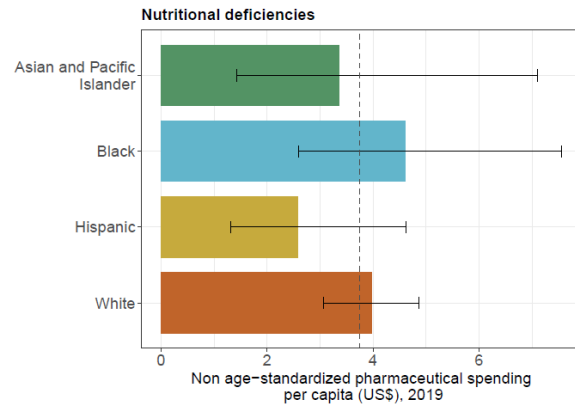
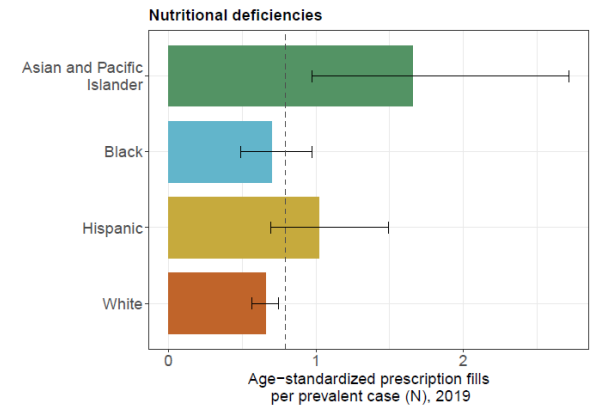
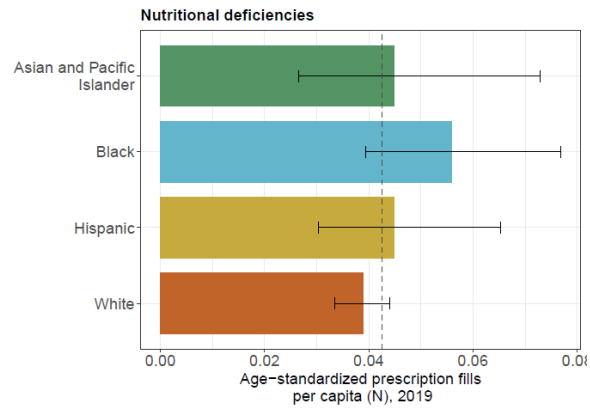
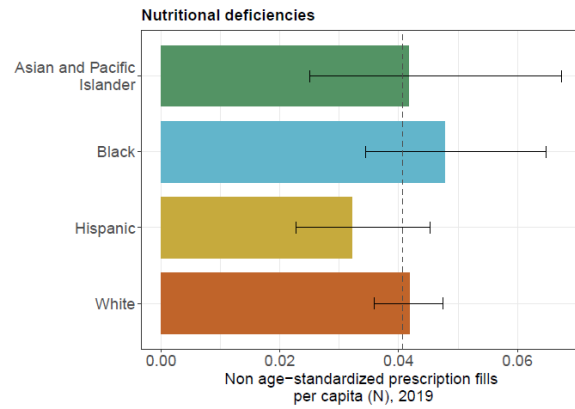
(16) Opioid use disorders



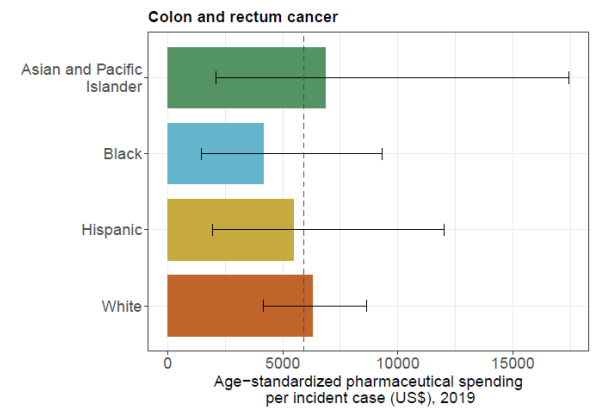
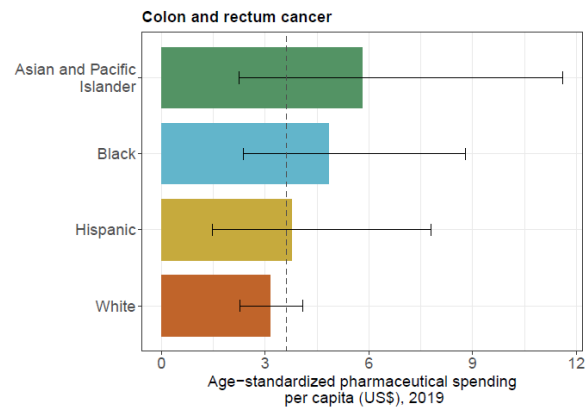
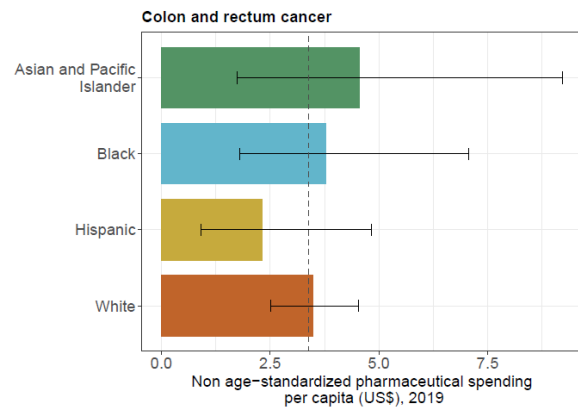
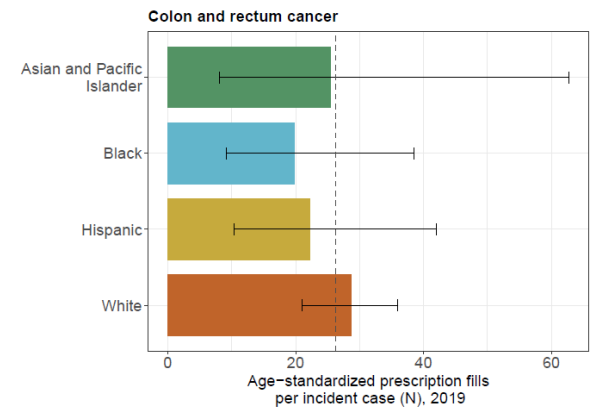
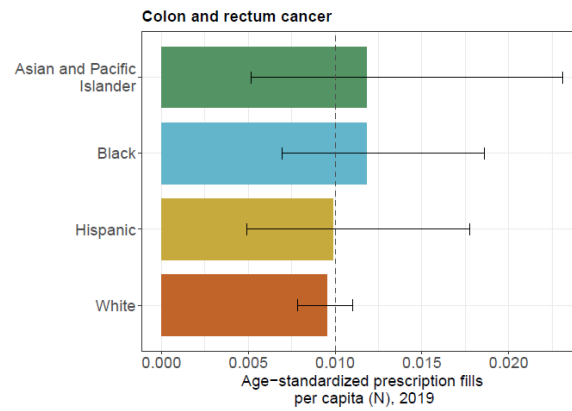
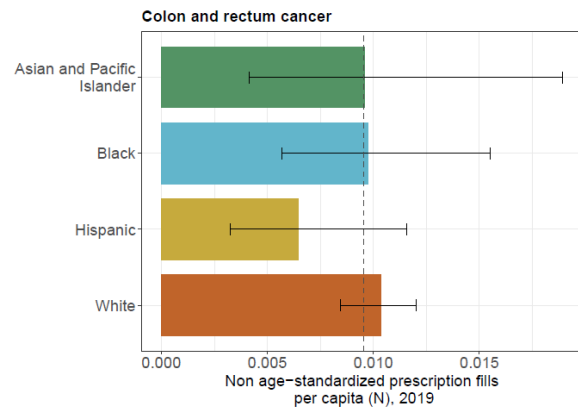
(17) Stroke



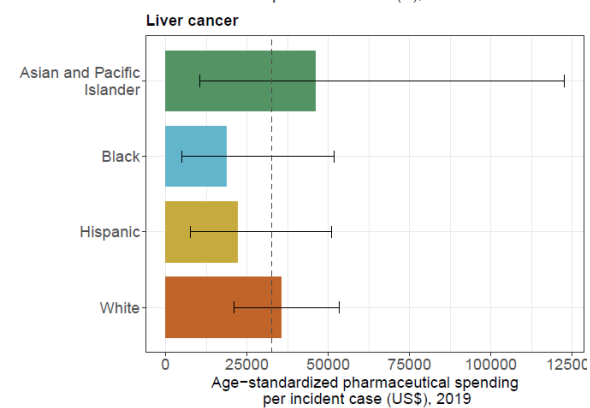
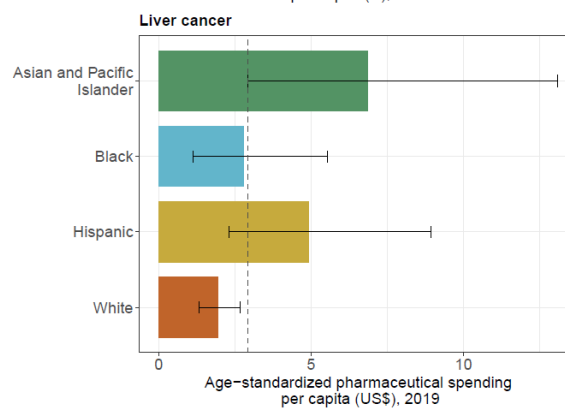
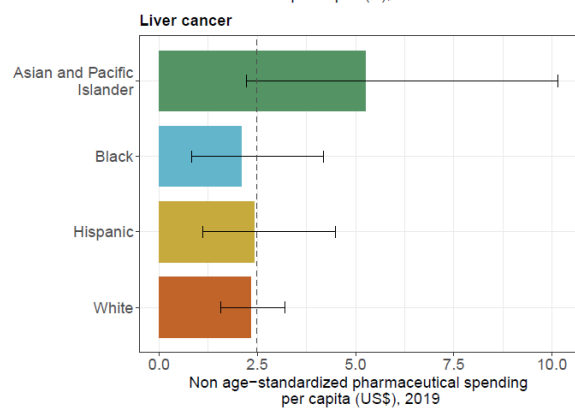
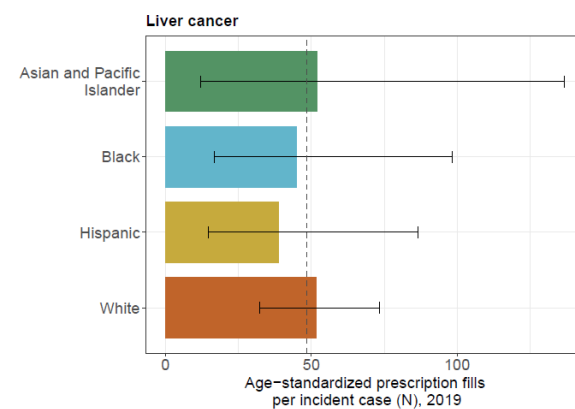
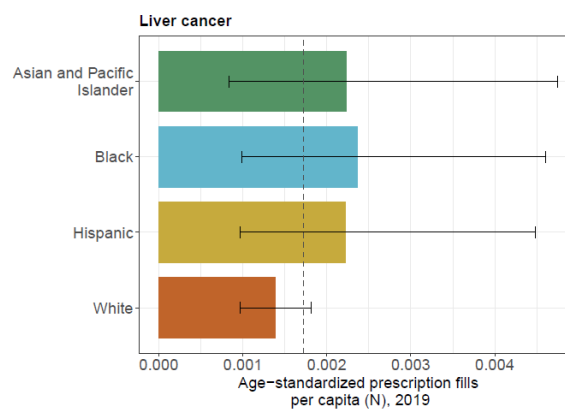
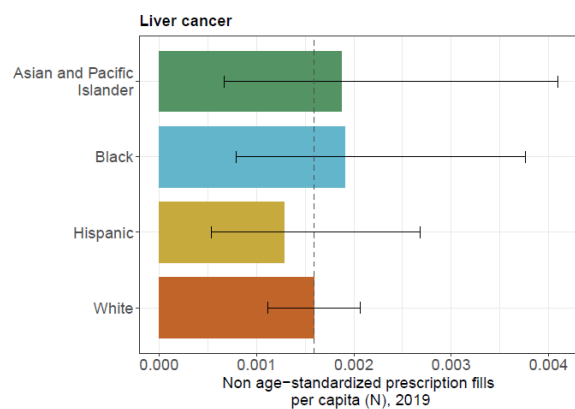
(18) Nutritional deficiencies



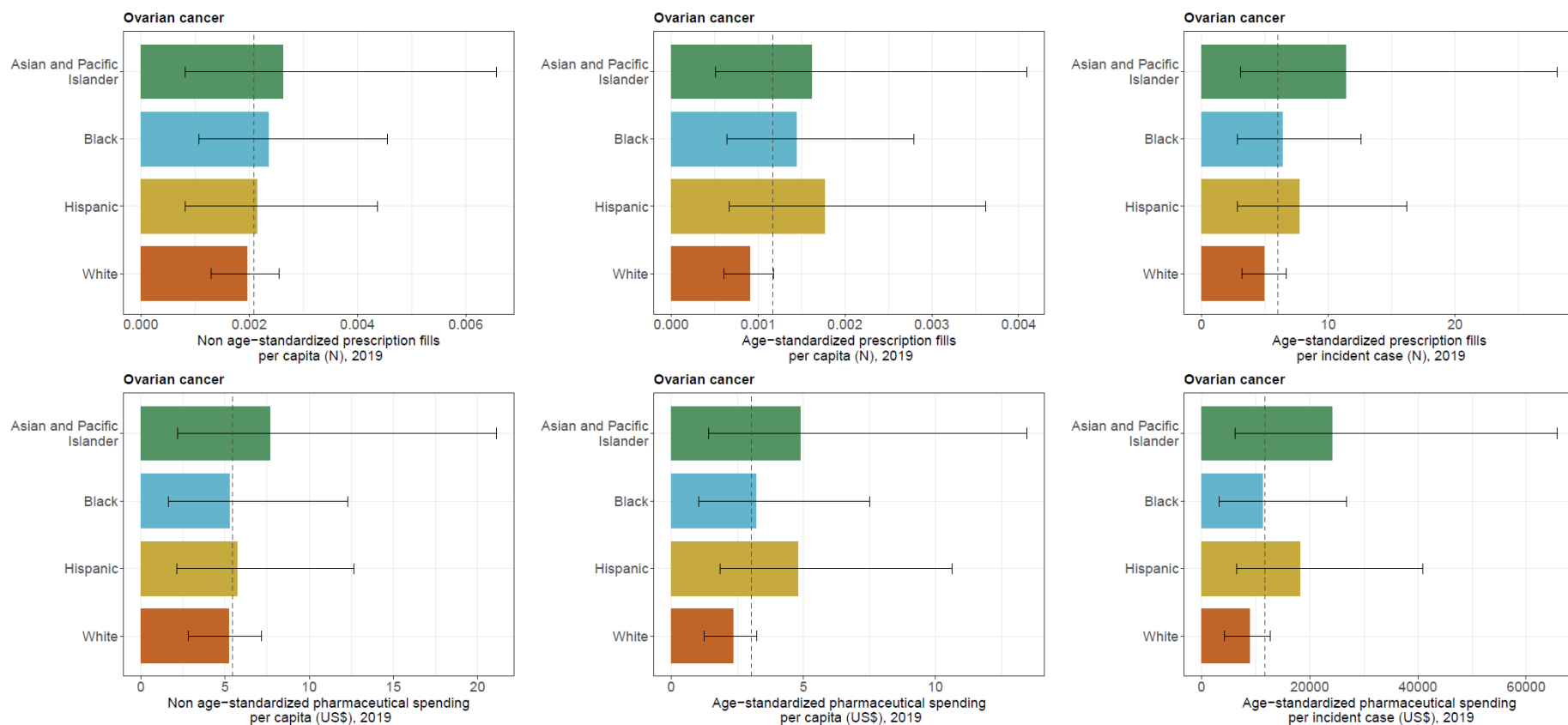
(19) Colon and rectum cancer



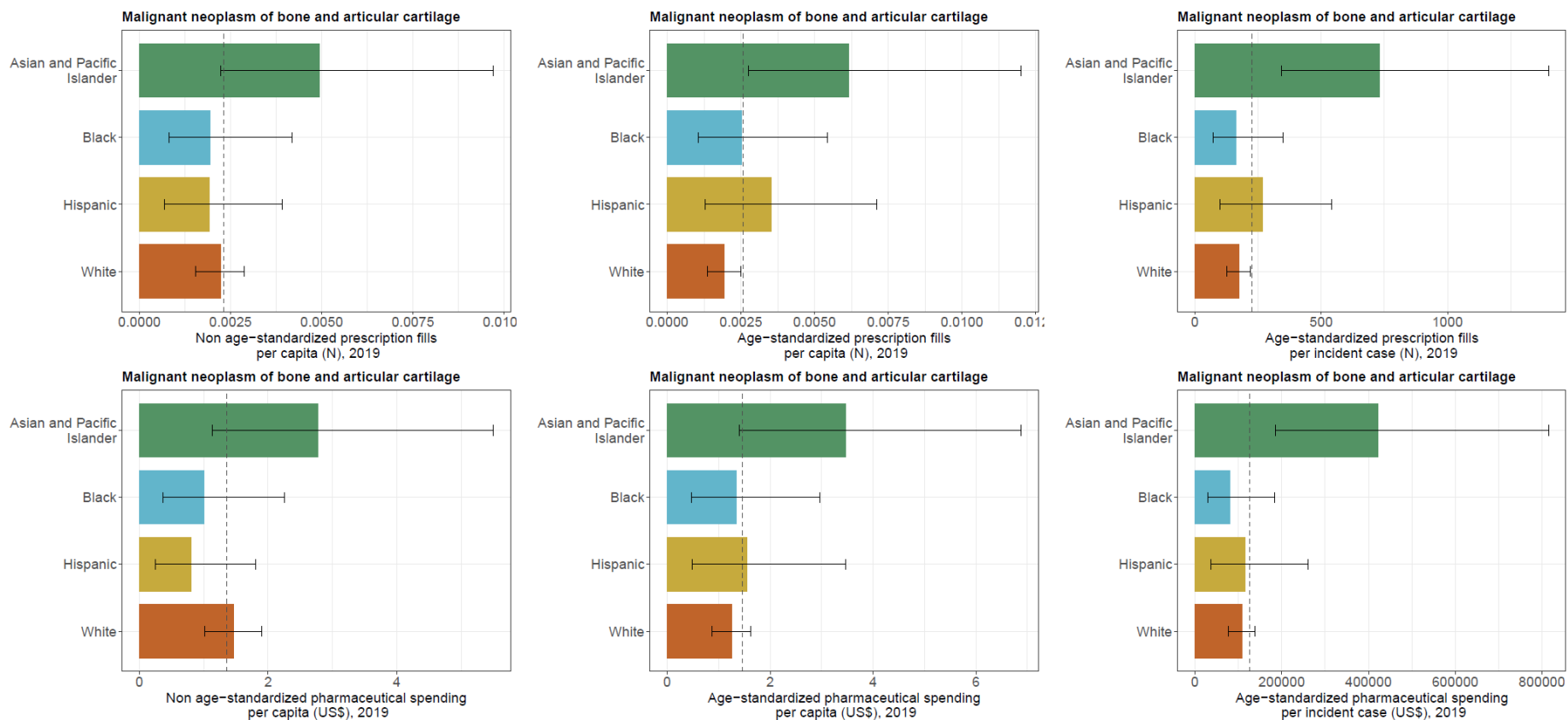
(20) Liver cancer



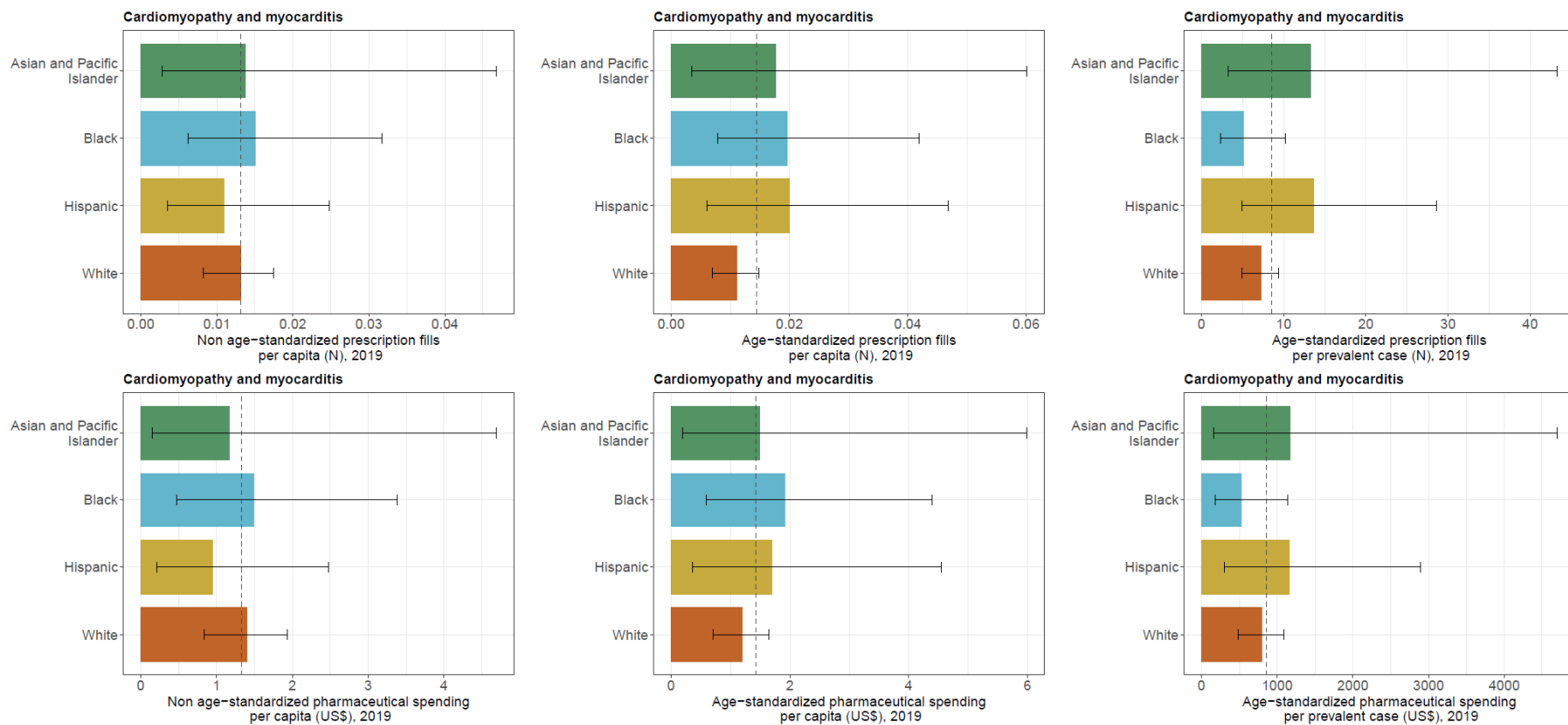
(21) Ovarian cancer



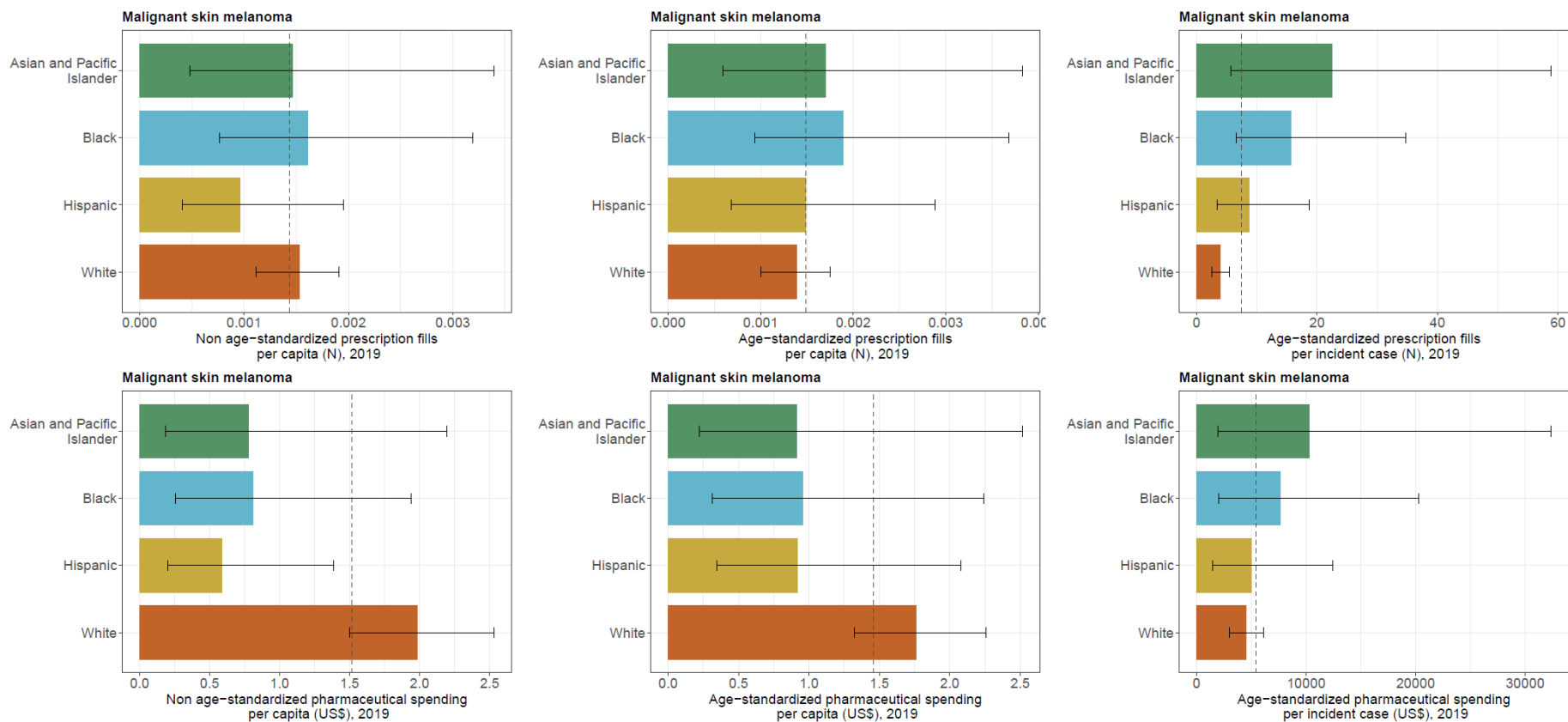
(22) Malignant neoplasm of bone and articular cartilage



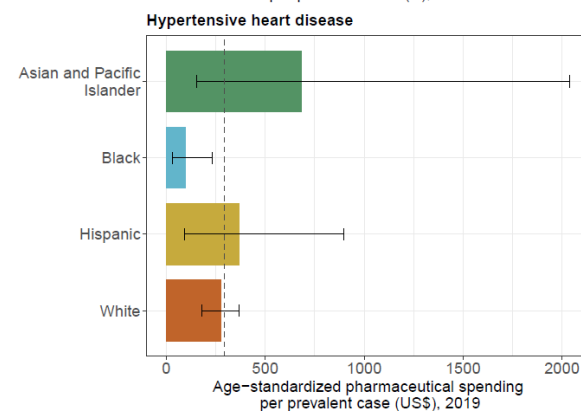
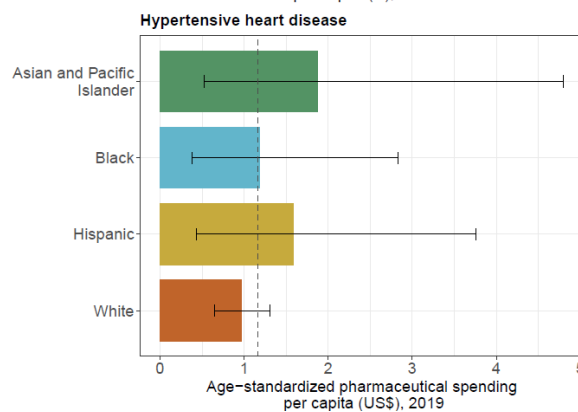
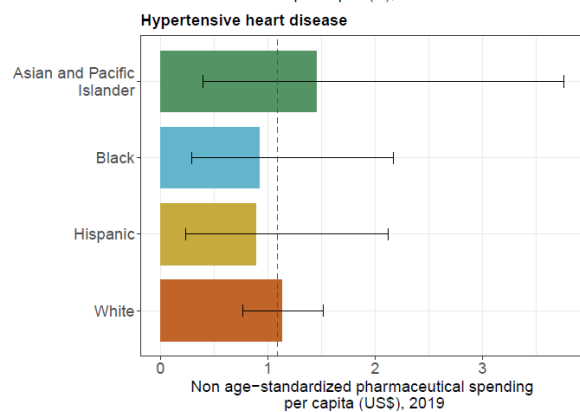
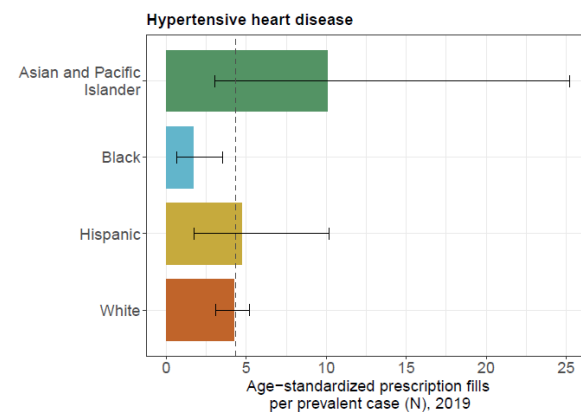
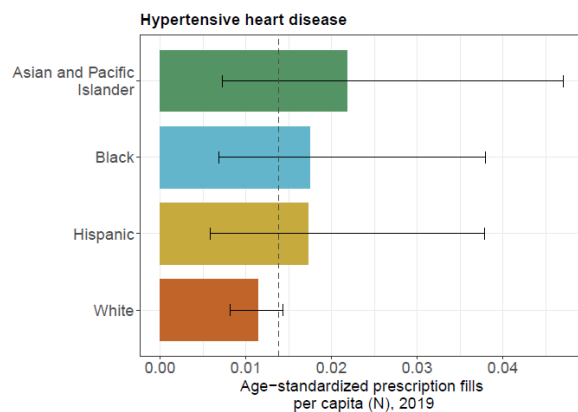
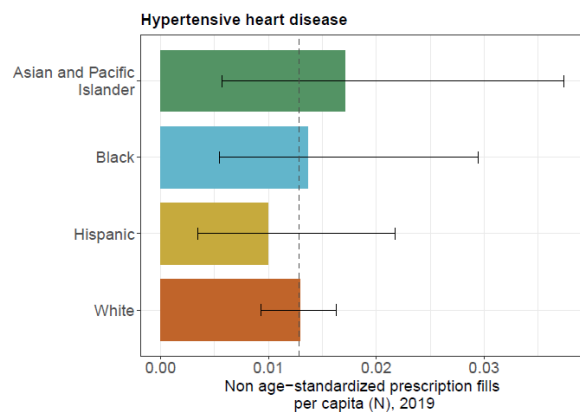
(23) Cardiomyopathy and myocarditis



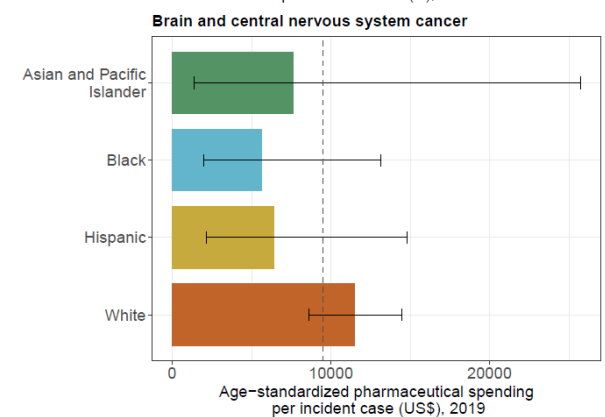
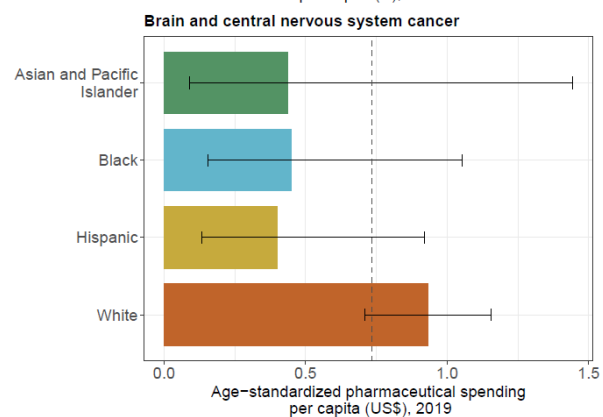
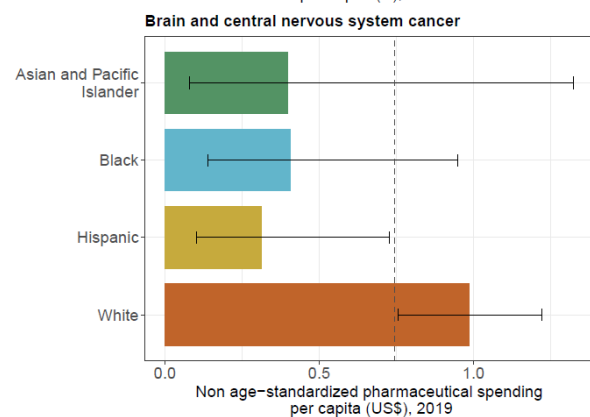
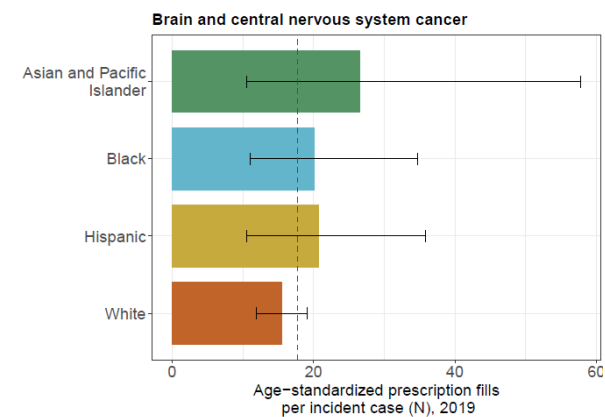
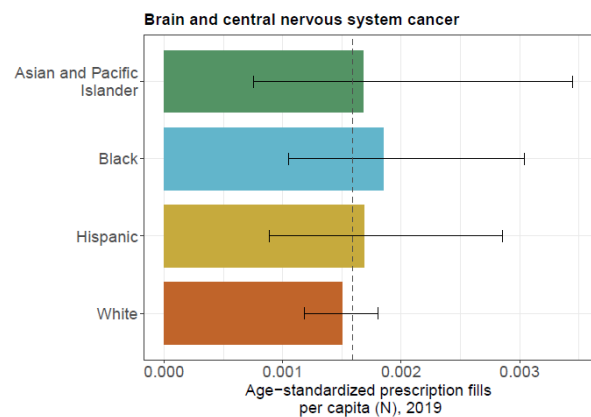
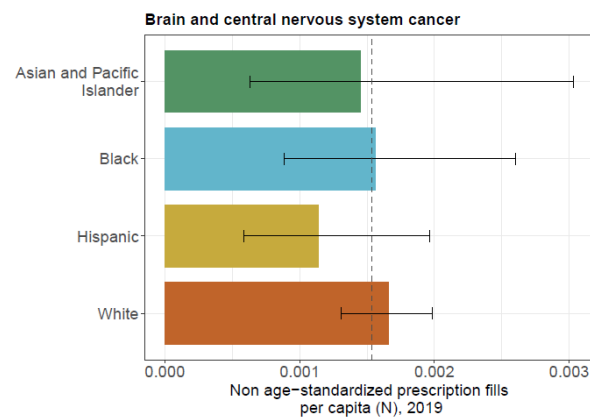
(24) Malignant skin melanoma



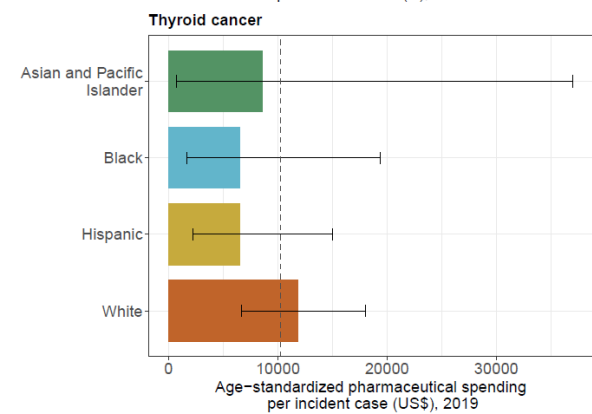
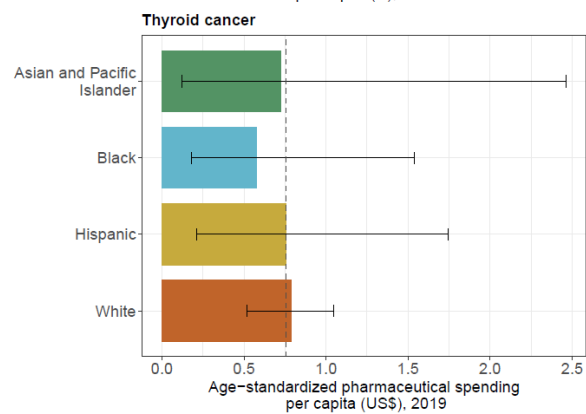
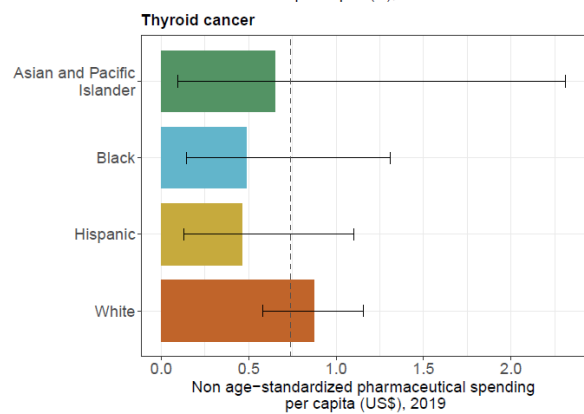
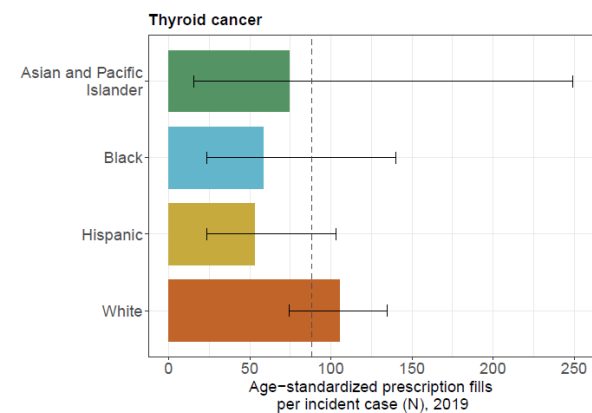
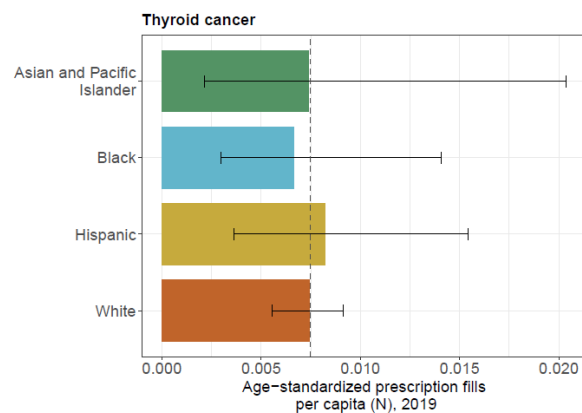
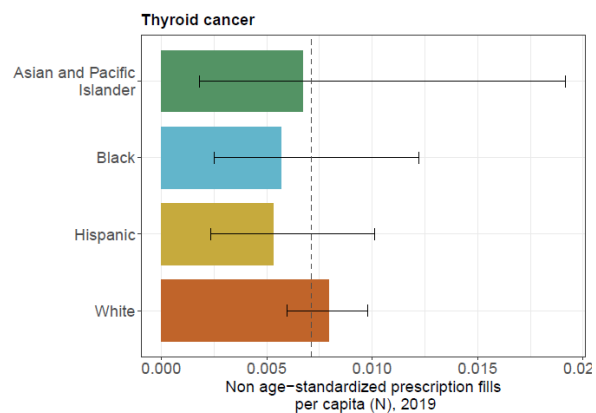
(25) Hypertensive heart disease



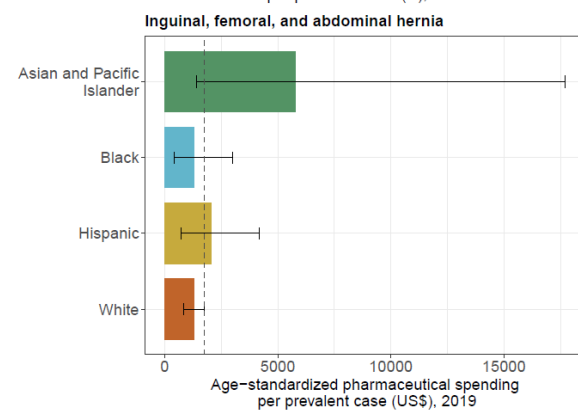
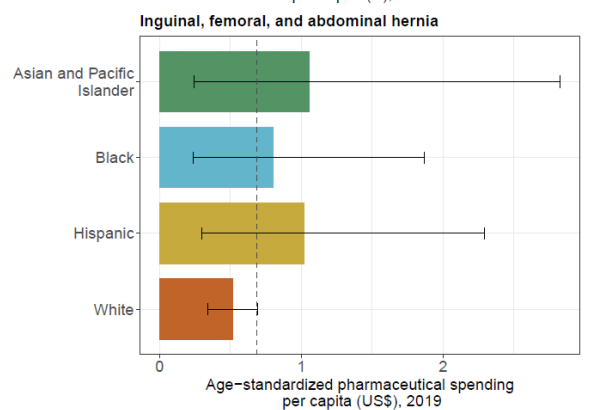
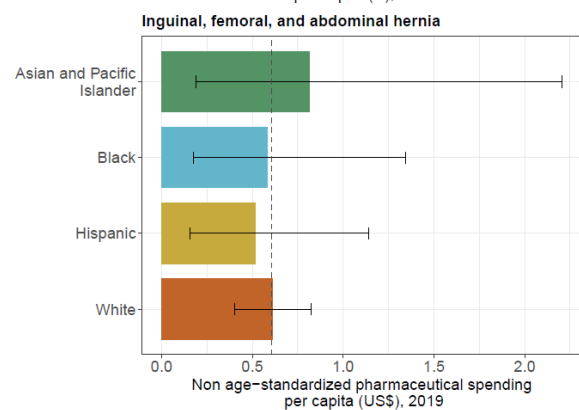
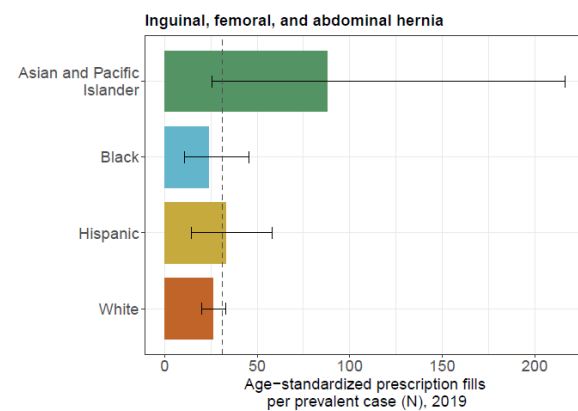
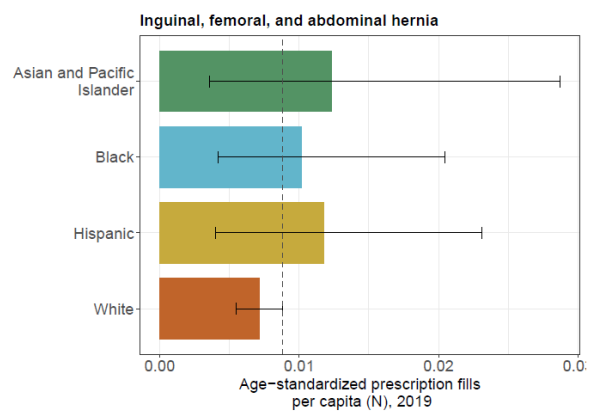
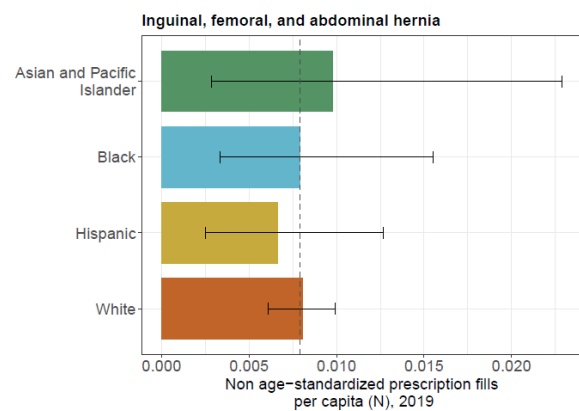
(26) Brain and central nervous system cancer



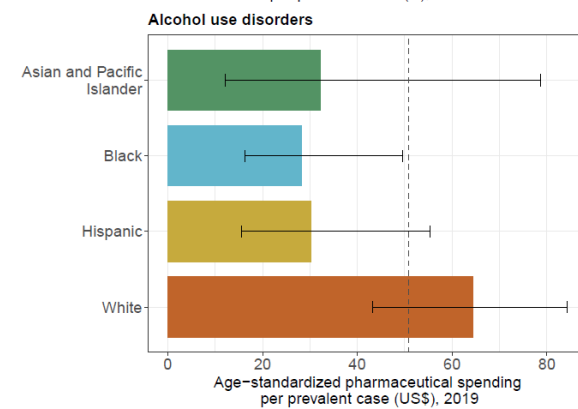
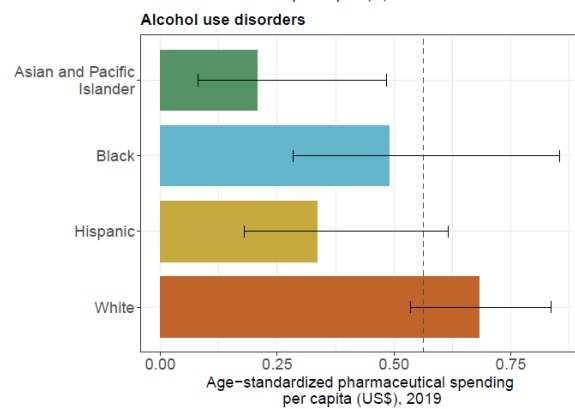
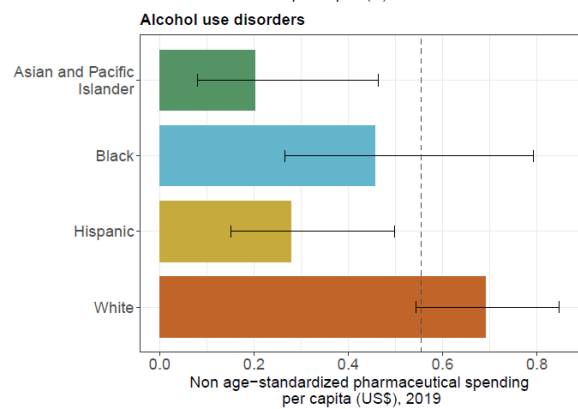
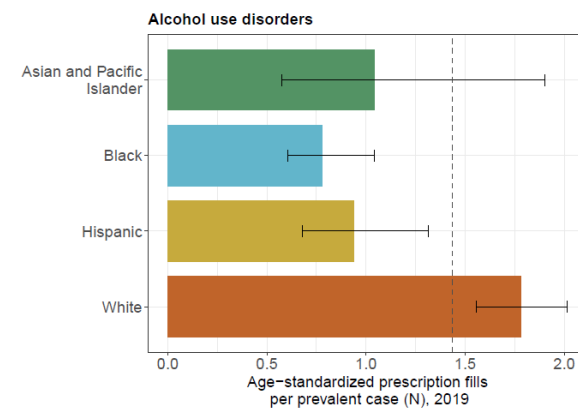
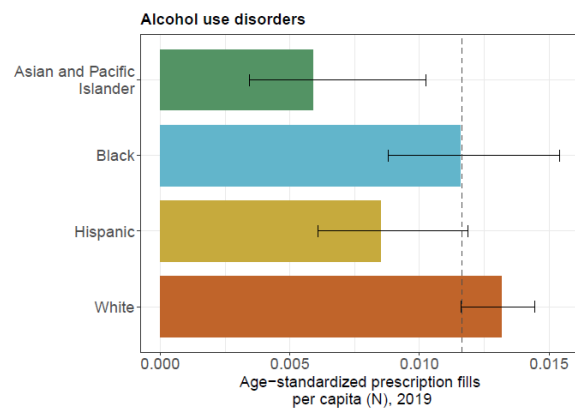
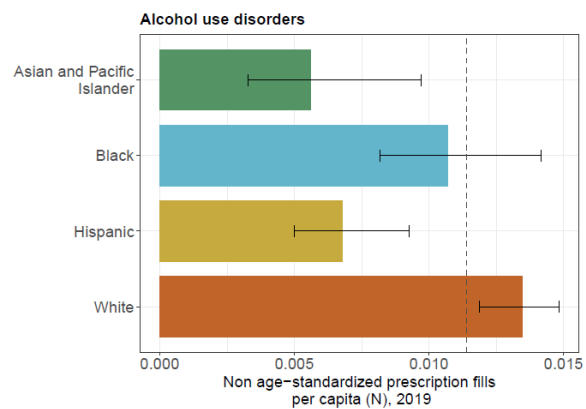
(27) Thyroid cancer



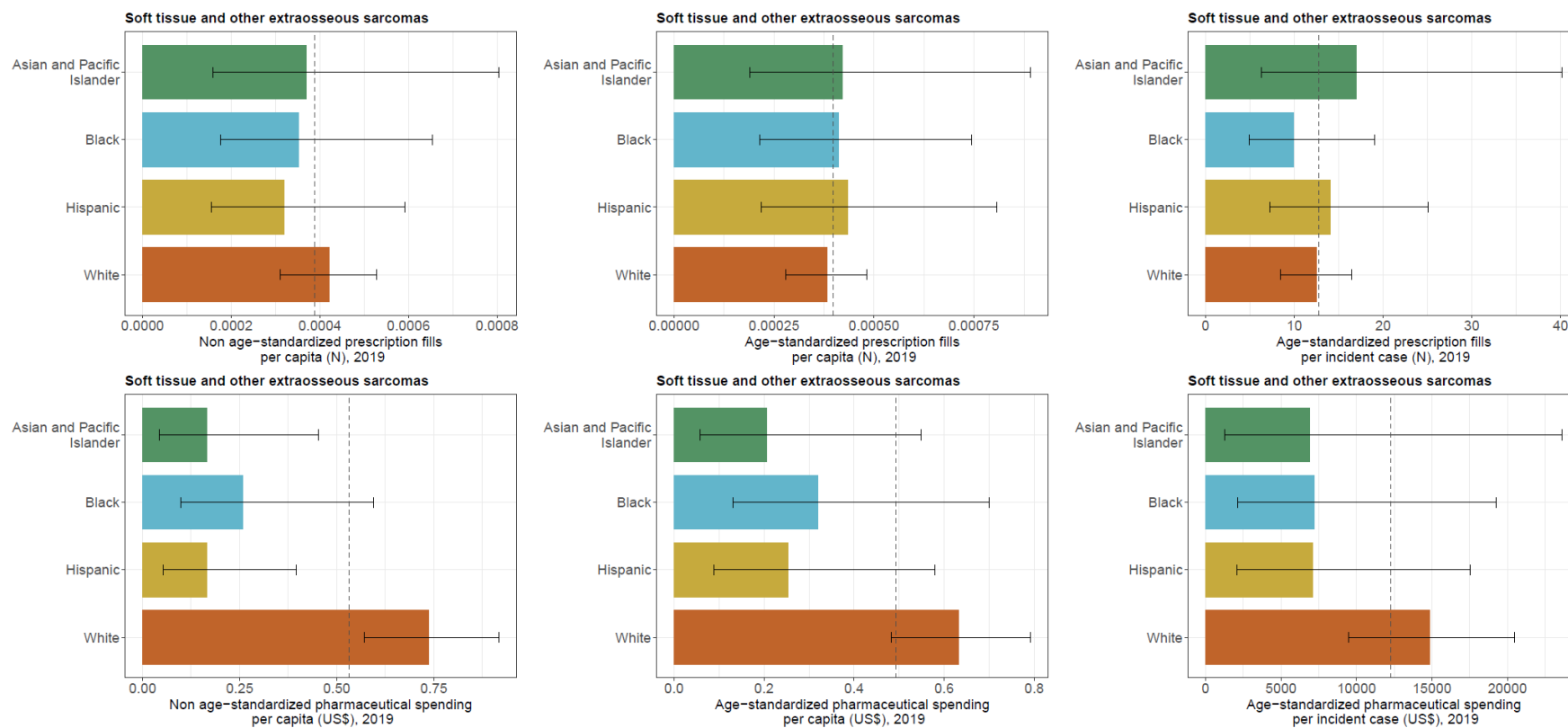
(28) Inguinal, femoral, and abdominal hernia



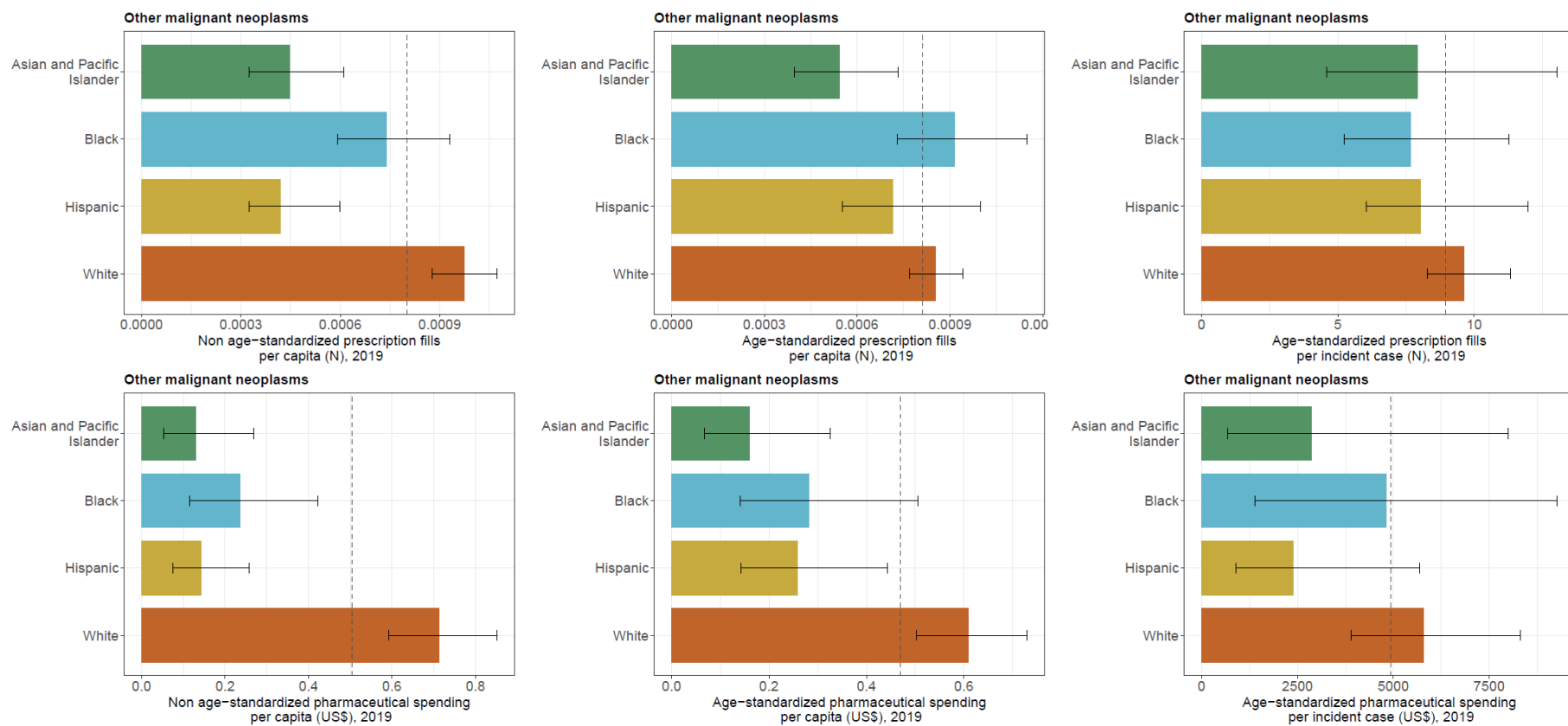
(29) Alcohol use disorders



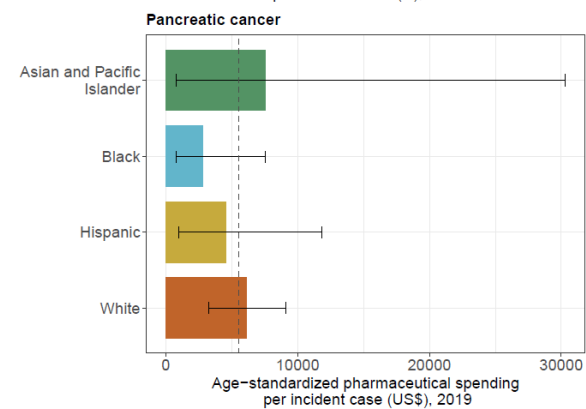
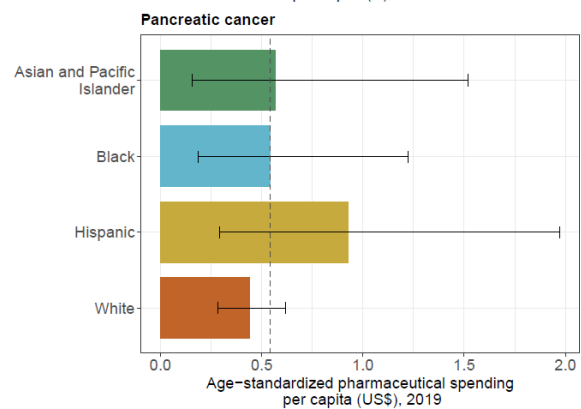
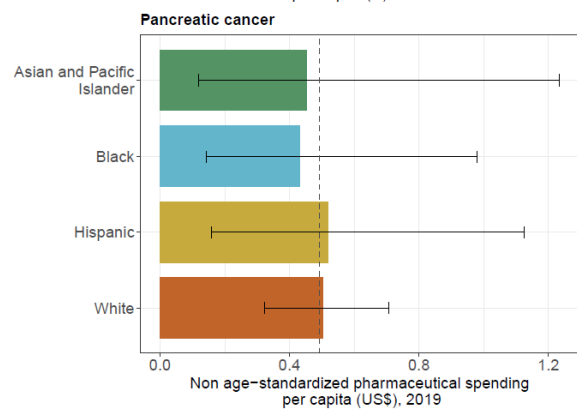
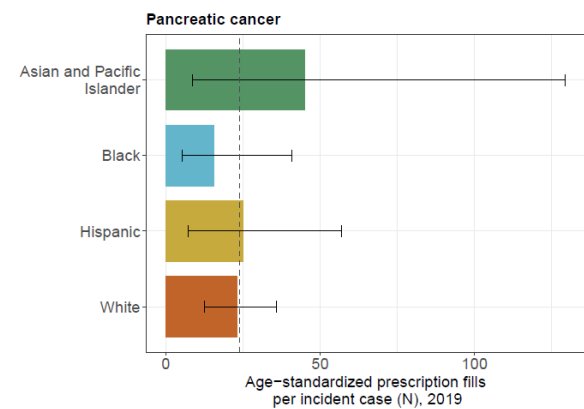
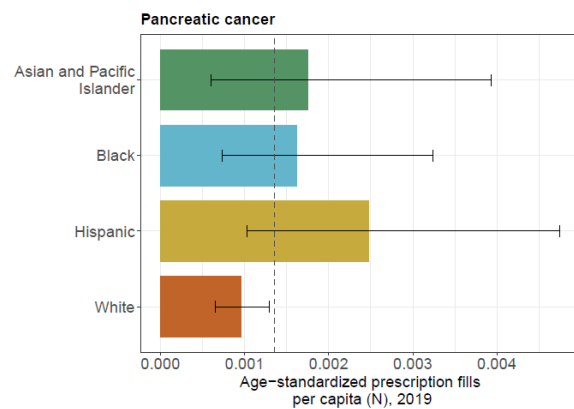
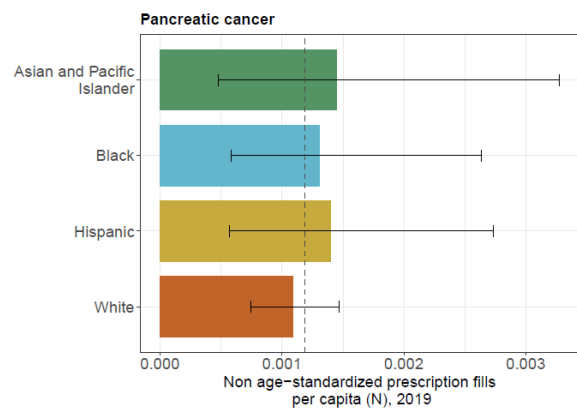
(30) Soft tissue and other extraosseous sarcomas



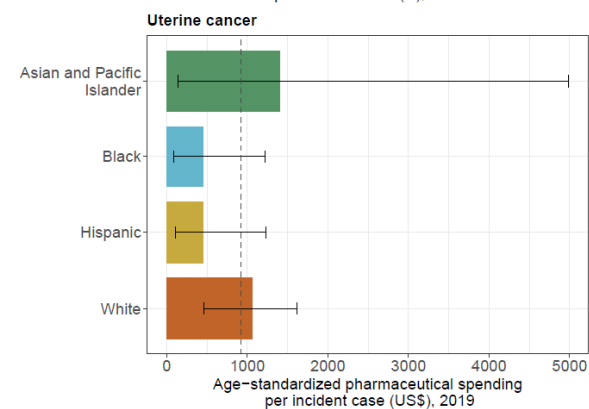
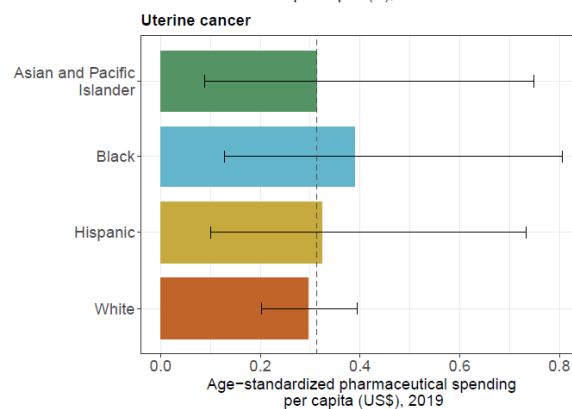
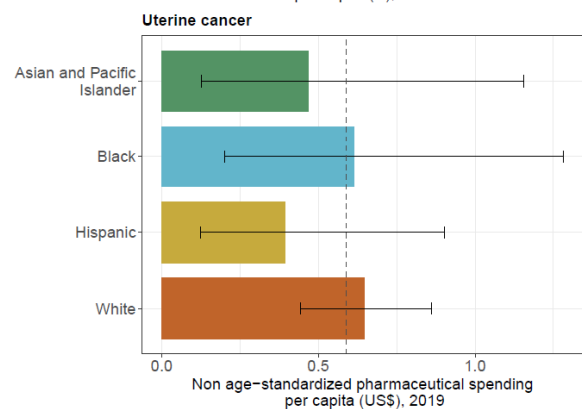
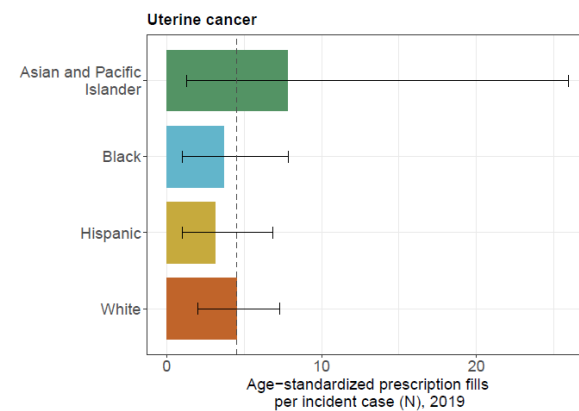
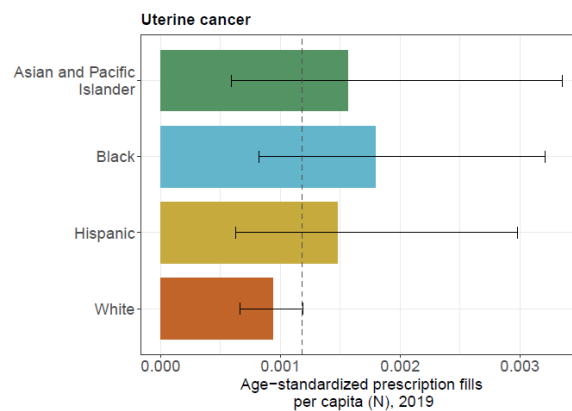
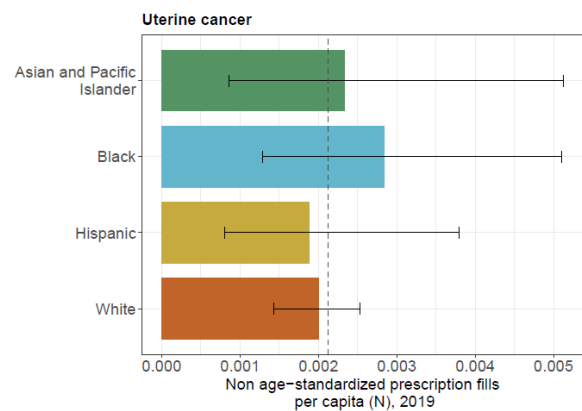
(31) Other malignant neoplasms



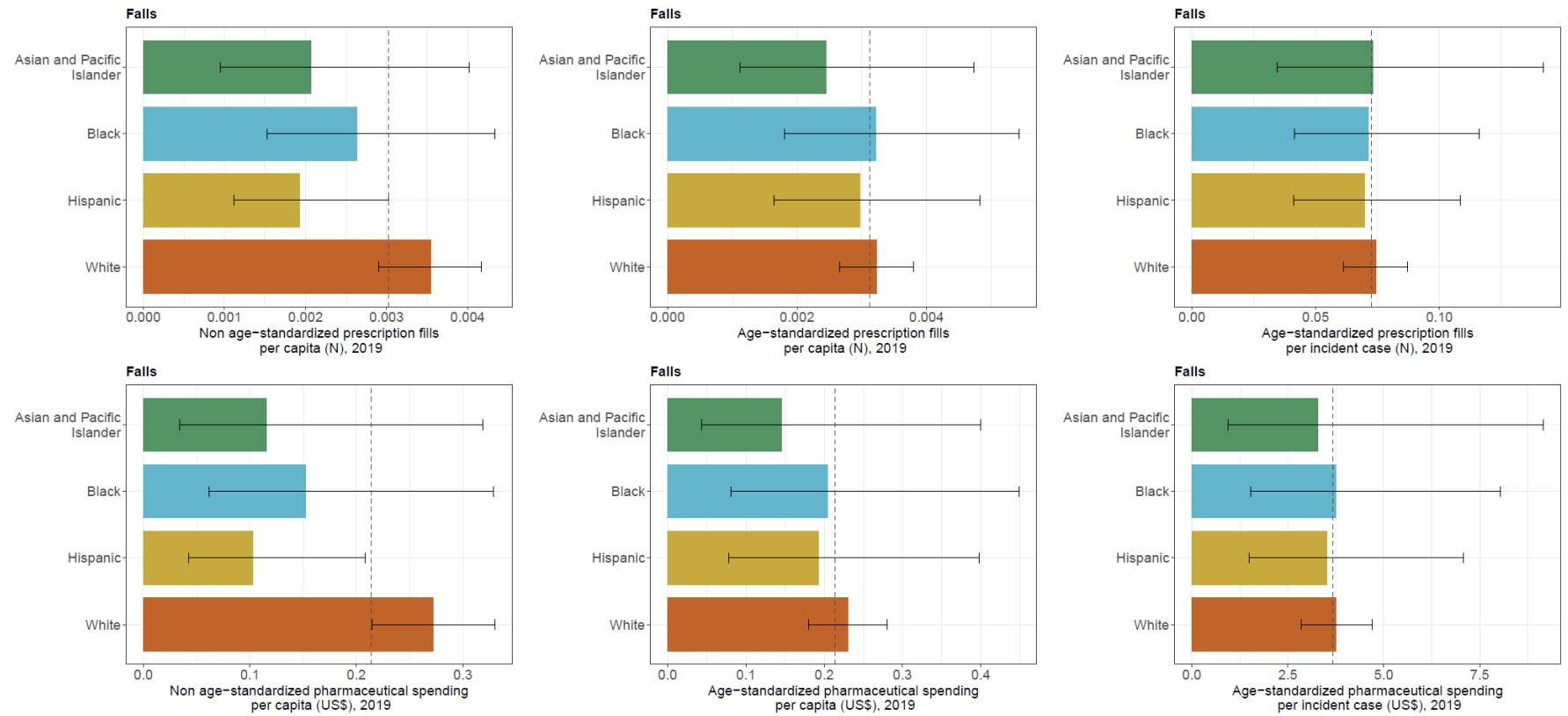
(32) Pancreatic cancer



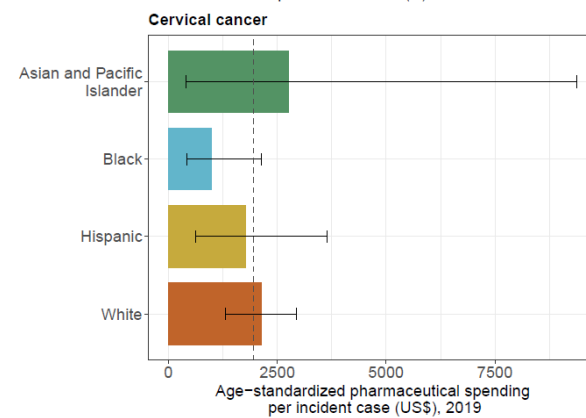
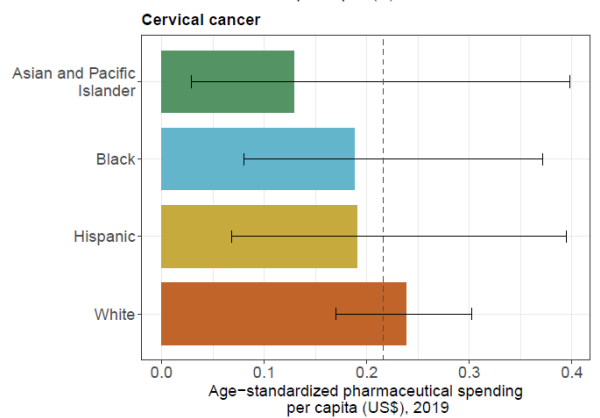
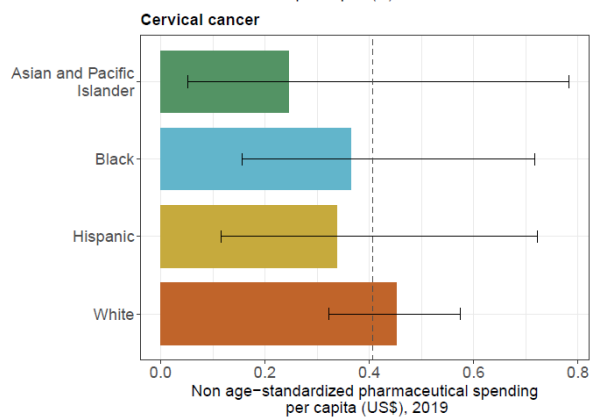
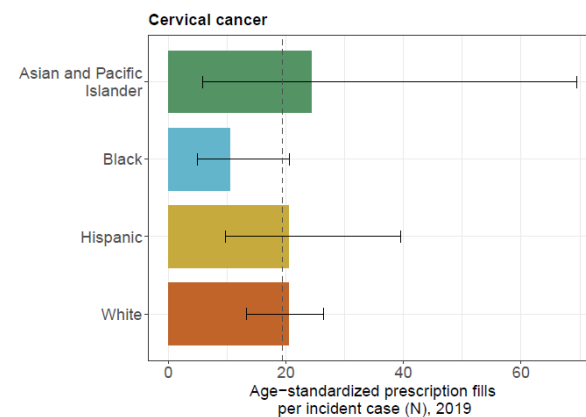
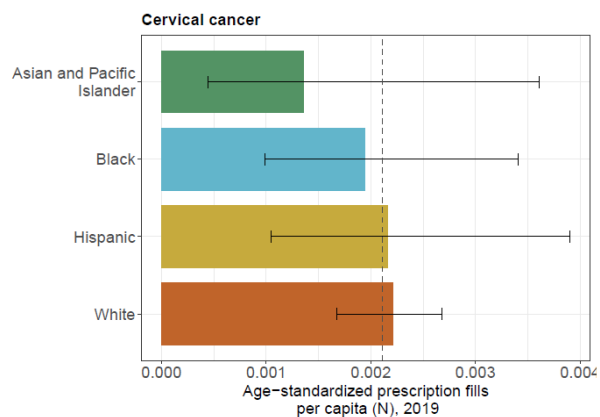
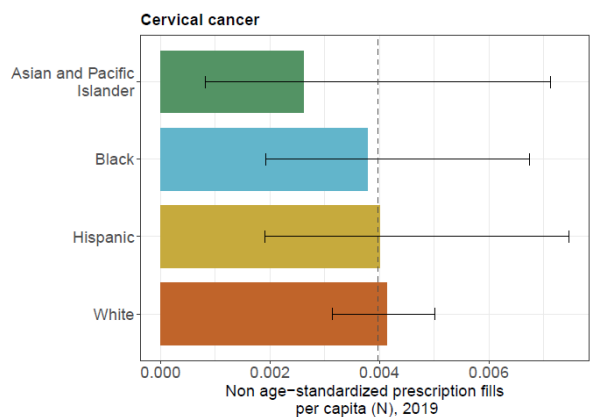
(33) Uterine cancer



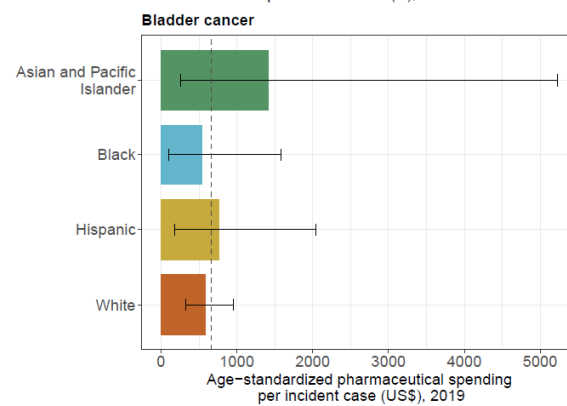
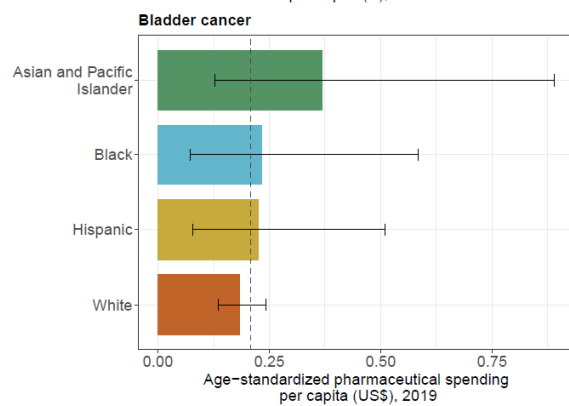
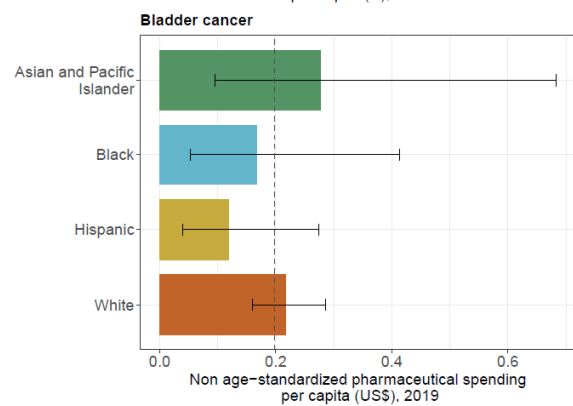
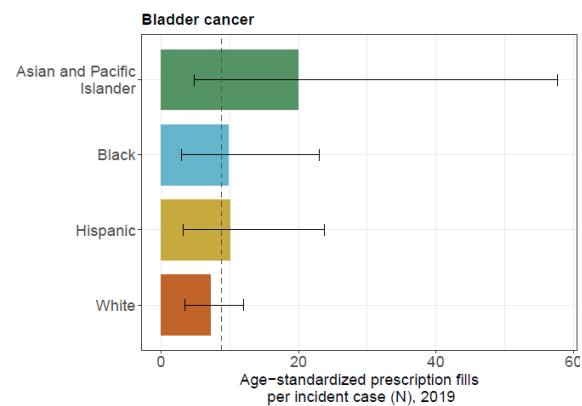
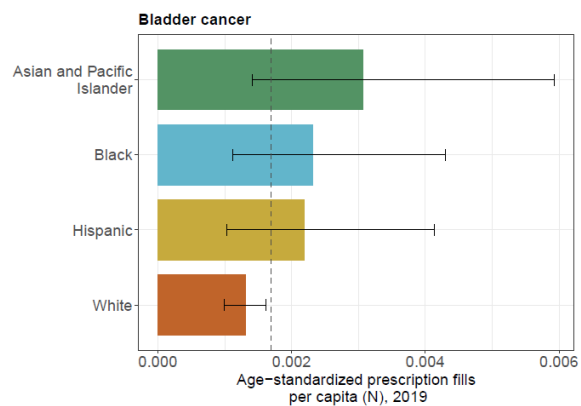
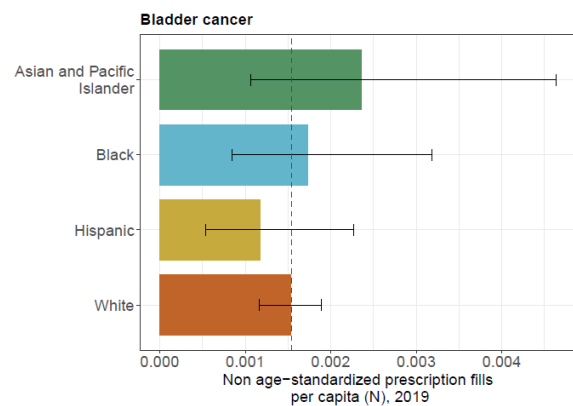
(34) Falls



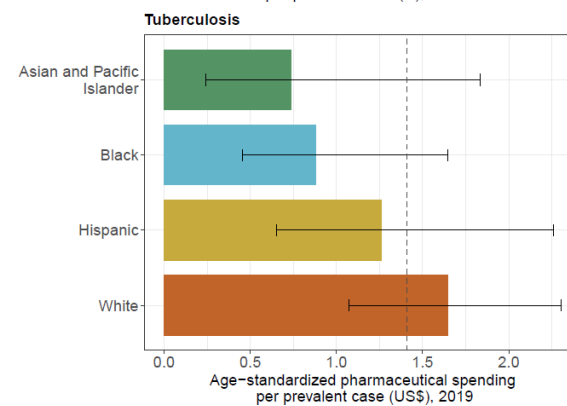
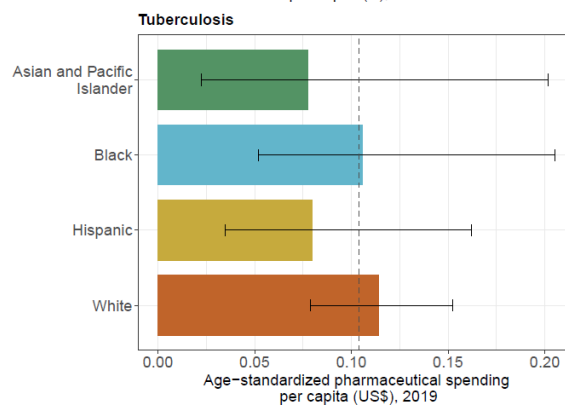
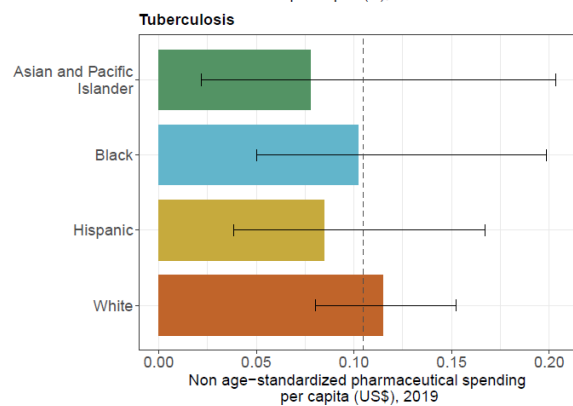
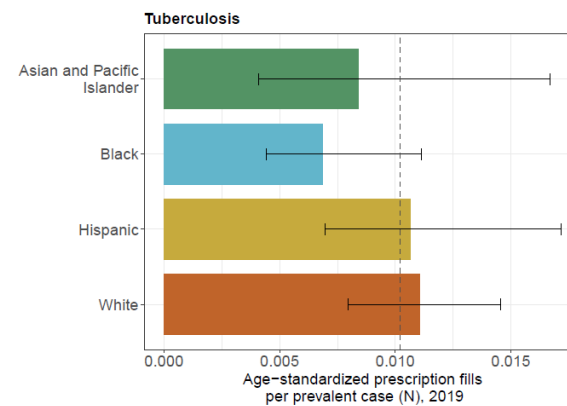
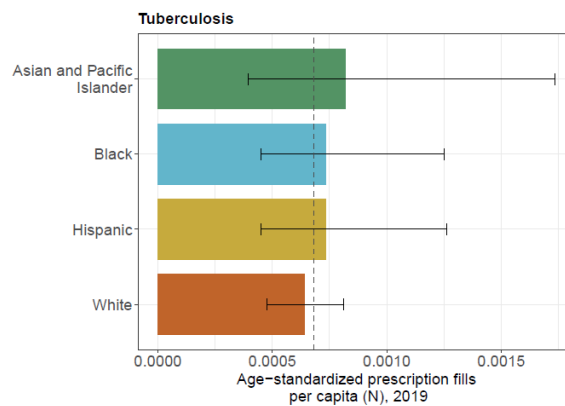
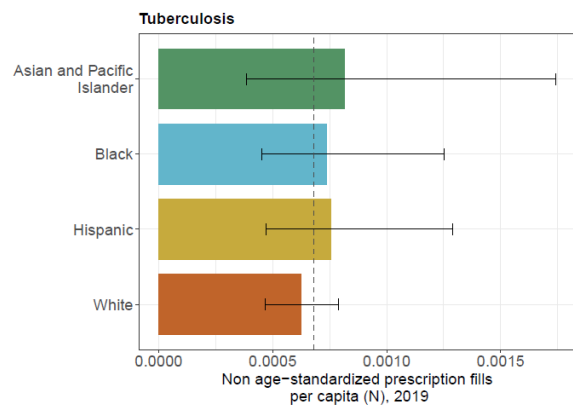
(35) Cervical cancer



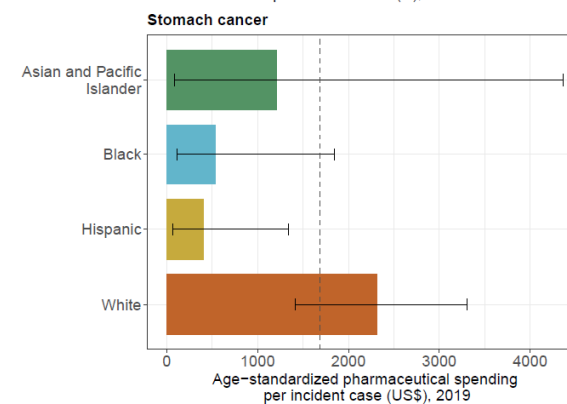
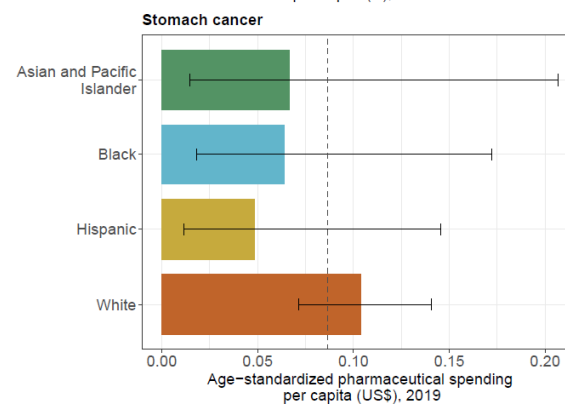
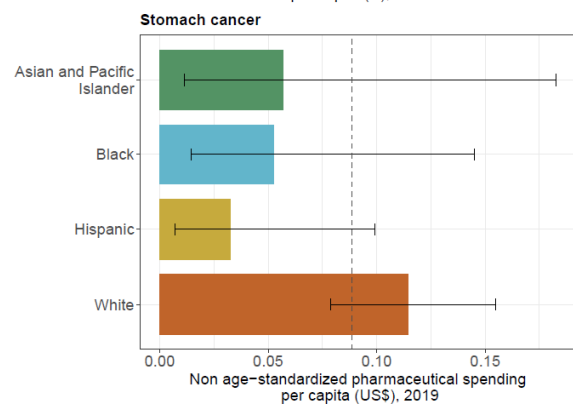
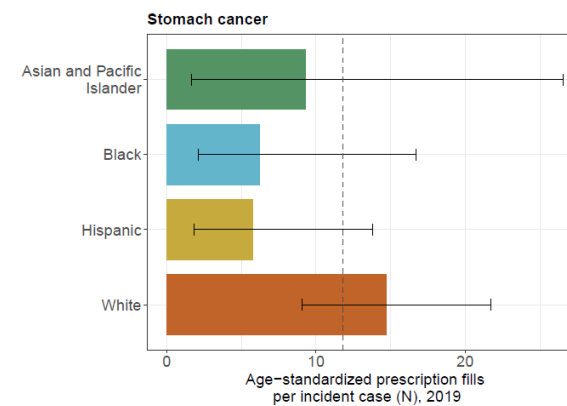
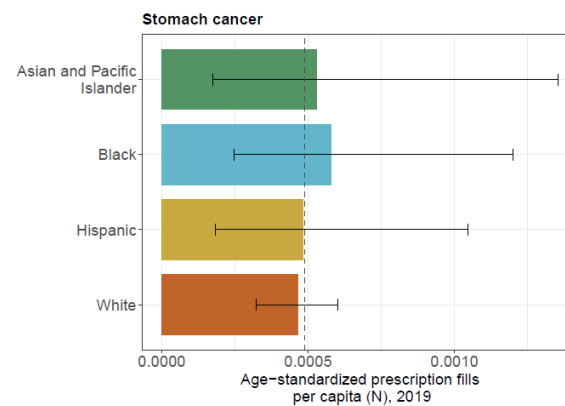
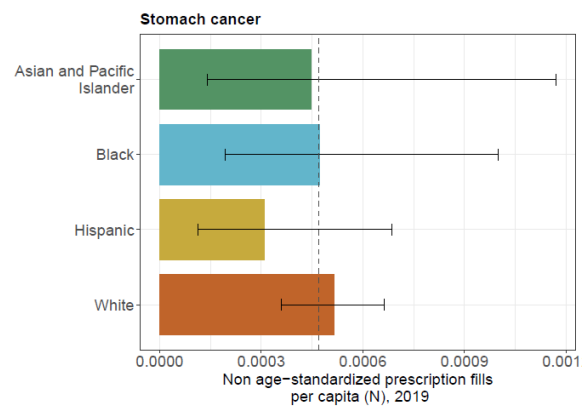
(36) Bladder cancer



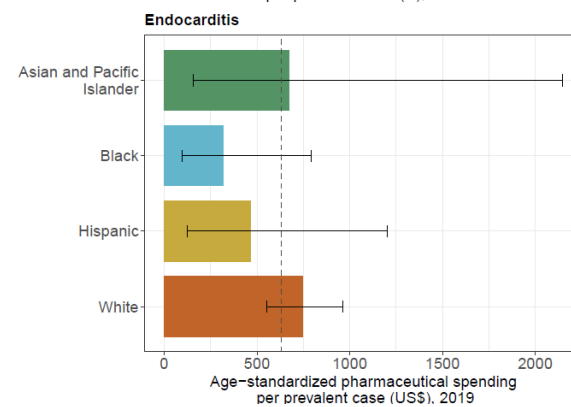
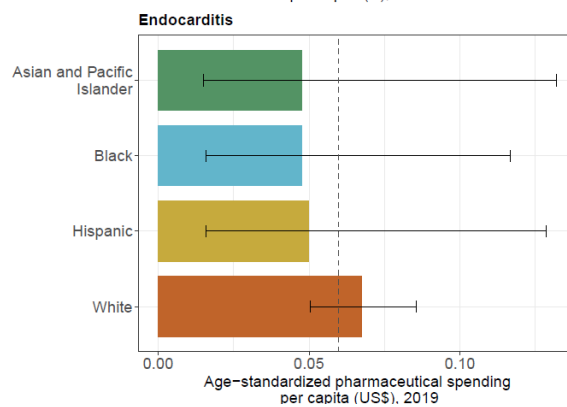
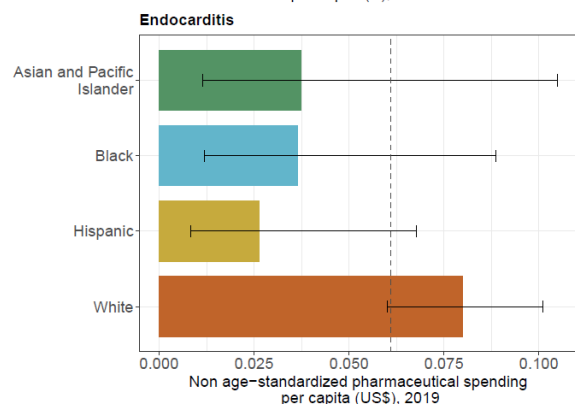
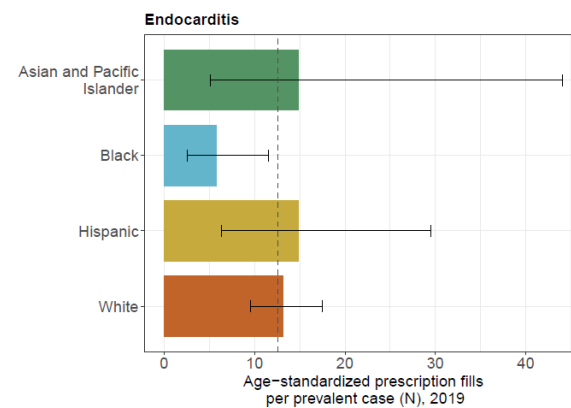
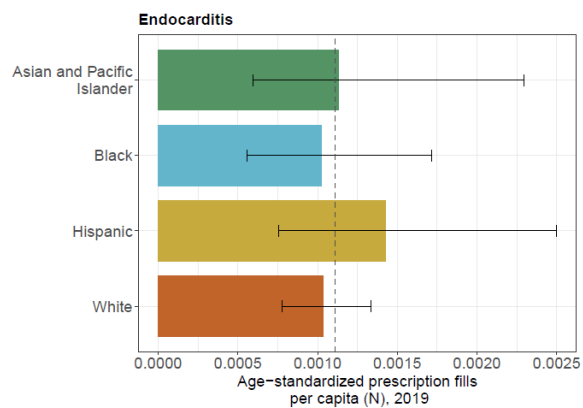
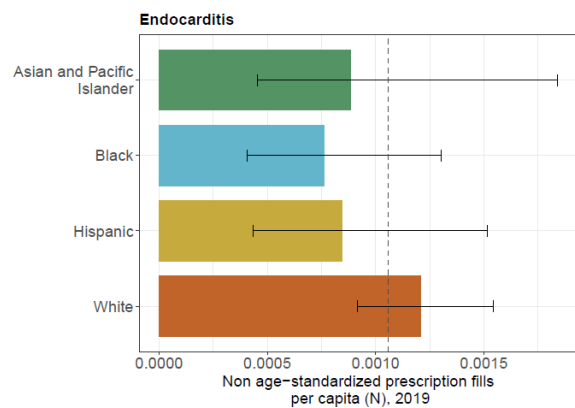
(37) Tuberculosis



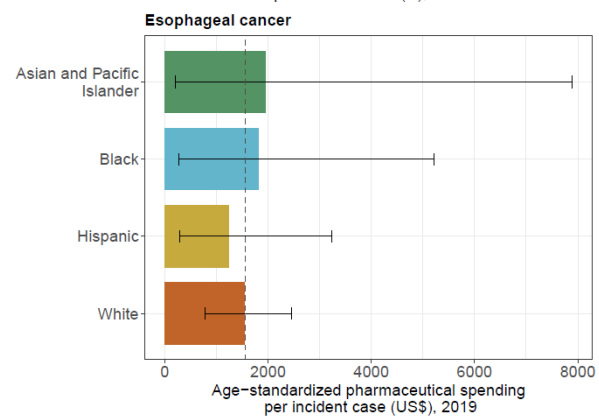
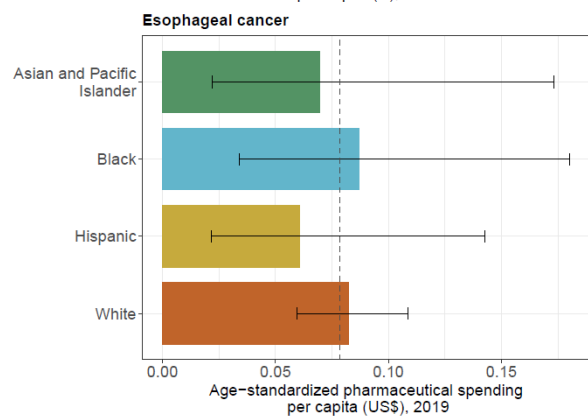
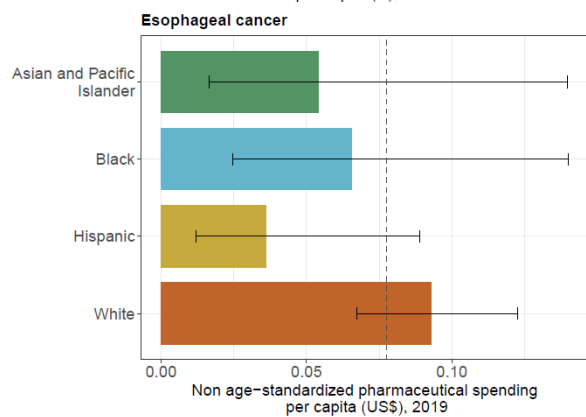
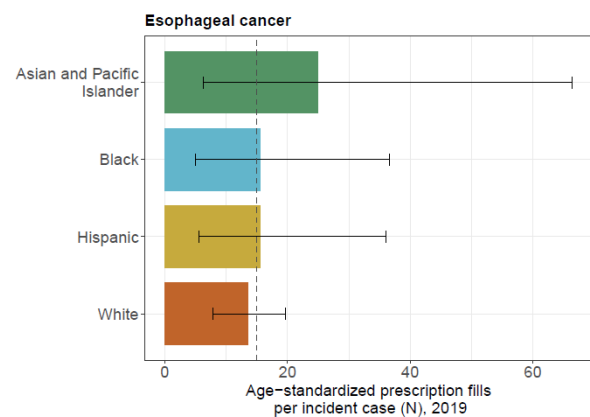
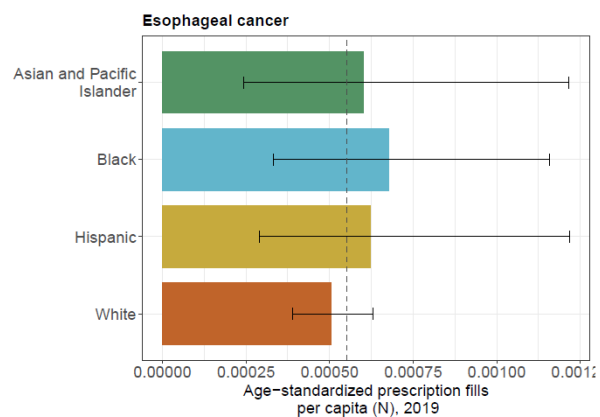
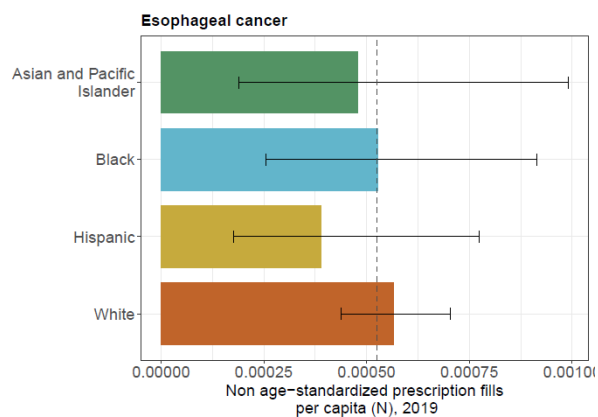
(38) Stomach cancer



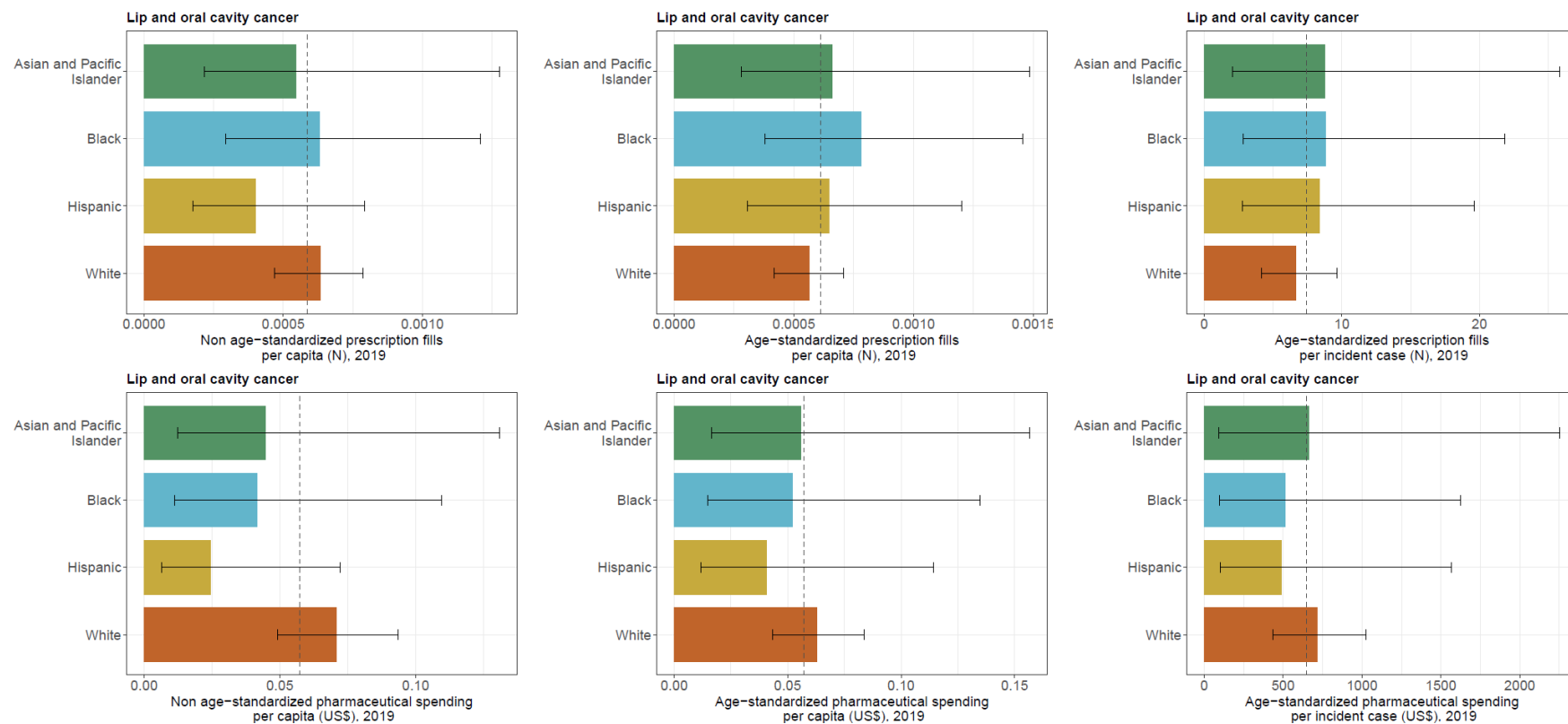
(39) Endocarditis



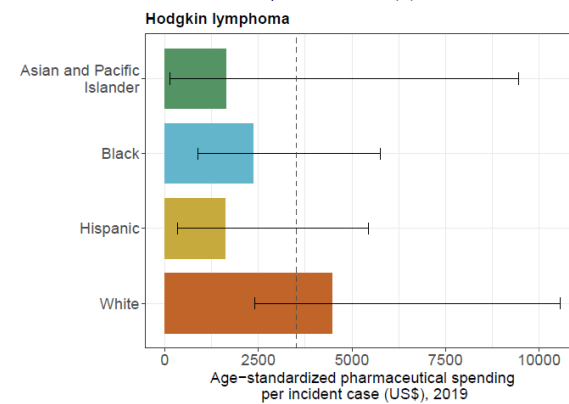
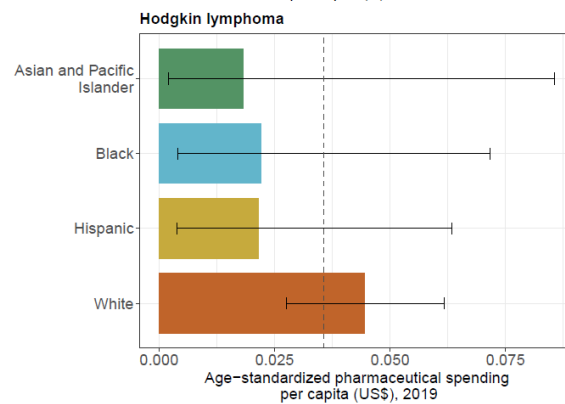
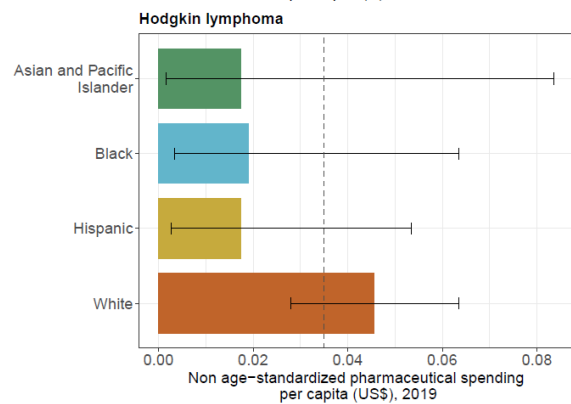
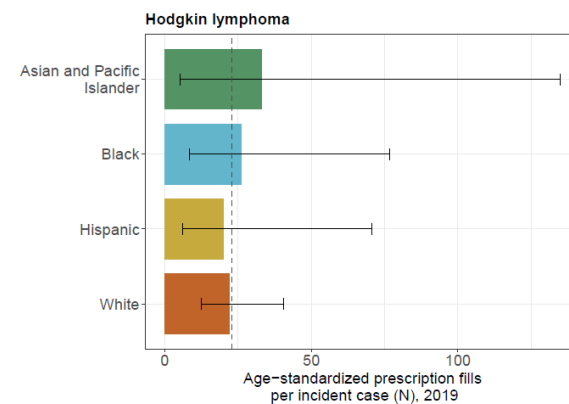
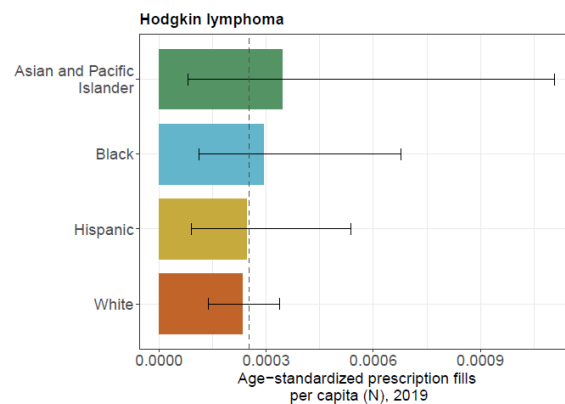
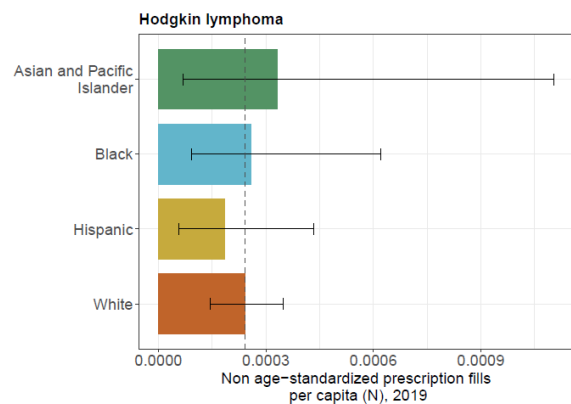
(40) Esophageal cancer



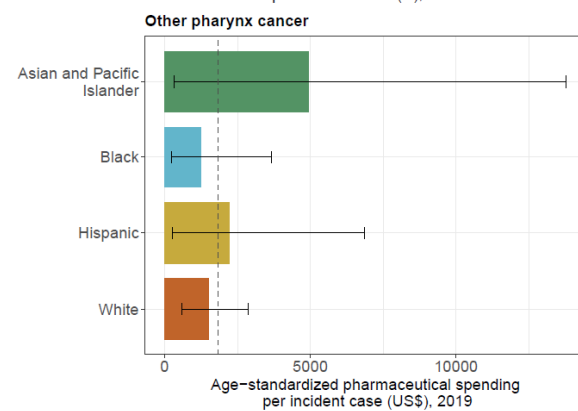
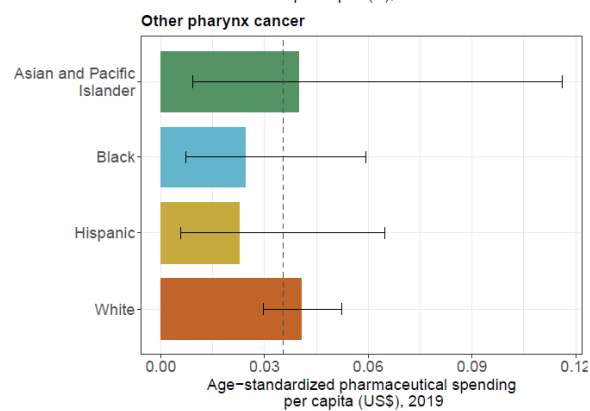
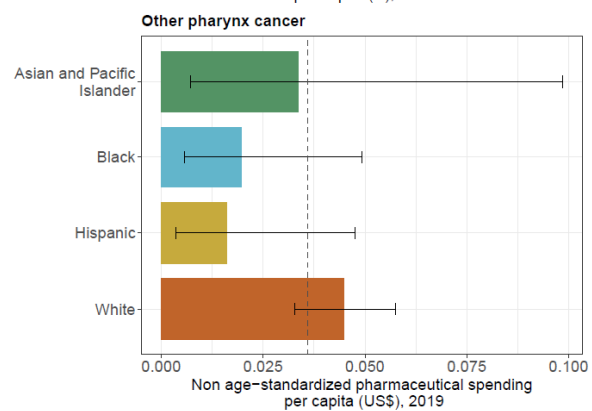
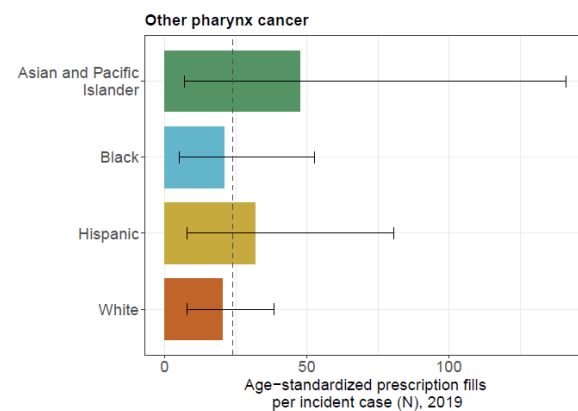
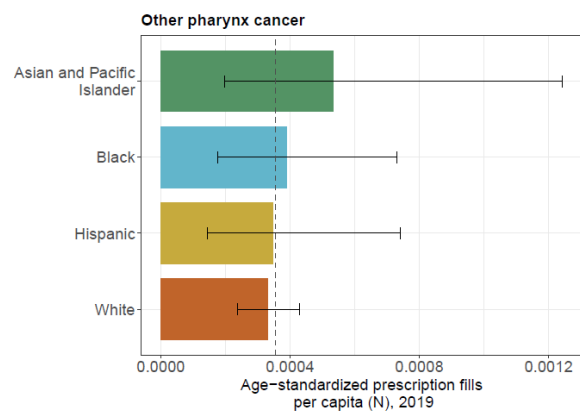
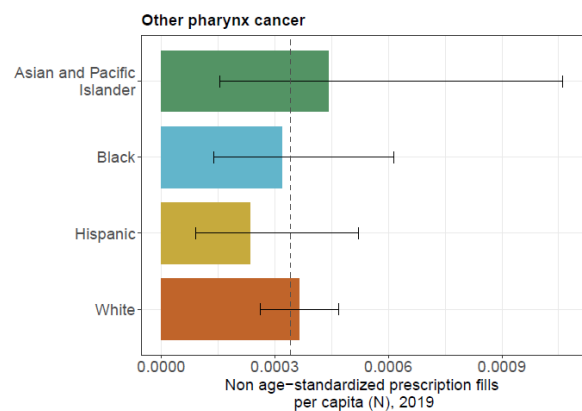
(41) Lip and oral cavity cancer



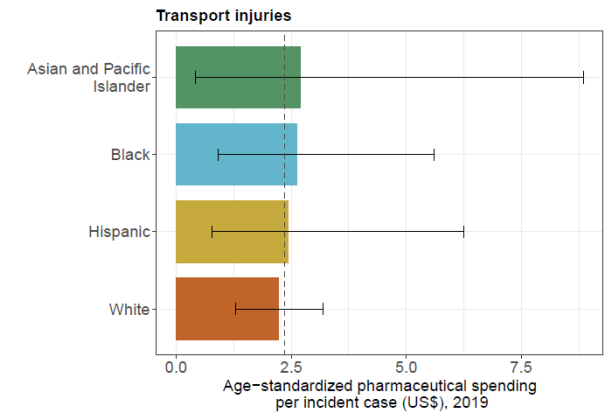
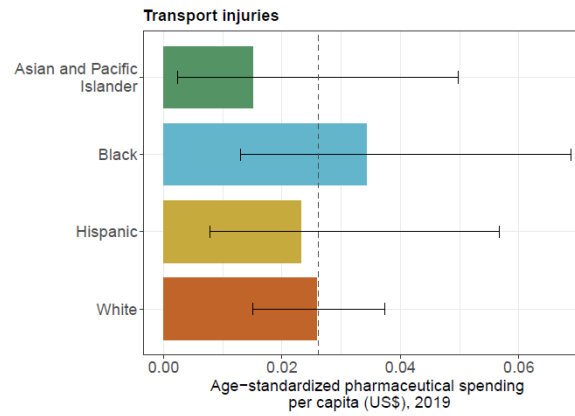
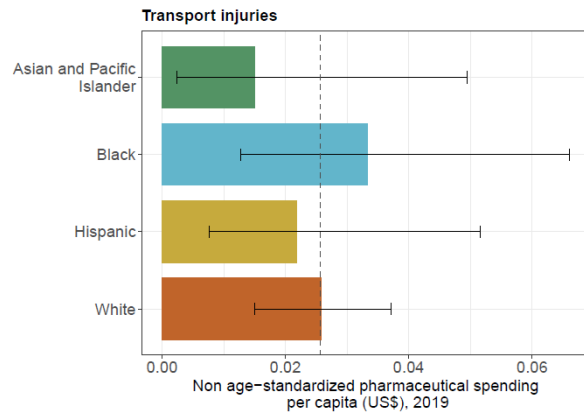
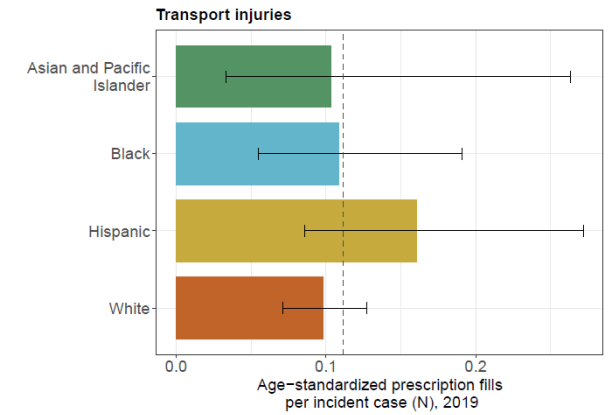
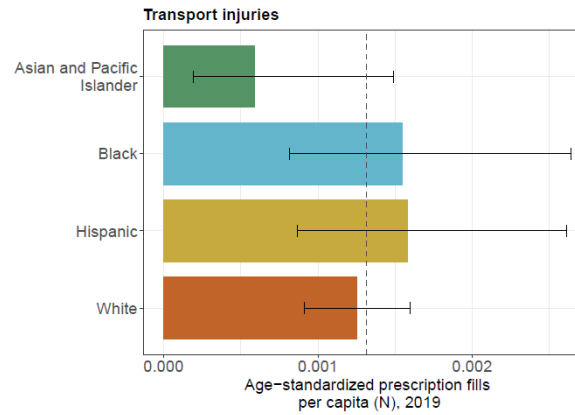
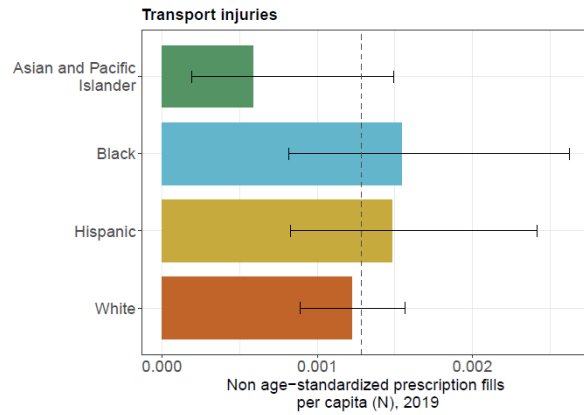
(42) Hodgkin lymphoma



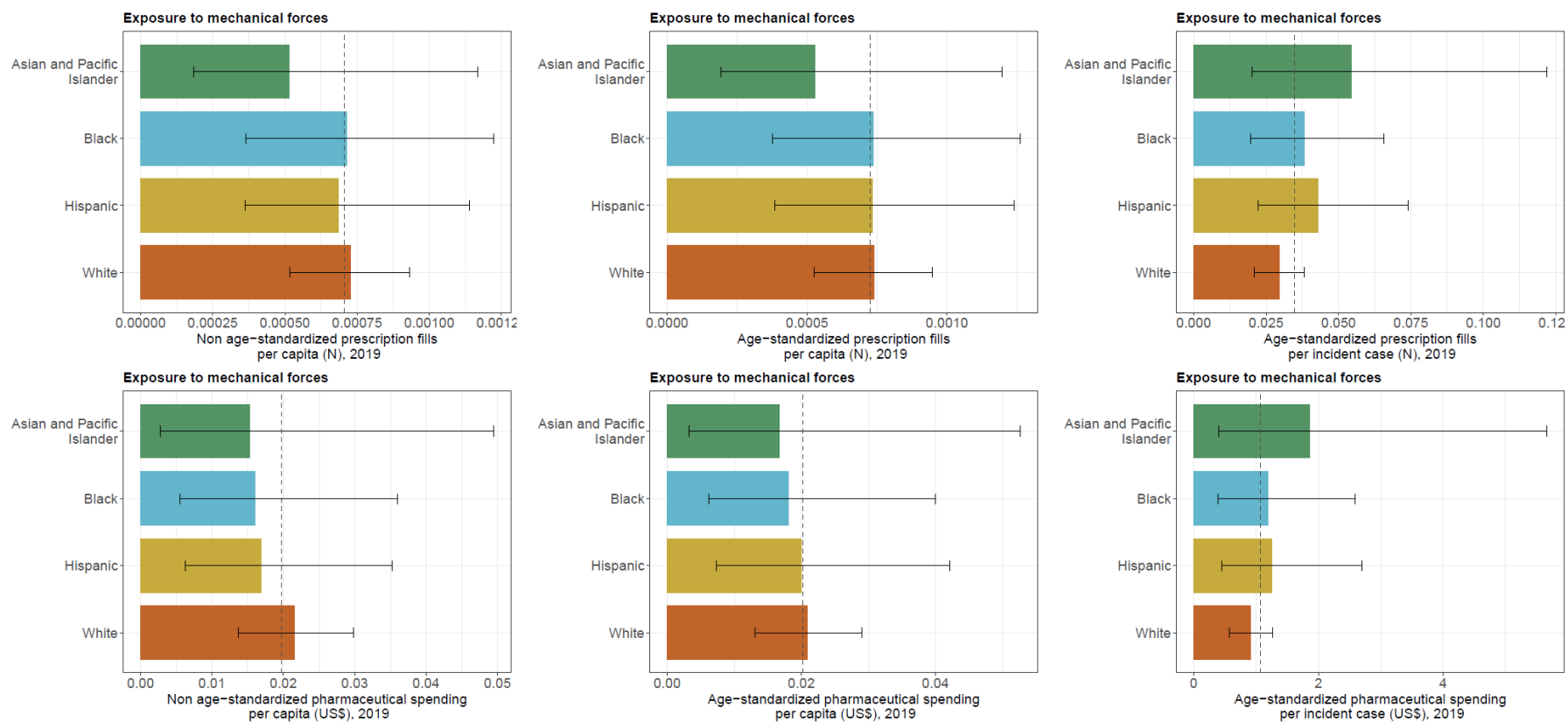
(43) Other pharynx cancer



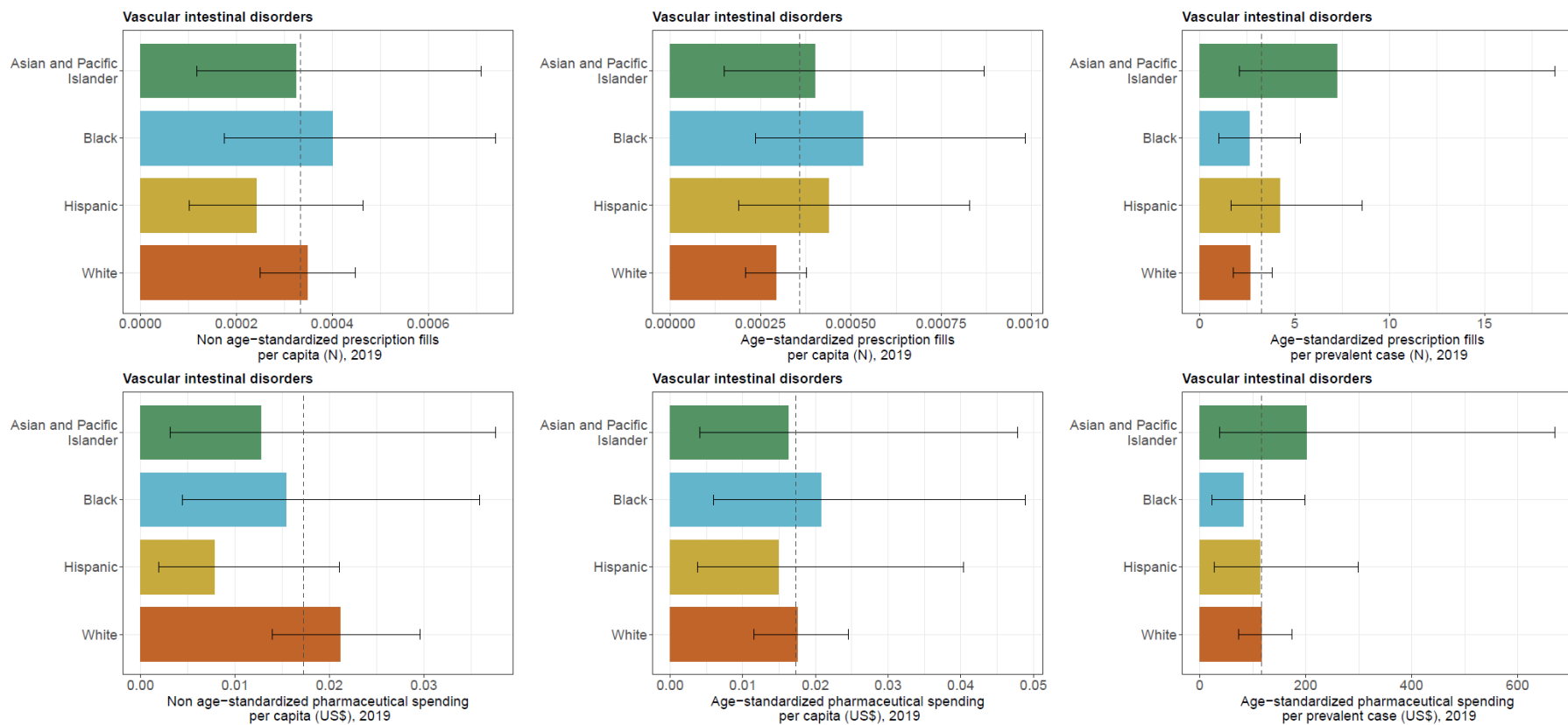
(44) Transport injuries



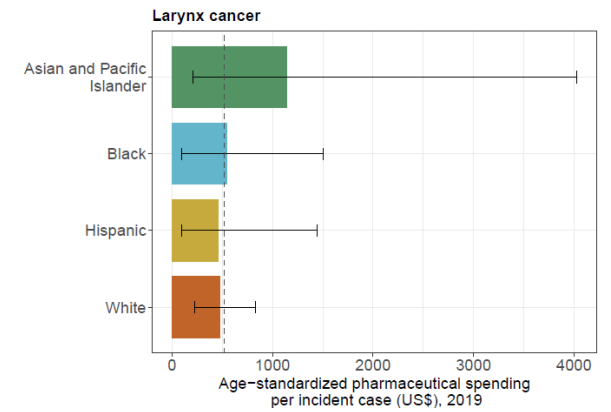
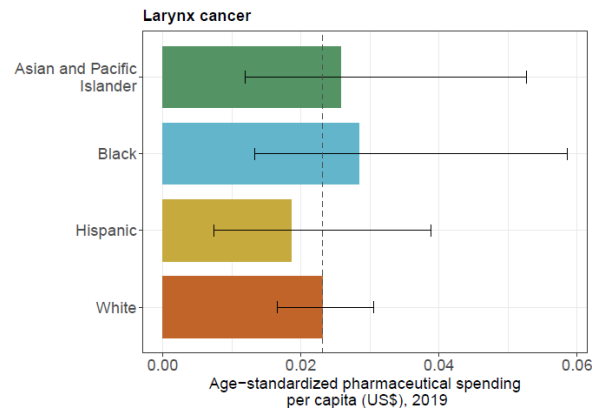
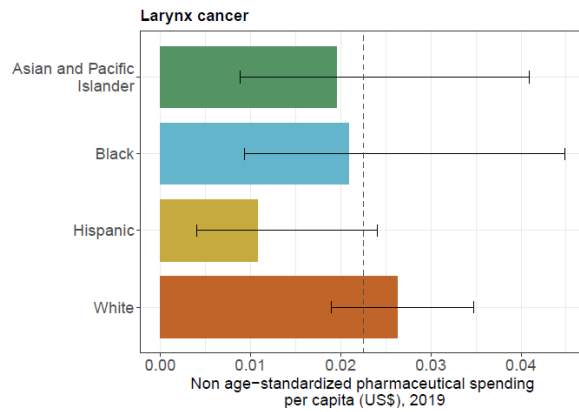
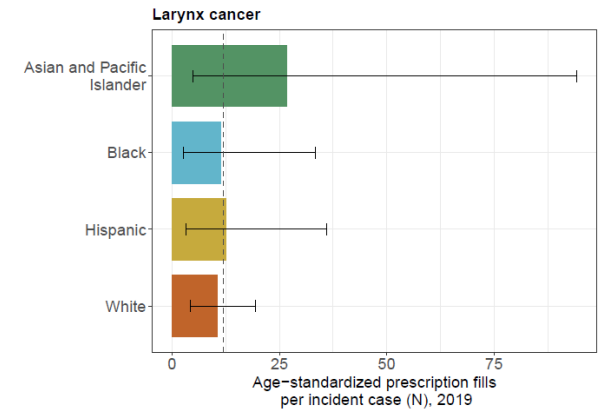
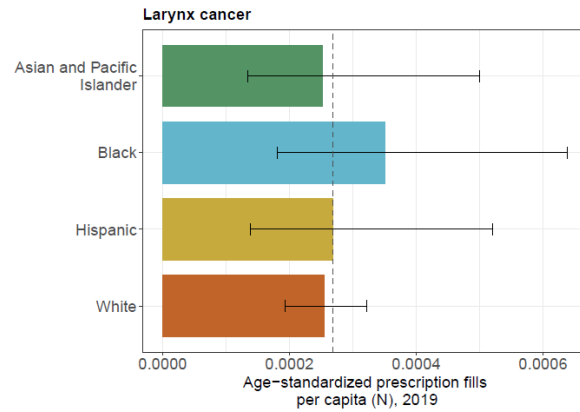
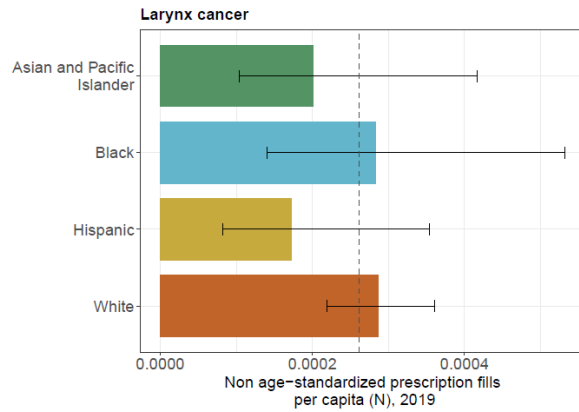
(45) Exposure to mechanical forces



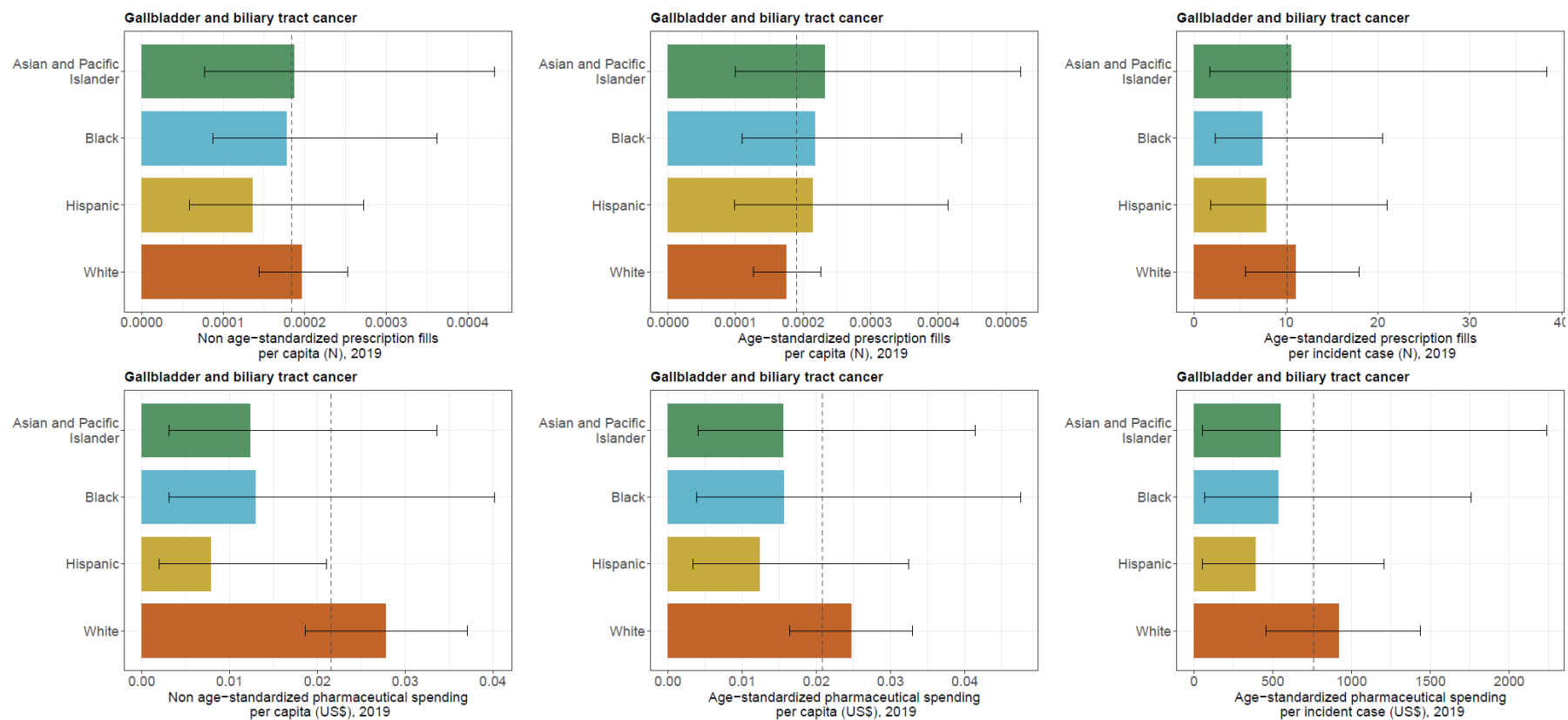
(46) Vascular intestinal disorders



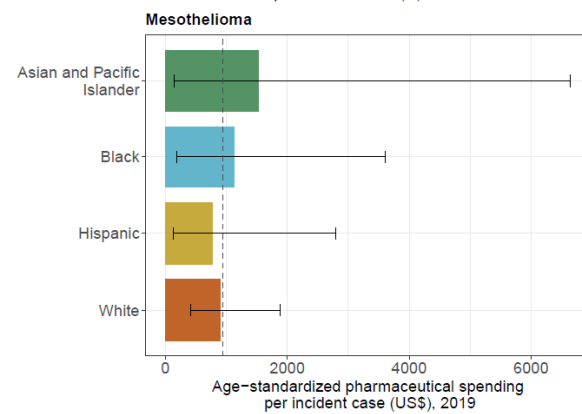
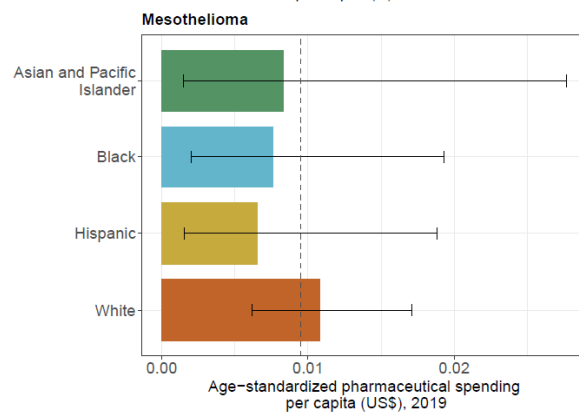
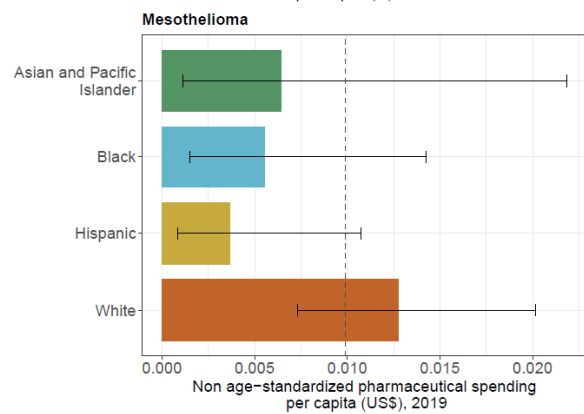
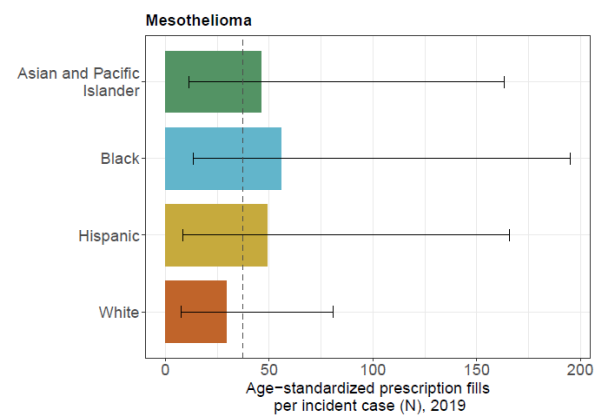
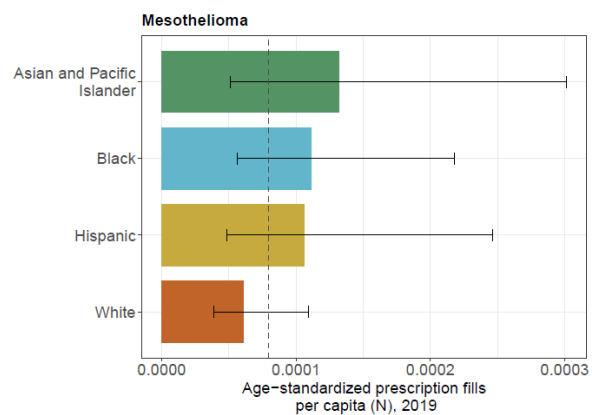
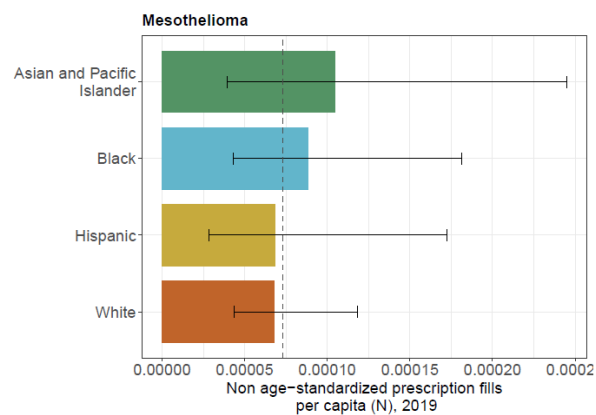
(47) Larynx cancer



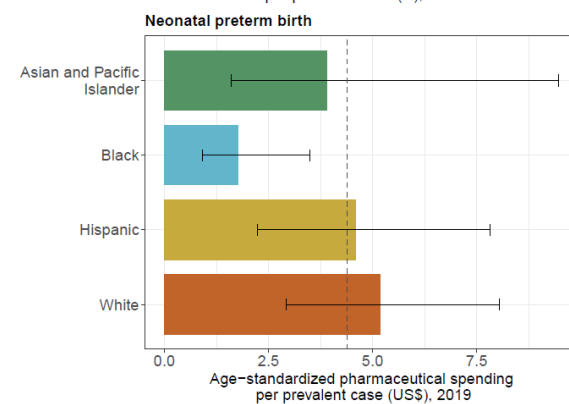
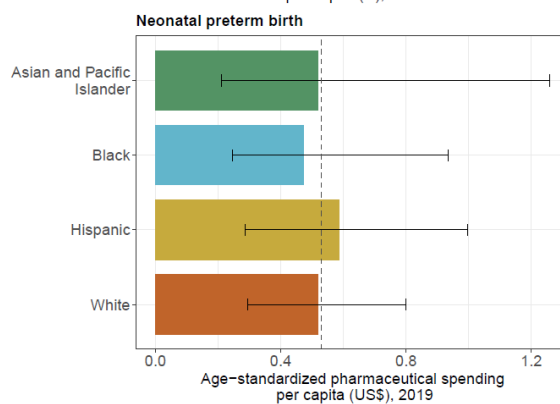
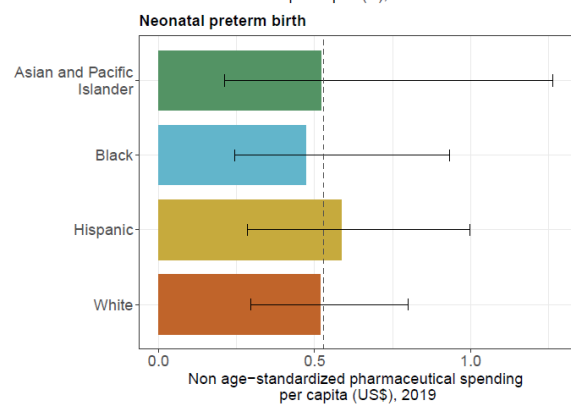
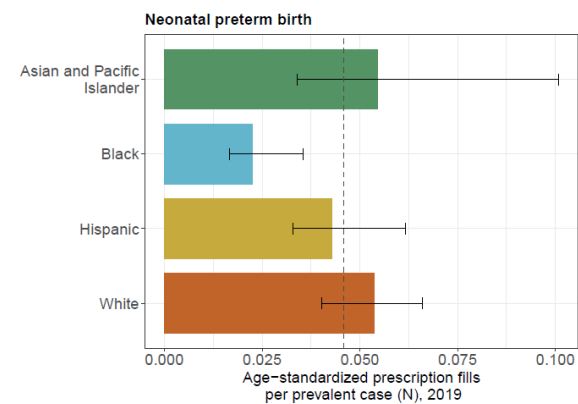
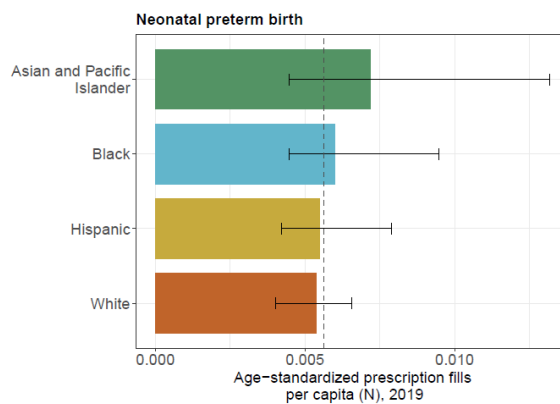
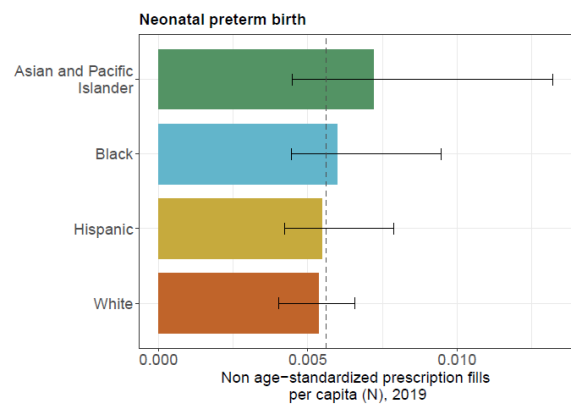
(48) Gallbladder and biliary tract cancer



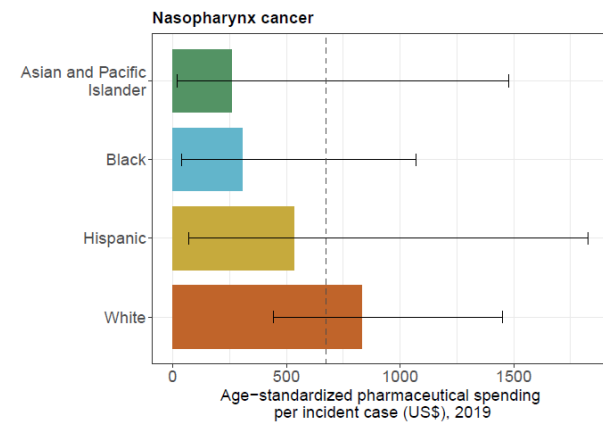
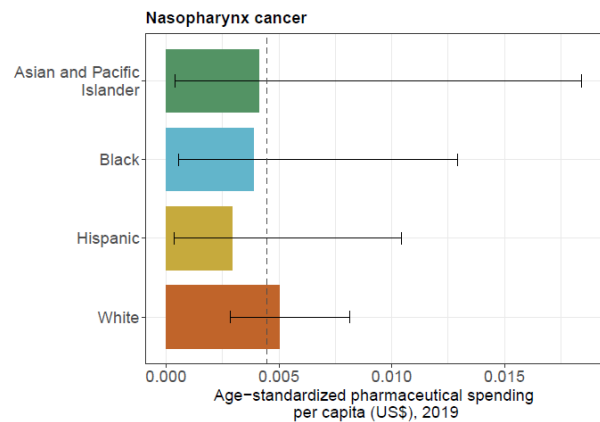
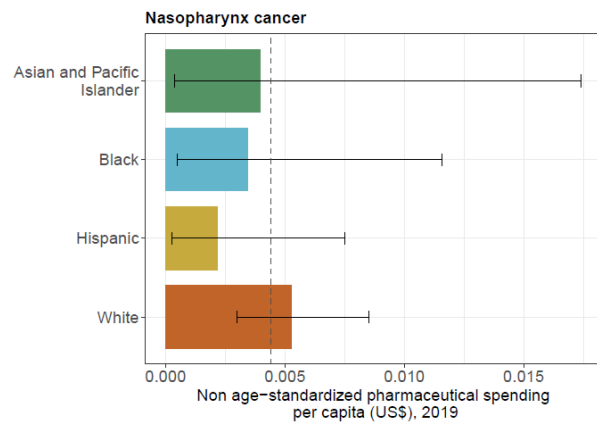
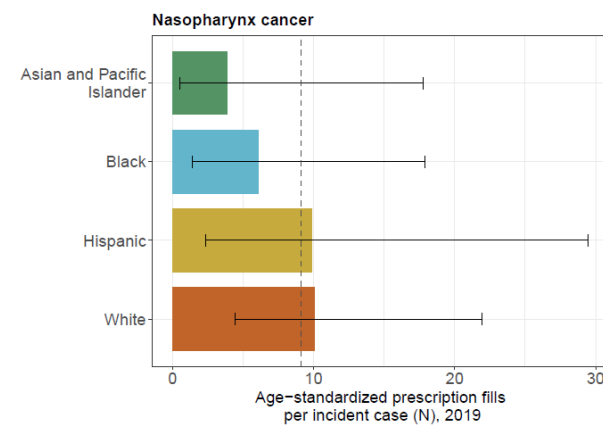
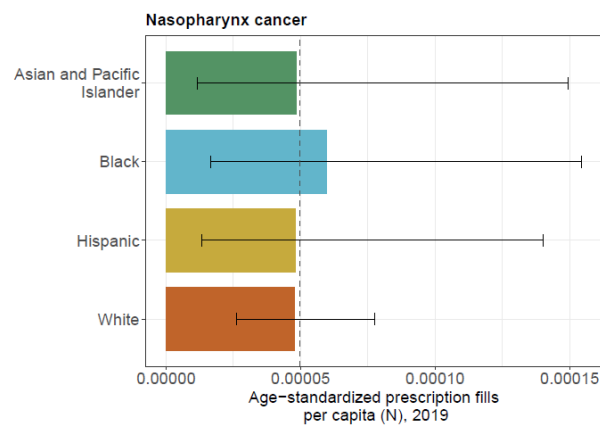
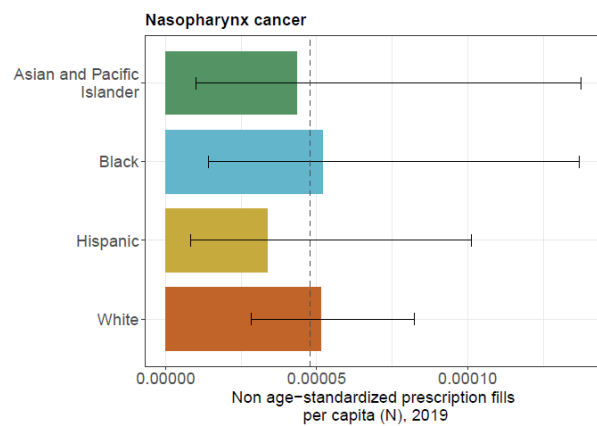
(49) Mesothelioma



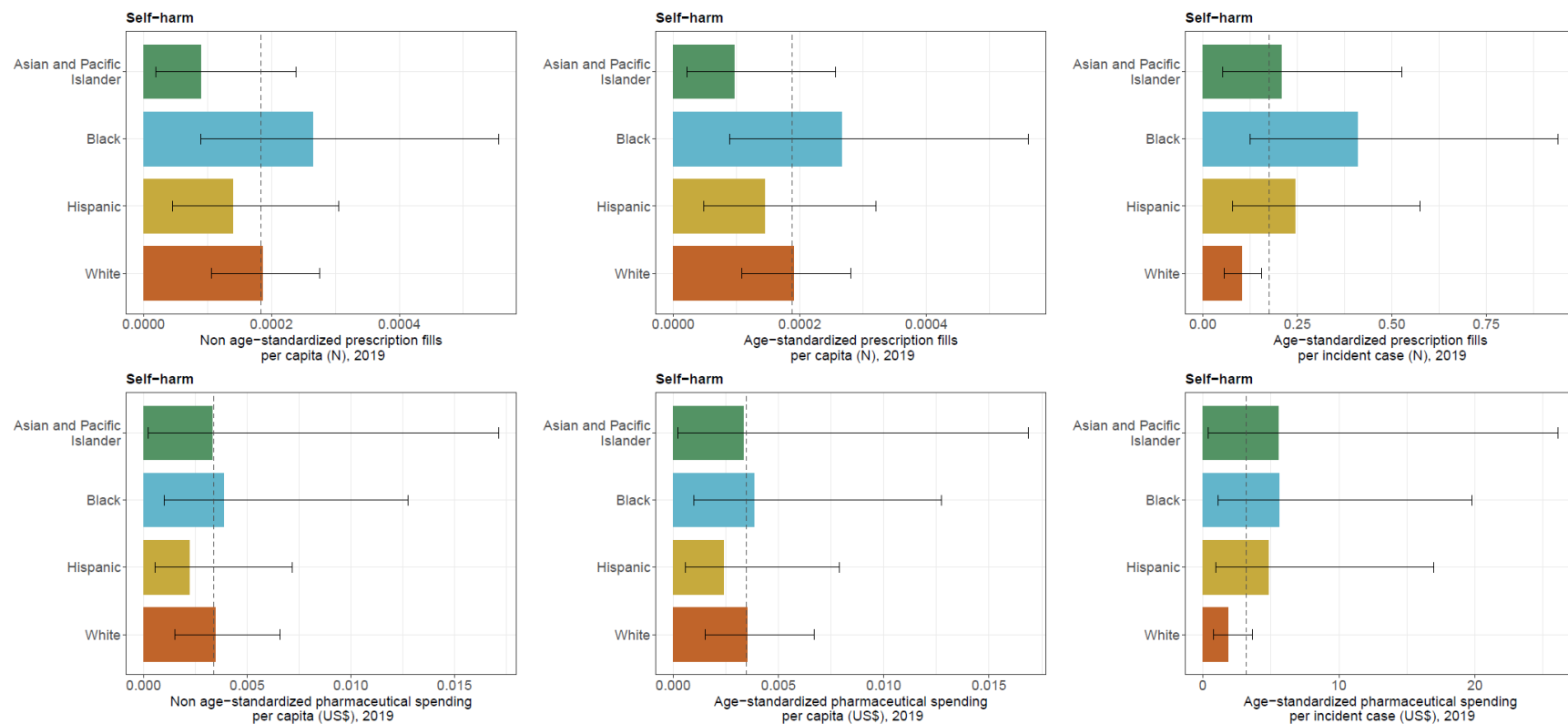
(50) Neonatal preterm birth



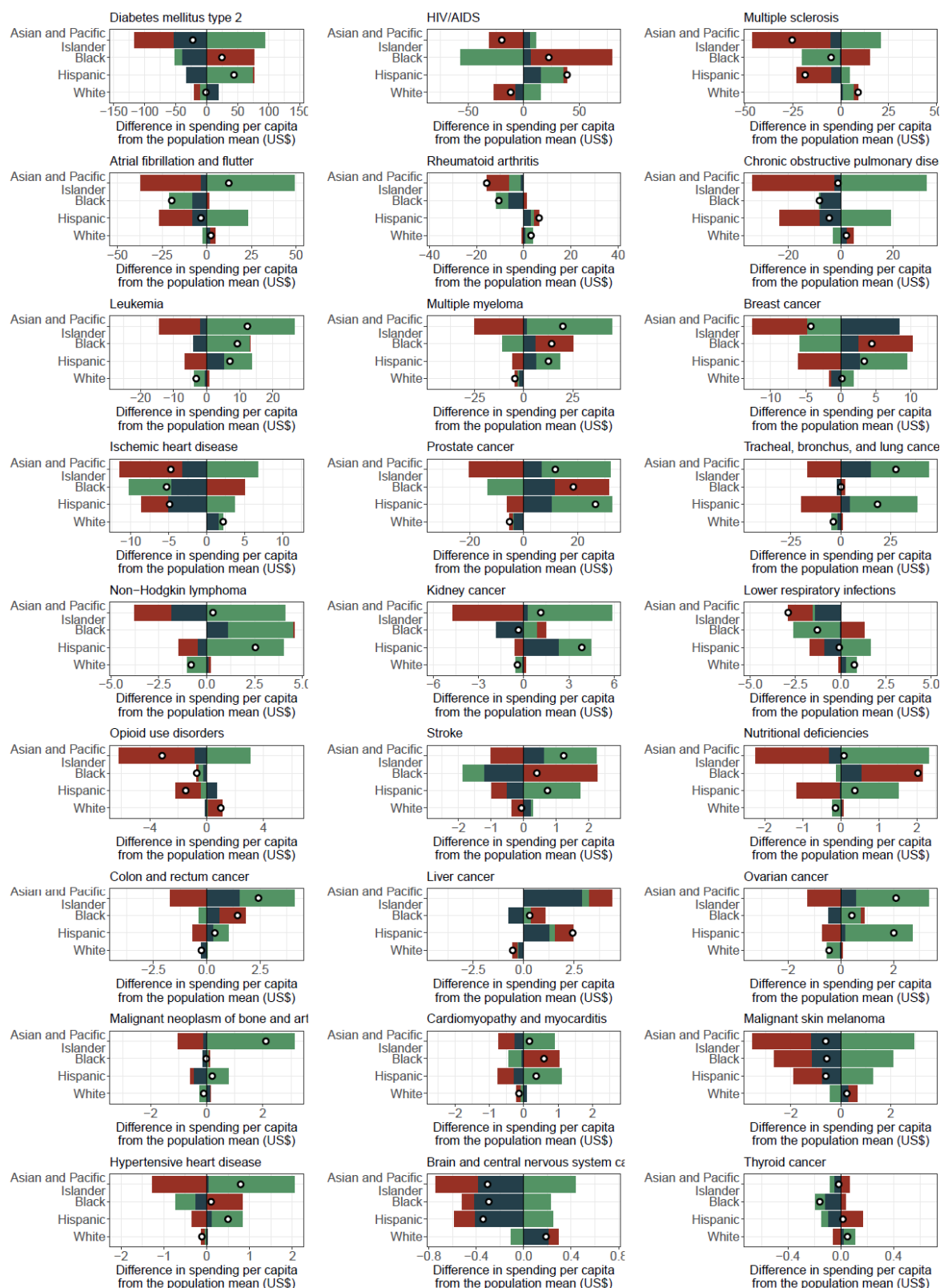
(51) Nasopharynx cancer

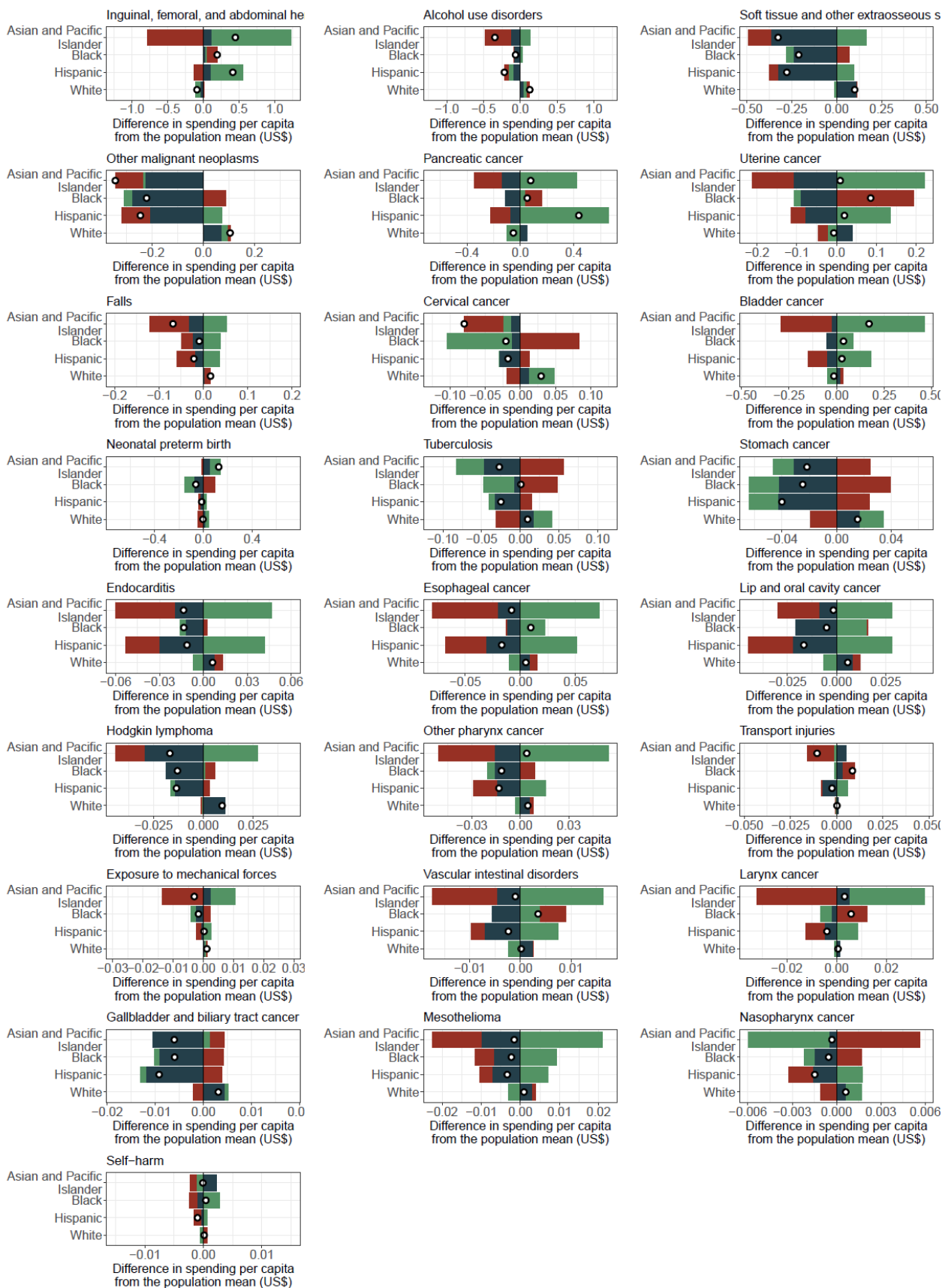


(52) Self-harm



D. Figure B3.2: Decomposition of pharmaceutical spending per capita, for 52 conditions





B4 STATE-LEVEL SUPPLEMENTAL RESULTS

A. Table B4.1: *Pharmaceutical spending and utilization per capita by race and ethnicity (age-standardized)*

For each US state and the District of Columbia, we present the total per-capita pharmaceutical spending and utilization by race/ethnicity. States are ranked by total (national) retail pharmaceutical spending. All groups other than Hispanic are non-Hispanic ethnicity. The mean values across 100 draws are presented in the top row, with the 95% Uncertainty Interval below.

State	Rank (all-population per-capita drug spending, age-std)	Rank (all-population per-capita prescription fills, age-std)	Spending per capita (US\$)				Prescription fills per capita (n)			
			Asian and Pacific Islander	Black	Hispanic	White	Asian and Pacific Islander	Black	Hispanic	White
Alaska	44	48	381.2	603.1	574.3	914.6	3.6	4.9	4.6	6.9
			(333.7, 451)	(519.6, 706.7)	(481.4, 698.2)	(893.9, 929.8)	(3.4, 3.9)	(4.5, 5.3)	(4.1, 5)	(6.6, 7.3)
			516.1	1081.7	605	1363.4	4.8	9.6	5.5	11.9
Alabama	8	2	(1003.5,			(1331.4,				(11.4,
			(406.4, 752.3)	1169.3)	(504.9, 792.2)	1392.5)	(4.3, 5.4)	(8.8, 10.2)	(5, 5.8)	12.4)
			461.3	928.5	3231.8	982.9	5.4	8.7	26.2	9.4
Arkansas	26	10	(847.9,		(2888.3,				(23.6,	
			(358, 650.8)	1029.2)	3581.5)	(955, 1005)	(4.9, 6)	(8, 9.3)	28.4)	(9, 9.8)
			725.6	848.4	767.8	976.1	5.6	7	6.4	7.3
Arizona	35	38	(948.2,							
			(633.5, 925)	(746.6, 997.2)	(705.6, 850.7)	1005.5)	(5, 6.1)	(6.5, 7.4)	(5.9, 6.9)	(7, 7.6)
			734.4	931.9	891.5	1103.9	5.2	6.3	6.2	6.8
California	36	47	(1048.4,							
			(672, 836.7)	(850.6, 1057)	(836.6, 965.7)	1159)	(4.6, 5.6)	(5.9, 6.7)	(5.8, 6.5)	(6.5, 7.1)
			480.6	599	779.5	796.3	3.8	5.1	6.2	6
Colorado	45	50	(408.9, 635.6)	(520.6, 687.6)	(715.8, 865.2)	(776.8, 813.1)	(3.4, 4.1)	(4.8, 5.4)	(5.7, 6.6)	(5.7, 6.2)
			1451.6	1181.7	1405.5	1405.3	10	8.3	10.2	9.1
			(1260,	(1086.9,	(1278.7,	(1370.5,				
Connecticut	5	11	1780.5)	1306.8)	1555.9)	1434.3)	(9.1, 11)	(7.7, 8.7)	(9.4, 11)	(8.7, 9.4)
			1250	1512.5	1420.3	2569	10.7	11	9.8	17.7
			(919.8,	(1379.8,	(1174.8,	(2350.5,		(10.4,		(16.5,
District of Columbia	1	1	1838.6)	1671.3)	1868.4)	2735.9)	(9.1, 12.3)	11.8)	(8.9, 10.7)	18.8)

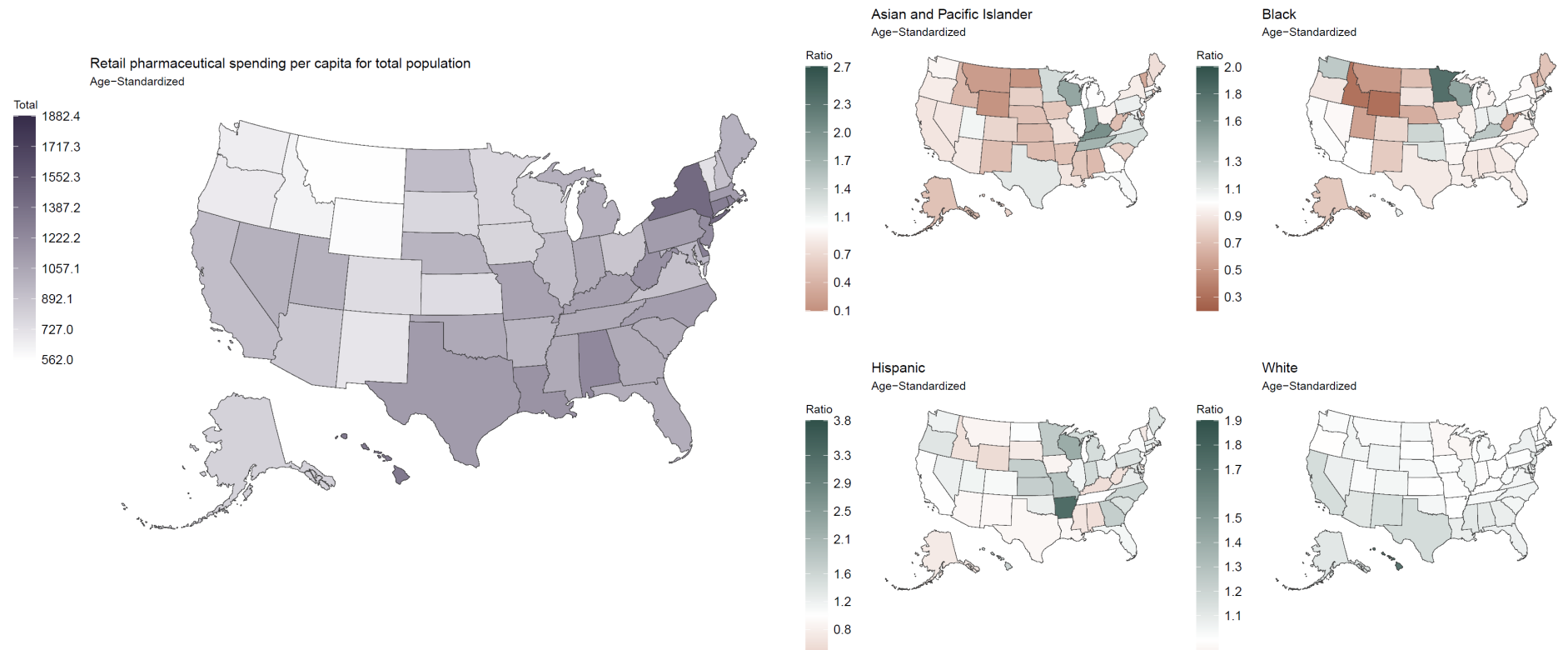
			488.3	1468.7	714	1413.3	3.4	9.9	5.5	8.4
Delaware	4	16	(1355.7, (409.4, 628.7)	(1597)	(627, 867.9)	(1378.4, 1451.7)	(3.1, 3.7)	(9.1, 10.5)	(5.2, 5.7)	(8, 8.8)
			1074	943.9	1105.4	1108.6	7.4	7	7.9	7.4
Florida	15	24	(859.1, (953, 1345)	(1048.7)	(1026.9, 1213.7)	(1064.8, 1151.8)	(6.6, 8.1)	(6.4, 7.3)	(7.3, 8.4)	(7.1, 7.6)
			1029.3	926.1	1674.4	1147.4	7.6	7.4	12	8.5
Georgia	27	30	(905.5, 1311.1)	(847.7, 1006.8)	(1483.7, 1898.6)	(1093.8, 1197.1)	(6.9, 8.4)	(6.9, 7.9)	(10.8, 13.1)	(8.1, 8.9)
			877.9	1633.7	2166.6	2316.2	6.1	11.8	14.6	13.6
Hawaii	3	12	(1351.7, (791, 955.5)	(2088.6)	(1829.6, 2579.3)	(2179, 2498.3)	(5.6, 6.6)	(10.3, 13.1)	(13.1, 15.9)	(12.2, 14.7)
Iowa	40	23	405.9	592.8	590.2	805.7	4.2	6.2	5.9	7.8
			(327.9, 620.4)	(533.2, 676.4)	(512.1, 715.4)	(794.8, 812.6)	(3.8, 4.5)	(5.8, 6.5)	(5.6, 6.2)	(7.4, 8.1)
Idaho	49	45	236	199.6	361.8	663.7	2.2	2.2	3.9	6.2
			(178.2, 346.3)	(177.5, 219.4)	(307.9, 443.6)	(653.7, 670.5)	(2, 2.4)	(2, 2.4)	(3.7, 4.1)	(5.8, 6.5)
			942.8	859.5	1019.6	990.6	7.1	6.8	7.2	7
Illinois	31	40	(833, 1149.9)	(785.2, 961.6)	(920.1, 1133.3)	(952.3, 1018.8)	(6.3, 7.8)	(6.3, 7.1)	(6.7, 7.7)	(6.7, 7.2)
			1944.7	1086.9	1498.7	1033.1	12.9	8.1	10.9	7.7
Indiana	22	26	(1667.6, 2450.8)	(999.6, 1196.5)	(1333.8, 1702.8)	(1013.8, 1048.1)	(11.4, 14.5)	(7.4, 8.6)	(9.8, 11.8)	(7.3, 8.1)
			319.9	891	1257.9	751.6	3.1	8.4	11.9	6.8
Kansas	43	35	(811.2, (262.9, 428.1)	(1006.7)	(1134.2, 1398.1)	(739.9, 761)	(2.9, 3.4)	(7.6, 8.9)	(10.8, 12.8)	(6.4, 7)
			2287.8	1378.3	518.9	1104	17.5	11.9	5	10.2
Kentucky	16	5	(1952.8, 2881.8)	(1259.9, 1533.7)	(435.7, 636.2)	(1088.1, 1115.5)	(15.3, 19.9)	(10.8, 12.7)	(4.7, 5.3)	(9.7, 10.7)
			954.3	1083.5	956.9	1277.8	8.5	9.7	7.9	10.7
Louisiana	11	6	(802.1, 1213.1)	(1006.7, 1165.5)	(852.5, 1113)	(1237.7, 1315.7)	(7.8, 9.2)	(8.9, 10.2)	(7.4, 8.4)	(10.3, 11.2)
			1024.7	1169	1227.5	1151.2	8	9	9.6	8.2
Massachusetts	12	19	(894.6, 1304.4)	(1049.1, 1330.4)	(1093.9, 1416.1)	(1114.8, 1179.5)	(7.3, 8.7)	(8.4, 9.4)	(8.9, 10.3)	(7.8, 8.5)
Maryland	29	44	1037.8	870.9	1181.3	1069.2	6.8	5.6	7.6	6.2

			(934, 1236.2)	(804.7, 949.7)	(1063.1, 1364.4)	(1016.7, 1108.7)	(6.1, 7.5)	(5.2, 5.8)	(7, 8.1)	(6, 6.5)
			588.5	682.4	964.7	999.2	4.8	5.3	6.3	8.4
Maine	14	7	(406.6, 900.5)	(578.7, 988.6)	(710.7, 1367.8)	(990.2, 1005.2)	(4.3, 5.3)	(5, 5.6)	(5.8, 6.7)	(8, 8.7)
			1002.8	939.7	1393.2	1027.2	7.3	7.7	10	7.8
Michigan	21	22	(868.7, 1284.4)	(861.8, 1044.9)	(1219.8, 1637.5)	(1007.2, 1043.1)	(6.4, 8)	(7.1, 8.1)	(9.1, 10.7)	(7.4, 8.1)
			1038.8	1349.7	1207	767.8	8.5	11.4	9.5	6.3
Minnesota	41	43	(909.1, 1329.6)	(1228.1, 1511.5)	(1019.2, 1469.4)	(744.1, 781.2)	(7.5, 9.3)	(10.3, 12)	(8.6, 10.3)	(6, 6.6)
			930.4	1059.8	1856.3	1097.1	7.6	8.9	13.4	8.9
Missouri	17	13	(785.2, 1319.3)	(968.3, 1193.8)	(1591.4, 2204.2)	(1073.1, 1115.6)	(6.9, 8.4)	(8.1, 9.4)	(12, 14.6)	(8.4, 9.2)
			536	905.6	578.9	1166.7	4.4	8.6	5	10.4
Mississippi	20	9	(432.8, 781.4)	(851.3, 965.3)	(1130.4, 480.4, 742.6)	1199.5)	(4.1, 4.9)	(8, 9)	(4.7, 5.3)	(9.9, 10.8)
			102.2	256	387.6	581	1.2	2.4	3.2	5.5
Montana	50	46	(88.2, 113)	(228.9, 280.5)	(310.5, 556.7)	(575.7, 585.8)	(1.1, 1.3)	(2.2, 2.6)	(2.9, 3.4)	(5.1, 5.7)
			1475.8	993.6	1669.4	1195.9	10.7	7.7	12.5	8.9
North Carolina	13	18	(1314.3, 1863.6)	(906.3, 1112.4)	(1468.5, 1899.5)	(1158, 1229.6)	(9.5, 11.8)	(7.1, 8.1)	(11.4, 13.5)	(8.5, 9.3)
			120.1	628.4	807.8	962.4	1.5	9.1	6.5	8.6
North Dakota	34	20	(111.3, 133.8)	(530.5, 759.2)	(639, 1104.2)	(948.5, 972.4)	(1.4, 1.7)	(8.5, 9.6)	(5.7, 7.1)	(8.1, 9)
			453	588.7	1630.3	1012.2	3.8	5.1	13.1	8.1
Nebraska	30	25	(350.5, 653.6)	(523.7, 656.2)	(1446.8, 1796.7)	(996.8, 1025.6)	(3.5, 4.2)	(4.8, 5.4)	(11.6, 14.3)	(7.6, 8.4)
New Hampshire	28	31	498.2	575.9	698.5	930.5	3.2	4	4.2	6.9
			(371.2, 712.6)	(487.1, 763.6)	(547, 954.1)	(917.9, 938.7)	(2.8, 3.6)	(3.8, 4.2)	(3.9, 4.5)	(6.6, 7.1)
			1211.6	1155.6	1134.3	1358	8	7.6	7.5	7.8
New Jersey	9	29	(1085.5, 1429.9)	(1061.6, 1292.4)	(1045.9, 1236.9)	(1312.6, 1402)	(7.2, 8.7)	(7.1, 8)	(7, 8)	(7.5, 8.1)
			337.4	514.9	640.6	813.7	3.4	5.2	6.3	7.4
New Mexico	46	34	(281.1, 468.4)	(452.7, 577.6)	(607.3, 675.7)	(787.2, 839.2)	(3.1, 3.7)	(4.9, 5.5)	(5.8, 6.7)	(7, 7.6)
Nevada	25	28	817.3	961.2	1152.5	1092	6.3	8.1	9.2	8.2

			(865.4, 1117.2)	(1061.3, 1264.4)	(1040.5, 1137.5)	(5.6, 6.8)	(7.6, 8.5)	(8.6, 9.7)	(7.9, 8.5)
			1215.8	1386.9	1428.8	1545.9	9.5	9.6	10.1
			(1126.1, 1399.7)	(1295.6, 1530.2)	(1332.5, 1553)	(1496.5, 1584)	(8.8, 10.1)	(9, 10.1)	(9.5, 10.7)
New York	2	8	810.2	934.6	975.6	891.3	6.5	7.9	7.8
			(696.1, 1053.2)	(858.9, 1041.2)	(849.9, 1142.5)	(875.3, 903.1)	(5.9, 7.2)	(7.3, 8.3)	(7.1, 8.5)
Ohio	32	27	470.3	1184.2	1311.3	1092.7	4.6	9.5	10.5
			(389.2, 634.9)	(1081.7, 1307.5)	(1173, 1481.5)	(1071.4, 1113.8)	(4.2, 5)	(8.6, 10.1)	(9.6, 11.3)
Oklahoma	24	21	547.8	576.2	890.3	676.1	5.1	5.4	8.2
					(806.8, 1016.1)	(661.8, 685.6)	(4.6, 5.6)	(5, 5.7)	(7.6, 8.7)
Oregon	47	42	1300.4	1076.2	1424.9	1137.9	9.6	8.3	10.7
			(1147.7, 1631.5)	(980.4, 1204.2)	(1267.4, 1678.2)	(1114.4, 1156.1)	(8.6, 10.6)	(7.7, 8.7)	(9.9, 11.5)
Pennsylvania	10	14	830.9	884.9	964.4	1359.4	6.9	7.6	8.6
			(703.3, 1099.5)	(794.8, 1005.3)	(873.6, 1121.4)	(1329.1, 1379.9)	(6.3, 7.5)	(7.2, 8.1)	(8.1, 9.1)
Rhode Island	6	4	540.4	1003.4	1436.5	1085.6	3.9	7.9	11.1
			(928.8, 1107.2)	(1234, 1659.5)	(1055.2, 1116.8)	(3.5, 4.3)	(7.2, 8.3)	(10, 12.2)	(8.3, 9)
South Carolina	18	17	411.1	624	485.8	795.8	3.4	13.4	5
			(327.3, 578.6)	(537.5, 732.1)	(397.3, 660.2)	(782.9, 806.4)	(3, 3.6)	(12.1, 14.6)	(4.8, 5.3)
South Dakota	42	41	1836.3	961.3	874.5	1074.8	13.4	8.1	7
			(1568.3, 2349.6)	(868.6, 1089.5)	(744.5, 1104.7)	(1045.2, 1097.1)	(11.7, 15.2)	(7.4, 8.6)	(6.3, 7.6)
Tennessee	19	15	1355.9	1011.4	986.2	1337.7	9.9	7.6	7.3
			(1237.1, 1600)	(914.5, 1125.4)	(927.5, 1065.5)	(1284.1, 1387.3)	(8.8, 10.9)	(7, 8)	(6.9, 7.7)
Texas	23	32	1098.8	537.3	1142.9	1024.4	8.3	4.8	9.2
			(929.4, 1411.4)	(455.6, 672)	(998.4, 1306.4)	(995, 1045.5)	(7.3, 9.2)	(4.5, 5.1)	(8.4, 10)
Utah	38	39	1096.6	843.5	1065	954.5	8.6	6.5	7.9
Virginia	33	37							7

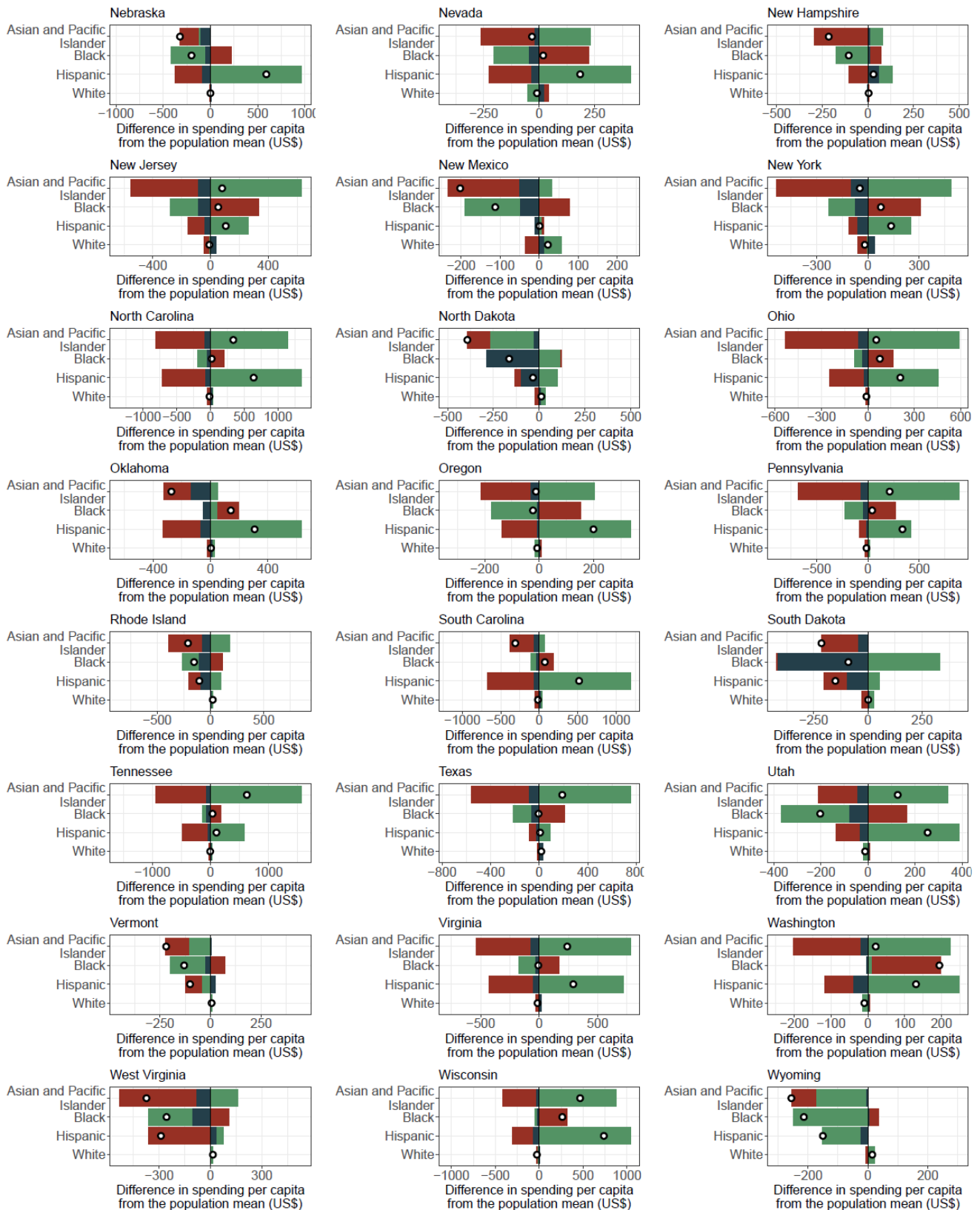
			(994.5, 1290.7)	(783.6, 913.8)	(957, 1235.6)	(923.7, 978.1)	(7.7, 9.5)	(6.1, 6.8)	(7.1, 8.6)	(6.8, 7.3)
			209.7	446.9	401.2	748.6	2.2	4.2	3.1	6.5
Vermont	39	33	(193.5, 230.5)	(401.5, 511.9)	(312, 559.3)	(744.6, 751.5)	(2, 2.3)	(3.9, 4.4)	(2.8, 3.3)	(6.1, 6.8)
			607.9	851.4	733.1	679.4	5.3	7	6.5	5.6
Washington	48	51	(535.6, 725.3)	(742.9, 958.3)	(635.6, 858.5)	(658.2, 695.5)	(4.7, 5.8)	(6.4, 7.3)	(6, 7.1)	(5.3, 5.8)
			1437.1	1204.7	1745.1	782.8	10.3	9.2	13.7	6.4
			(1237.7, 1845.5)	(1093.9, 1340.3)	(1504.9, 2009.9)	(766.9, 796.1)	(9.1, 11.4)	(8.4, 9.7)	(12.3, 14.9)	(6.1, 6.7)
Wisconsin	37	36	572.2	668.8	558.7	1212	4.1	6.2	4.1	10.2
						(1205.2, 1216.4)				
West Virginia	7	3	(438.3, 944.5)	(588.7, 807.6)	(449.5, 820.4)	1216.4)	(3.7, 4.6)	(5.8, 6.6)	(3.8, 4.4)	(9.7, 10.5)
			69.9	165	283.9	618.2	0.9	1.5	2.6	5.7
Wyoming	51	49	(63.8, 75.7)	(151.2, 182.5)	(248.2, 332.4)	(614.4, 621.2)	(0.7, 1)	(1.3, 1.7)	(2.4, 2.8)	(5.3, 6)

B. Figure B4.1: State maps of pharmaceutical spending per capita (age-standardized), 2019



C. Figure B4.2: State-level decomposition (aggregated across 52 health conditions)





Appendix C: Supplementary Methods for Chapter 4

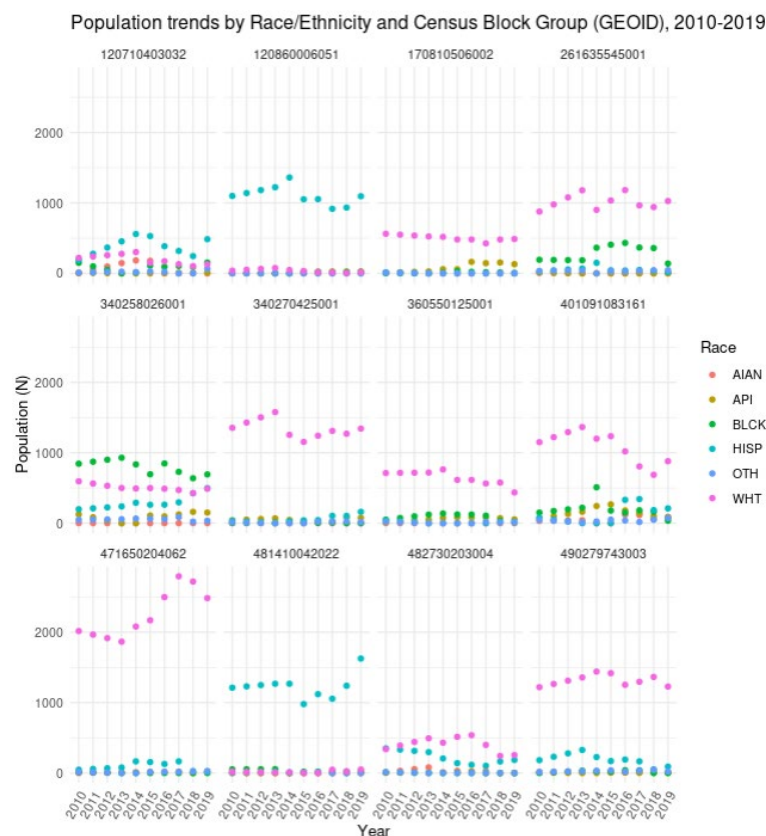
Distance metric

Census Population

For 2013–2019, we used 5-year rolling American Community Survey (ACS) estimates of population by race and ethnicity at the census block group (CBG) level. For 2010, we used data from the decennial census. For 2011 and 2012 (years for which neither data source was available) we linearly interpolated population estimates separately for each racial and ethnic group, as well as for the total population. Across the full study period (2010–2019), our analytic universe included 217,812 unique CBGs. Of these, 837 (0.34%) had no recorded population and were assumed to have zero population in those years.

We classified race and ethnicity into six mutually exclusive categories: Hispanic (any race); and among non-Hispanic individuals, White alone, Black alone, American Indian or Alaska Native (AIAN) alone, Asian or Pacific Islander (API) alone (including both Asian and Native Hawaiian/Pacific Islander), and “Other” (other single race or multiple races). There were no missing values for race or ethnicity; by design, the sum of the six categories equaled the total population for each CBG-year.

Appendix Figure C.1: Population sizes for random sample of 12 CBGs, with interpolated values for 2011 and 2012



**This figure demonstrates that the interpolation works as expected.*

Isochrones

We used Mapbox to generate 10- and 20-minute walk- and drive-time isochrones surrounding each pharmacy. Coordinates were measured to five decimal places, corresponding to approximately 1.1 meters of spatial precision. To ensure capture of locations in close proximity to isochrone boundaries, we applied a 250-meter buffer to each polygon. Travel time estimates did not account for traffic conditions and instead reflect low-traffic (e.g., nighttime) driving conditions.

Centroids

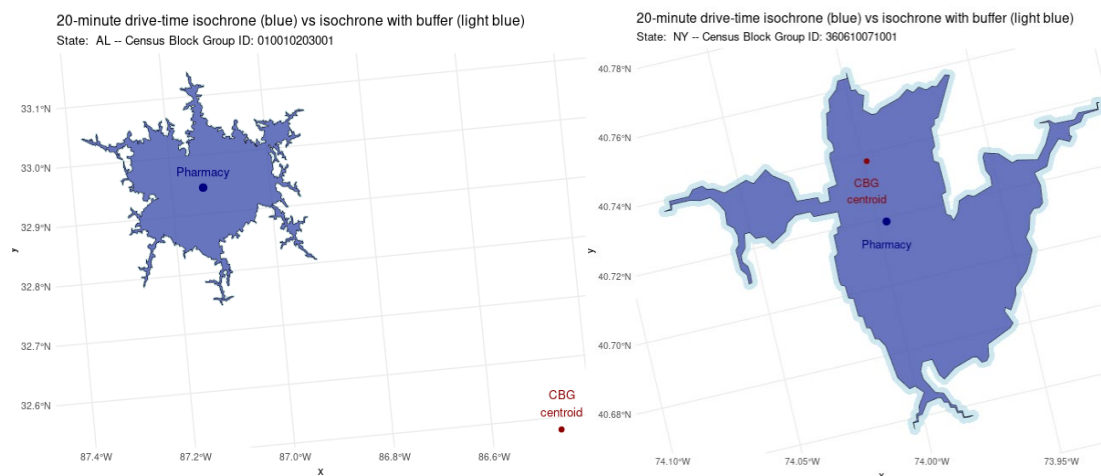
We obtained the CBG boundaries for 2010-2019 through the National Historical Geographic Information Systems (NHGIS).²⁰³ Centroids were calculated separately for each year. First, we convert to a projected coordinate reference system (CRS), which is recommended for finding centroids. Specifically, we used the Albers Equal Area projection (ESPG 5070) for use in the contiguous US; the Alaska Albers Equal Area Conic (ESPG 3338) for Alaska, and the Hawaii Albers Equal Area Conic (ESPG 3563) for Hawaii. We then obtained the centroid of the CBG. We then converted back to World Geodetic System 1984 (ESPG 4236) for mapping coordinates.

Assessing overlap between isochrone and centroids

For each CBG-year, we assessed whether the centroid of a CBG fell within the boundaries of the pharmacy drive-time and walk-time isochrones. CBG-level populations, disaggregated by race and ethnicity, were assigned to a given distance range if the corresponding centroid intersected the isochrone. All spatial intersections were assessed using the EPSG:4326 CRS.

Appendix Figure C.2: 20-minute drive-time isochrone and buffer for a single pharmacy

Alabama vs. Midtown Manhattan



Aggregation

Finally, we aggregated results to the county and state levels by calculating the percentage of the population within 10- or 20-minute walk or drive to a pharmacy, stratified by race and ethnicity.

References

1. Essien U, Dusetzina SB, Matthews K, Mallick A, Qato D. Achieving Pharmaco-equity - a Policy Prescription for Reducing Health Disparities. Policy Roundtable presented at: AcademyHealth Annual Research Meeting; June 7, 2022; Washington, D.C.
<https://academyhealth.confex.com/academyhealth/2022arm/meetingapp.cgi/Session/30296>
2. Duwamish Tribe. History of the Duwamish People. Accessed June 23, 2025.
<https://www.duwamishtribe.org/history>
3. Duwamish Tribe. Land Acknowledgement Guidance. Accessed June 23, 2025.
<https://www.duwamishtribe.org/land-acknowledgement>
4. Gee GC, Ro A, Shariff-Marco S, Chae D. Racial Discrimination and Health Among Asian Americans: Evidence, Assessment, and Directions for Future Research. *Epidemiologic Reviews*. 2009;31(1):130-151. doi:10.1093/epirev/mxp009
5. Lett E, Adekunle D, McMurray P, et al. Health Equity Tourism: Ravaging the Justice Landscape. *J Med Syst*. 2022;46(3):17. doi:10.1007/s10916-022-01803-5
6. Herman AA. Toward a Conceptualization of Race in Epidemiologic Research. *Ethnicity & Disease, Inc*. 1996;6(1/2):7-20.
7. Senior PA, Bhopal R. Ethnicity as a variable in epidemiological research. *BMJ*. 1994;309(6950):327-330. doi:10.1136/bmj.309.6950.327
8. Kaufman JS. Commentary: Considerations for Use of Racial/Ethnic Classification in Etiologic Research. *American Journal of Epidemiology*. 2001;154(4):291-298. doi:10.1093/aje/154.4.291
9. Office of Management and Budget. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. Published online October 30, 1997.
https://obamawhitehouse.archives.gov/omb/fedreg_1997standards
10. Braveman P. Health Disparities and Health Equity: Concepts and Measurement. *Annu Rev Public Health*. 2006;27(1):167-194. doi:10.1146/annurev.publhealth.27.021405.102103
11. Carter-Pokras O, Baquet C. What is a “Health Disparity”? *Public Health Rep*. 2002;117(5):426-434. doi:10.1093/phr/117.5.426
12. Whitehead M. The Concepts and Principles of Equity and Health. *Int J Health Serv*. 1992;22(3):429-445. doi:10.2190/986L-LHQ6-2VTE-YRRN
13. Whitehead M, Dahlgren G. *Concepts and Principles for Tackling Social Inequities in Health*: WHO Collaborating Centre for Policy Research on Social Determinants of Health
https://www.euro.who.int/__data/assets/pdf_file/0010/74737/E89383.pdf
14. Agency for Healthcare Research and Quality. *2022 National Healthcare Quality and Disparities Report*.; 2023. Accessed May 11, 2025.
<https://www.ahrq.gov/research/findings/nhqrdr/nhqrdr22/index.html>

15. Braveman PA, Arkin E, Proctor D, Kauh T, Holm N. Systemic And Structural Racism: Definitions, Examples, Health Damages, And Approaches To Dismantling: Study examines definitions, examples, health damages, and dismantling systemic and structural racism. *Health Affairs*. 2022;41(2):171-178. doi:10.1377/hlthaff.2021.01394
16. Essien UR, Dusetzina SB, Gellad WF. A Policy Prescription for Reducing Health Disparities—Achieving Pharmacoequity. *JAMA*. 2021;326(18):1793. doi:10.1001/jama.2021.17764
17. Dwyer-Lindgren L, Kendrick P, Kelly YO, et al. Life expectancy by county, race, and ethnicity in the USA, 2000–19: a systematic analysis of health disparities. *The Lancet*. 2022;400(10345):25-38. doi:10.1016/S0140-6736(22)00876-5
18. Arias E, Xu J. *United States Life Tables, 2019*. National Center for Health Statistics; National Center for Health Statistics. Accessed October 10, 2024. <https://stacks.cdc.gov/view/cdc/113096>
19. Caraballo C, Massey DS, Ndumele CD, et al. Excess Mortality and Years of Potential Life Lost Among the Black Population in the US, 1999-2020. *JAMA*. 2023;329(19):1662. doi:10.1001/jama.2023.7022
20. Mahajan S, Caraballo C, Lu Y, et al. Trends in Differences in Health Status and Health Care Access and Affordability by Race and Ethnicity in the United States, 1999-2018. *JAMA*. 2021;326(7):637. doi:10.1001/jama.2021.9907
21. Briesacher B, Limcangco R, Gaskin D. Racial and Ethnic Disparities in Prescription Coverage and Medication Use. *Health Care Financing Review*. 25(2).
22. Buxbaum JD, Chernew ME, Fendrick AM, Cutler DM. Contributions Of Public Health, Pharmaceuticals, And Other Medical Care To US Life Expectancy Changes, 1990-2015: Study examines the conditions most responsible for changing US life expectancy and how public health, pharmaceuticals, other medical care, and other factors may have contributed to the changes. *Health Affairs*. 2020;39(9):1546-1556. doi:10.1377/hlthaff.2020.00284
23. Centers for Medicare & Medicaid Services. National Health Expenditure Accounts (NHEA) Data. Accessed May 12, 2025. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical>
24. IQVIA Institute. Medicine Spending and Affordability in the United States: Understanding Patients' Costs for Medicines. Published online 2023. Accessed April 27, 2025. <https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/medicine-spending-and-affordability-in-the-us>
25. Tichy EM, Hoffman JM, Tadrous M, et al. National trends in prescription drug expenditures and projections for 2024. *American Journal of Health-System Pharmacy*. 2024;81(14):583-598. doi:10.1093/ajhp/zxae105
26. Dieleman JL, Chen C, Crosby SW, et al. US Health Care Spending by Race and Ethnicity, 2002-2016. *JAMA*. 2021;326(7):649. doi:10.1001/jama.2021.9937
27. Goodney P, Dzebisashvili N, Goodman DC, Bronner KK. *Variation in the Care of Surgical Conditions: Diabetes and Peripheral Arterial Disease: A Dartmouth Atlas of Health Care Series*.

The Dartmouth Institute for Health Policy and Clinical Practice; 2014. Accessed April 13, 2025.
https://data.dartmouthatlas.org/downloads/reports/Diabetes_report_10_14_14.pdf

28. Lefebvre KM, Metraux S. Disparities in Level of Amputation Among Minorities: Implications for Improved Preventative Care. *Journal of the National Medical Association*. 2009;101(7):649-655. doi:10.1016/S0027-9684(15)30973-1
29. Chan KS, Parikh MA, Thorpe RJ, Gaskin DJ. Health Care Disparities in Race-Ethnic Minority Communities and Populations: Does the Availability of Health Care Providers Play a Role? *J Racial and Ethnic Health Disparities*. 2020;7(3):539-549. doi:10.1007/s40615-019-00682-w
30. Smedley BD, Stith AY, Nelson AR, Institute of Medicine (U.S.), eds. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. National Academy Press; 2003.
31. Maina IW, Belton TD, Ginzberg S, Singh A, Johnson TJ. A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. *Social Science & Medicine*. 2018;199:219-229. doi:10.1016/j.socscimed.2017.05.009
32. Feagin J, Bennefield Z. Systemic racism and U.S. health care. *Social Science & Medicine*. 2014;103:7-14. doi:10.1016/j.socscimed.2013.09.006
33. Hall WJ, Chapman MV, Lee KM, et al. Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review. *Am J Public Health*. 2015;105(12):e60-e76. doi:10.2105/AJPH.2015.302903
34. Essien UR, Holmes DN, Jackson LR, et al. Association of Race/Ethnicity With Oral Anticoagulant Use in Patients With Atrial Fibrillation: Findings From the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation II. *JAMA Cardiol*. 2018;3(12):1174. doi:10.1001/jamacardio.2018.3945
35. Nong P, Raj M, Creary M, Kardia SLR, Platt JE. Patient-Reported Experiences of Discrimination in the US Health Care System. *JAMA Netw Open*. 2020;3(12):e2029650. doi:10.1001/jamanetworkopen.2020.29650
36. Pletcher MJ, Kertesz SG, Kohn MA, Gonzales R. Trends in Opioid Prescribing by Race/Ethnicity for Patients Seeking Care in US Emergency Departments. *JAMA*. 2008;299(1). doi:10.1001/jama.2007.64
37. Hoffman KM, Trawalter S, Axt JR, Oliver MN. Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proc Natl Acad Sci U S A*. 2016;113(16):4296-4301. doi:10.1073/pnas.1516047113
38. Buchmueller TC, Levy HG. The ACA's Impact On Racial And Ethnic Disparities In Health Insurance Coverage And Access To Care: An examination of how the insurance coverage expansions of the Affordable Care Act have affected disparities related to race and ethnicity. *Health Affairs*. 2020;39(3):395-402. doi:10.1377/hlthaff.2019.01394
39. Berchick ER, Hood E, Barnett JC. Health Insurance Coverage in the United States: 2017.

40. Qato DM, Daviglus ML, Wilder J, Lee T, Qato D, Lambert B. 'Pharmacy Deserts' Are Prevalent In Chicago's Predominantly Minority Communities, Raising Medication Access Concerns. *Health Affairs*. 2014;33(11):1958-1965. doi:10.1377/hlthaff.2013.1397
41. Guadamuz JS, Wilder JR, Mouslim MC, Zenk SN, Alexander GC, Qato DM. Fewer Pharmacies In Black And Hispanic/Latino Neighborhoods Compared With White Or Diverse Neighborhoods, 2007–15: Study examines pharmacy “deserts” in Black and Hispanic/Latino neighborhoods compared with white or diverse neighborhoods. *Health Affairs*. 2021;40(5):802-811. doi:10.1377/hlthaff.2020.01699
42. Guadamuz JS, Alexander GC, Zenk SN, Kanter GP, Wilder JR, Qato DM. Access to pharmacies and pharmacy services in New York City, Los Angeles, Chicago, and Houston, 2015-2020. *Journal of the American Pharmacists Association*. 2021;61(6):e32-e41. doi:10.1016/j.japh.2021.07.009
43. Chisholm-Burns MA, Spivey CA, Gatwood J, Wiss A, Hohmeier K, Erickson SR. Evaluation of racial and socioeconomic disparities in medication pricing and pharmacy access and services. *American Journal of Health-System Pharmacy*. 2017;74(10):653-668. doi:10.2146/ajhp150872
44. Luo J, Kulldorff M, Sarpatwari A, Pawar A, Kesselheim AS. Variation in Prescription Drug Prices by Retail Pharmacy Type: A National Cross-sectional Study. *Ann Intern Med*. 2019;171(9):605. doi:10.7326/M18-1138
45. Gellad WF, Choudhry NK, Friedberg MW, Brookhart MA, Haas JS, Shrank WH. Variation in Drug Prices at Pharmacies: Are Prices Higher in Poorer Areas? *Health Services Research*. 2009;44(2p1):606-617. doi:10.1111/j.1475-6773.2008.00917.x
46. PhRMA Foundation. Addressing Disparities in Medication Access and Adherence. Published online May 2022. Accessed October 3, 2022. https://phrma.org/-/media/Project/PhRMA/PhRMA-Org/PhRMA-Refresh/Report-PDFs/A-C/Addressing-Disparity-Report_v3p1.pdf
47. Doshi JA, Li P, Huo H, Pettit AR, Armstrong KA. Association of Patient Out-of-Pocket Costs With Prescription Abandonment and Delay in Fills of Novel Oral Anticancer Agents. *JCO*. 2018;36(5):476-482. doi:10.1200/JCO.2017.74.5091
48. Panzone J, Welch C, Morgans A, et al. Association of Race With Cancer-Related Financial Toxicity. *JCO Oncology Practice*. 2022;18(2):e271-e283. doi:10.1200/OP.21.00440
49. Devine JW, Lim D, Lugo A, Farley JF. Prevalence and patterns of catastrophic spending for antidiabetic medication in 2020. *JMCP*. 2023;29(10):1158-1164. doi:10.18553/jmcp.2023.29.10.1158
50. Ratner NL, Davis EB, Lhotka LL, Wille SM, Walls ML. Patient-Centered Care, Diabetes Empowerment, and Type 2 Diabetes Medication Adherence Among American Indian Patients. *Clinical Diabetes*. 2017;35(5):281-285. doi:10.2337/cd17-0008
51. Halbert CH, Armstrong K, Gandy OH, Shaker L. Racial Differences in Trust in Health Care Providers. *Arch Intern Med*. 2006;166(8):896. doi:10.1001/archinte.166.8.896
52. Armstrong K, Putt M, Halbert CH, et al. Prior experiences of racial discrimination and racial differences in health care system distrust. *Med Care*. 2013;51(2):144-150. doi:10.1097/MLR.0b013e31827310a1

53. Bazargan M, Cobb S, Assari S. Discrimination and Medical Mistrust in a Racially and Ethnically Diverse Sample of California Adults. *Ann Fam Med*. 2021;19(1):4-15. doi:10.1370/afm.2632
54. Chakravarty S. Did the Medicare Prescription Drug Program Lead to New Racial and Ethnic Disparities? Examining Long-term Changes in Prescription Drug Access among Minority Populations. *Social Work in Public Health*. 2020;35(5):248-260. doi:10.1080/19371918.2020.1785981
55. Wilder ME, Kulie P, Jensen C, et al. The Impact of Social Determinants of Health on Medication Adherence: a Systematic Review and Meta-analysis. *J Gen Intern Med*. 2021;36(5):1359-1370. doi:10.1007/s11606-020-06447-0
56. Singer AW, Weiser SD, McCoy SI. Does Food Insecurity Undermine Adherence to Antiretroviral Therapy? A Systematic Review. *AIDS Behav*. 2015;19(8):1510-1526. doi:10.1007/s10461-014-0873-1
57. Wroth TH, Pathman DE. Primary Medication Adherence in a Rural Population: The Role of the Patient-Physician Relationship and Satisfaction with Care. *The Journal of the American Board of Family Medicine*. 2006;19(5):478-486. doi:10.3122/jabfm.19.5.478
58. Hensley C, Heaton PC, Kahn RS, Luder HR, Frede SM, Beck AF. Poverty, Transportation Access, and Medication Nonadherence. *Pediatrics*. 2018;141(4):e20173402. doi:10.1542/peds.2017-3402
59. LaVeist TA, Nuru-Jeter A, Jones KE. The association of doctor-patient race concordance with health services utilization. *J Public Health Pol*. 2003;24(3):312-323. doi:10.2307/3343378
60. Takeshita J, Wang S, Loren AW, et al. Association of Racial/Ethnic and Gender Concordance Between Patients and Physicians With Patient Experience Ratings. *JAMA Netw Open*. 2020;3(11):e2024583. doi:10.1001/jamanetworkopen.2020.24583
61. Saha S, Beach MC. Impact of Physician Race on Patient Decision-Making and Ratings of Physicians: a Randomized Experiment Using Video Vignettes. *J Gen Intern Med*. 2020;35(4):1084-1091. doi:10.1007/s11606-020-05646-z
62. Jetty A, Jabbarpour Y, Pollack J, Huerto R, Woo S, Petterson S. Patient-Physician Racial Concordance Associated with Improved Healthcare Use and Lower Healthcare Expenditures in Minority Populations. *J Racial and Ethnic Health Disparities*. 2022;9(1):68-81. doi:10.1007/s40615-020-00930-4
63. Mora H, Obayemi A, Holcomb K, Hinson M. The National Deficit of Black and Hispanic Physicians in the US and Projected Estimates of Time to Correction. *JAMA Netw Open*. 2022;5(6):e2215485. doi:10.1001/jamanetworkopen.2022.15485
64. Essien UR, Agbafé V, Norris KC. Diversifying the Medical Pathway in a Post-Affirmative Action World. *JAMA*. 2023;330(14):1325. doi:10.1001/jama.2023.17089
65. Appolon G, Tang S, Gabriel N, et al. Association Between Redlining and Spatial Access to Pharmacies. *JAMA Netw Open*. 2023;6(8):e2327315. doi:10.1001/jamanetworkopen.2023.27315

66. Zhang Y, Ancker JS, Hall J, Khullar D, Wu Y, Kaushal R. Association Between Residential Neighborhood Social Conditions and Health Care Utilization and Costs. *Medical Care*. 2020;58(7):586-593. doi:10.1097/MLR.0000000000001337
67. Guadamuz JS, Alexander GC, Zenk SN, Qato DM. Assessment of Pharmacy Closures in the United States From 2009 Through 2015. *JAMA Intern Med*. 2020;180(1):157. doi:10.1001/jamainternmed.2019.4588
68. Guadamuz JS, Alexander GC, Kanter GP, Qato DM. More US Pharmacies Closed Than Opened In 2018–21; Independent Pharmacies, Those In Black, Latinx Communities Most At Risk: Study examines US pharmacy closures at the county level, 2018–21. *Health Affairs*. 2024;43(12):1703-1711. doi:10.1377/hlthaff.2024.00192
69. Amstislavski P, Matthews A, Sheffield S, Maroko AR, Weedon J. Medication deserts: survey of neighborhood disparities in availability of prescription medications. *Int J Health Geogr*. 2012;11(1):48. doi:10.1186/1476-072X-11-48
70. Qato D. Availability of Buprenorphine at Retail Pharmacies in the U.S., 2011-2016. In: *Structural Racism and Disparities in Medications for Opioid Use Disorder*. ; 2022. <https://academyhealth.confex.com/academyhealth/2022arm/meetingapp.cgi/Paper/53832>
71. Groos M, Wallace M, Hardeman R, Theall KP. Measuring inequity: a systematic review of methods used to quantify structural racism. 2018;11(2).
72. Yearby R. Racial Disparities in Health Status and Access to Healthcare: The Continuation of Inequality in the United States Due to Structural Racism. *American J Econ Sociol*. 2018;77(3-4):1113-1152. doi:10.1111/ajes.12230
73. Chalasani R, Krishnamurthy S, Suda KJ, Newman TV, Delaney SW, Essien UR. Pursuing Pharmacoequity: Determinants, Drivers, and Pathways to Progress. *Journal of Health Politics, Policy and Law*. Published online July 14, 2022:10041135. doi:10.1215/03616878-10041135
74. Berenbrok LA, Tang S, Gabriel N, et al. Access to community pharmacies: A nationwide geographic information systems cross-sectional analysis. *Journal of the American Pharmacists Association*. 2022;62(6):1816-1822.e2. doi:10.1016/j.japh.2022.07.003
75. Walker RE, Keane CR, Burke JG. Disparities and access to healthy food in the United States: A review of food deserts literature. *Health & Place*. 2010;16(5):876-884. doi:10.1016/j.healthplace.2010.04.013
76. US Department of Agriculture Economic Research Service. Food Access Research Atlas. Accessed May 13, 2025. <https://www.ers.usda.gov/data-products/food-access-research-atlas/go-to-the-atlas>
77. Qato DM, Alexander GC, Chakraborty A, Guadamuz JS, Jackson JW. Association Between Pharmacy Closures and Adherence to Cardiovascular Medications Among Older US Adults. *JAMA Netw Open*. 2019;2(4):e192606. doi:10.1001/jamanetworkopen.2019.2606
78. Anderson KE, Sahu M, DiStefano MJ, Asche CV, Mattingly TJ. Pharmacy Closures and Anticonvulsant Medication Prescription Fills. *JAMA*. 2024;332(21):1847. doi:10.1001/jama.2024.19993

79. Urick BY, Adams JK, Bruce MR. State Telepharmacy Policies and Pharmacy Deserts. *JAMA Netw Open*. 2023;6(8):e2328810. doi:10.1001/jamanetworkopen.2023.28810
80. Pathak S, Haynes M, Qato DM, Urick BY. Telepharmacy and Quality of Medication Use in Rural Areas, 2013–2019. *Prev Chronic Dis*. 2020;17:200012. doi:10.5888/pcd17.200012
81. Guadamuz JS, McCormick CD, Choi S, Urick B, Alexander GC, Qato DM. Telepharmacy and medication adherence in urban areas. *Journal of the American Pharmacists Association*. 2021;61(2):e100-e113. doi:10.1016/j.japh.2020.10.017
82. Dieleman JL, Beauchamp M, Crosby SW, et al. Tracking US Health Care Spending by Health Condition and County. *JAMA*. Published online February 14, 2025. doi:10.1001/jama.2024.26790
83. Dieleman JL, Weil M, Beauchamp M, et al. Drivers of Variation in Health Care Spending Across US Counties. *JAMA Health Forum*. 2025;6(2):e245220. doi:10.1001/jamahealthforum.2024.5220
84. Dwyer-Lindgren L, Kendrick P, Kelly YO, et al. Cause-specific mortality by county, race, and ethnicity in the USA, 2000–19: a systematic analysis of health disparities. *The Lancet*. 2023;402(10407):1065-1082. doi:10.1016/S0140-6736(23)01088-7
85. 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
86. Horst C. *Predicting Medical Diagnoses from Pharmaceutical Claims Data*. University of Washington; 2017.
https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/40423/Horst_washington_0250O_17931.pdf?sequence=1
87. Dwyer-Lindgren L, Baumann MM, Li Z, et al. Ten Americas: a systematic analysis of life expectancy disparities in the USA. *The Lancet*. 2024;404(10469):2299-2313. doi:10.1016/S0140-6736(24)01495-8
88. Oslo: WHO Collaborating Centre for Drug Statistics Methodology. Purpose of the ATC/DDD system. October 11, 2022. Accessed February 16, 2023.
https://www.whocc.no/atc_ddd_methodology/purpose_of_the_atc_ddd_system/
89. Merative Micromedex. RED BOOK Flat File. Accessed February 7, 2026.
<https://www.merative.com/documents/micromedex-red-book-flat-file>
90. National Library of Medicine. Unified Medical Language System: RxNorm. Accessed February 16, 2023. <https://www.nlm.nih.gov/research/umls/rxnorm/index.html>
91. Ding Y, Miller GE, Hill S. RESEARCH FINDINGS #49: Pre-COVID-19 Retail Use and Expenditures for Drugs That Were Subsequently Used to Treat COVID-19. Agency for Healthcare Research and Quality. July 2021. Accessed February 16, 2023.
https://meps.ahrq.gov/data_files/publications/rf49/rf49.shtml
92. Peterson C, Grosse SD, Dunn A. A practical guide to episode groupers for cost-of-illness analysis in health services research. *SAGE Open Medicine*. 2019;7:205031211984020. doi:10.1177/2050312119840200

93. Merative. MarketScan Research Databases. Accessed January 22, 2025.
<https://www.merative.com/documents/merative-marketscan-research-databases>
94. Merative. *Micromedex® (Electronic Version)*. Accessed February 16, 2023.
<https://www.micromedexsolutions.com/>
95. Assistant Secretary for Planning and Evaluation (ASPE). Issue Brief: Trends in Prescription Drug Spending, 2016-2021. Published online September 2022. Accessed October 24, 2024.
<https://aspe.hhs.gov/sites/default/files/documents/88c547c976e915fc31fe2c6903ac0bc9/sdp-trends-prescription-drug-spending.pdf>
96. U.S. Department of Health & Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. National Health and Nutrition Examination Survey: Prescription Medications files [Data files and documentation]. Published online 2014, 2015–2016, 2017-2020, 2021-2023 2013. Accessed July 1, 2025. <https://www.cdc.gov/nchs/nhanes/index.htm>
97. Latkin CA, Edwards C, Davey-Rothwell MA, Tobin KE. The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. *Addictive Behaviors*. 2017;73:133-136.
doi:10.1016/j.addbeh.2017.05.005
98. Radley DC, Finkelstein SN, Stafford RS. Off-label Prescribing Among Office-Based Physicians. *Arch Intern Med*. 2006;166(9):1021. doi:10.1001/archinte.166.9.1021
99. Essien U. The Inflation Reduction Act: One step closer to pharmacoequity? Published online August 19, 2022. Accessed December 20, 2024. <https://www.statnews.com/2022/08/19/the-inflation-reduction-act-one-step-closer-to-pharmacoequity/>
100. Riley GF. Administrative and Claims Records as Sources of Health Care Cost Data. *Medical Care*. 2009;47(7_Supplement_1):S51-S55. doi:10.1097/MLR.0b013e31819c95aa
101. Gaskin D, Briesacher B, Limcangco R, Brigantti B. Exploring racial and ethnic disparities in prescription drug spending and use among medicare beneficiaries. *Am J Geriatr Pharmacother* . 2006;4(2):96-111. doi:<https://doi.org/10.1016/j.amjopharm.2006.06.008>
102. Mahmoudi E, Jensen GA. Has Medicare Part D Reduced Racial/Ethnic Disparities in Prescription Drug Use and Spending? *Health Serv Res*. 2014;49(2):502-525. doi:10.1111/1475-6773.12099
103. Zissimopoulos J, Goldman DP. Did Medicare Part D Reduce Disparities? Published online 2015.
104. Centers for Medicare & Medicaid Services (CMS). Health Expenditures by State of Residence, 1991-2020. Accessed May 10, 2024. <https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/state-residence>
105. Beauchamp M, Thomson A, et al. Tracking US State Health Spending by Race and Ethnicity. [Under review].
106. Assistant Secretary for Planning and Evaluation (ASPE). Report: How Increased Funding Can Advance the Mission of the Indian Health Service to Improve Health Outcomes for American Indians and Alaska Natives. Published online July 2022. Accessed October 24, 2024.

<https://aspe.hhs.gov/sites/default/files/documents/e7b3d02affdda1949c215f57b65b5541/aspe-ihs-funding-disparities-report.pdf>

107. United States Census Bureau. Race. Accessed October 24, 2024.
<https://www.census.gov/topics/population/race.html>
108. National Center for Health Statistics. U.S. Census Populations With Bridged Race Categories. Published online October 28, 2022. Accessed October 24, 2024.
https://www.cdc.gov/nchs/nvss/bridged_race.htm
109. Shadac. State Health Compare: Health Insurance Coverage Type. Accessed October 24, 2024.
<https://statehealthcompare.shadac.org/table/29/health-insurance-coverage-type-by-race-ethnicity#1/5,4,1,10,86,9,8,6,39,40,41,42,238,43,239/42/57,58>
110. Centers for Medicare & Medicaid Services. National Health Expenditure Accounts: Methodology Paper, 2022: Definitions, Sources, and Methods. Accessed October 10, 2024.
<https://www.cms.gov/files/document/definitions-sources-and-methods.pdf>
111. Das Gupta P. *Standardization and Decomposition of Rates: A User's Manual*. US Bureau of the Census; :23-186. <https://www2.census.gov/library/publications/1993/demographics/p23-186.pdf>
112. Chevan A, Sutherland M. Revisiting das gupta: Refinement and extension of standardization and decomposition. *Demography*. 2009;46(3):429-449. doi:10.1353/dem.0.0060
113. Stevens GA, Alkema L, Black RE, et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *The Lancet*. 2016;388(10062):e19-e23. doi:10.1016/S0140-6736(16)30388-9
114. Aburto JM, Tilstra AM, Floridi G, Dowd JB. Significant impacts of the COVID-19 pandemic on race/ethnic differences in US mortality. *Proc Natl Acad Sci USA*. 2022;119(35):e2205813119. doi:10.1073/pnas.2205813119
115. Hill L, Artiga S. *COVID-19 Cases and Deaths by Race/Ethnicity: Current Data and Changes Over Time*. KFF; 2022. Accessed September 3, 2024. <https://www.kff.org/racial-equity-and-health-policy/issue-brief/covid-19-cases-and-deaths-by-race-ethnicity-current-data-and-changes-over-time/>
116. Wittenauer R, Shah PD, Bacci JL, Stergachis A. Locations and characteristics of pharmacy deserts in the United States: a geospatial study. *Health Affairs Scholar*. 2024;2(4):qxae035. doi:10.1093/haschl/qxae035
117. Kaplan CM, Waters TM, Clear ER, Graves EE, Henderson S. The Impact of Prescription Drug Coverage on Disparities in Adherence and Medication Use: A Systematic Review. *Med Care Res Rev*. 2024;81(2):87-95. doi:10.1177/10775587231218050
118. Finkelstein A, Taubman S, Wright B, et al. The Oregon Health Insurance Experiment: Evidence from the First Year*. *The Quarterly Journal of Economics*. 2012;127(3):1057-1106. doi:10.1093/qje/qjs020
119. Ghosh A, Simon K, Sommers BD. The Effect of Health Insurance on Prescription Drug Use Among Low-Income Adults:Evidence from Recent Medicaid Expansions. *Journal of Health Economics*. 2019;63:64-80. doi:10.1016/j.jhealeco.2018.11.002

120. Sumarsono A, Lalani H, Segar MW, et al. Association of Medicaid Expansion With Rates of Utilization of Cardiovascular Therapies Among Medicaid Beneficiaries Between 2011 and 2018. *Circ: Cardiovascular Quality and Outcomes*. 2021;14(1):e007492. doi:10.1161/CIRCOUTCOMES.120.007492
121. H.R.5376 - 117th Congress (2021-2022): Inflation Reduction Act of 2022. Published online August 16, 2022. <https://www.congress.gov/bill/117th-congress/house-bill/5376>
122. Sullivan PS, DuBose SN, Castel AD, et al. Equity of PrEP uptake by race, ethnicity, sex and region in the United States in the first decade of PrEP: a population-based analysis. *The Lancet Regional Health - Americas*. 2024;33:100738. doi:10.1016/j.lana.2024.100738
123. Eberly LA, Yang L, Essien UR, et al. Racial, Ethnic, and Socioeconomic Inequities in Glucagon-Like Peptide-1 Receptor Agonist Use Among Patients With Diabetes in the US. *JAMA Health Forum*. 2021;2(12):e214182. doi:10.1001/jamahealthforum.2021.4182
124. Eberly LA, Yang L, Eneanya ND, et al. Association of Race/Ethnicity, Gender, and Socioeconomic Status With Sodium-Glucose Cotransporter 2 Inhibitor Use Among Patients With Diabetes in the US. *JAMA Netw Open*. 2021;4(4):e216139. doi:10.1001/jamanetworkopen.2021.6139
125. Kruse G, Lopez-Carmen VA, Jensen A, Hardie L, Sequist TD. The Indian Health Service and American Indian/Alaska Native Health Outcomes. *Annu Rev Public Health*. 2022;43(1):559-576. doi:10.1146/annurev-publhealth-052620-103633
126. Nead KT, Hinkston CL, Wehner MR. Cautions When Using Race and Ethnicity in Administrative Claims Data Sets. *JAMA Health Forum*. 2022;3(7):e221812. doi:10.1001/jamahealthforum.2022.1812
127. Dickman SL, Gaffney A, McGregor A, et al. Trends in Health Care Use Among Black and White Persons in the US, 1963-2019. *JAMA Netw Open*. 2022;5(6):e2217383. doi:10.1001/jamanetworkopen.2022.17383
128. Sahu M, et al. Prescription Drug Utilization and Spending by Race, Ethnicity, Payer, Health Condition, and US State. *Under review*.
129. Essien UR, Dusetzina SB, Gellad WF. A Policy Prescription for Reducing Health Disparities—Achieving Pharmacoequity. *JAMA*. 2021;326(18):1793. doi:10.1001/jama.2021.17764
130. Goode JV, Owen J, Page A, Gatewood S. Community-Based Pharmacy Practice Innovation and the Role of the Community-Based Pharmacist Practitioner in the United States. *Pharmacy*. 2019;7(3):106. doi:10.3390/pharmacy7030106
131. Valliant SN, Burbage SC, Pathak S, Urick BY. Pharmacists as accessible health care providers: quantifying the opportunity. *JMCP*. 2022;28(1):85-90. doi:10.18553/jmcp.2022.28.1.85
132. Shen AK, Peterson A. The pharmacist and pharmacy have evolved to become more than the corner drugstore: a win for vaccinations and public health. *Human Vaccines & Immunotherapeutics*. 2020;16(5):1178-1180. doi:10.1080/21645515.2019.1660119
133. Qato DM, Zenk S, Wilder J, Harrington R, Gaskin D, Alexander GC. The availability of pharmacies in the United States: 2007–2015. *PLoS One*. 2017;12(8):e0183172:13.

134. Constantin J, Ullrich F, Mueller KJ. *Rural and Urban Pharmacy Presence – Pharmacy Deserts*. RUPRI Center for Rural Health Policy Analysis; 2022. Accessed January 25, 2025. <https://rupri.public-health.uiowa.edu/publications/policybriefs/2022/Pharmacy%20Deserts.pdf>
135. Sharareh N, Zheutlin AR, Qato DM, Guadamuz J, Bress A, Vos RO. Access to community pharmacies based on drive time and by rurality across the contiguous United States. *Journal of the American Pharmacists Association*. 2024;64(2):476-482. doi:10.1016/j.japh.2024.01.004
136. Salako A, Ullrich F, Mueller KJ. Update: Independently Owned Pharmacy Closures in Rural America, 2003-2018. :6.
137. Hiscock R, Pearce J, Blakely T, Witten K. Is Neighborhood Access to Health Care Provision Associated with Individual-Level Utilization and Satisfaction? *Health Services Research*. 2008;43(6):2183-2200. doi:10.1111/j.1475-6773.2008.00877.x
138. Li X, Vojnovic I, Ligmann-Zielinska A. Spatial accessibility and travel to pharmacy by type in the Detroit region. *Journal of the American Pharmacists Association*. 2024;64(3):102052. doi:10.1016/j.japh.2024.102052
139. Syed ST, Sharp LK, Kim Y, et al. Relationship Between Medication Adherence and Distance to Dispensing Pharmacies and Prescribers Among an Urban Medicaid Population with Diabetes Mellitus. *Pharmacotherapy*. 2016;36(6):590-597. doi:10.1002/phar.1757
140. Neuner JM, Zhou Y, Fergestrom N, et al. Pharmacy deserts and patients with breast cancer receipt of influenza vaccines. *Journal of the American Pharmacists Association*. 2021;61(6):e25-e31. doi:10.1016/j.japh.2021.07.006
141. Wittenauer R, Shah PD, Bacci JL, Stergachis A. Pharmacy deserts and COVID-19 risk at the census tract level in the State of Washington. *Vaccine: X*. 2022;12:100227. doi:10.1016/j.jvacx.2022.100227
142. Pednekar P, Peterson AM, Heller D. Comparing Medication Adherence Rates Between Pharmacy Desert and Non-Desert Areas Among Elderly in Pennsylvania. *Value in Health*. 2016;19(3):A205-A206. doi:10.1016/j.jval.2016.03.1253
143. National Council for Prescription Drug Programs. National Council for Prescription Drug Programs (NCPDP) DataQ version 3.1. April 30, 2025. <https://www.ncdp.org/>
144. US Census Bureau. American Community Survey, 2013-2019 American Community Survey 5-Year Estimates. Tables B03002. Hispanic or Latino Origin by Race. Accessed using tidycensus package. Accessed May 2, 2025. <https://data.census.gov/table/ACSDT5Y2023.B08201>
145. US Census Bureau. Decennial Census, 2010. Tables P002005; using tidycensus package. Hispanic or Latino Origin by Race. Accessed March 21, 2025. <https://api.census.gov/data/2010/dec/pl/variables.html>
146. US Census Bureau. American Community Survey, 2011-2019 American Community Survey 5-Year Estimates. Tables B08201. Household Size by Vehicles Available. Accessed using tidycensus package. Accessed May 2, 2025. <https://data.census.gov/table/ACSDT5Y2023.B08201>

147. Wagstaff A, Paci P, Van Doorslaer E. On the measurement of inequalities in health. *Social Science & Medicine*. 1991;33(5):545-557. doi:10.1016/0277-9536(91)90212-U
148. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. 1st ed. Wiley; 1987. doi:10.1002/9780470316696
149. Pebesma E. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*. 2018;10(1):439. doi:10.32614/rj-2018-009
150. Croissant Y, Millo G. Panel Data Econometrics in R: The plm Package. *J Stat Soft*. 2008;27(2). doi:10.18637/jss.v027.i02
151. Zhou Y, Beyer KMM, Laud PW, et al. An adapted two-step floating catchment area method accounting for urban–rural differences in spatial access to pharmacies. *Journal of Pharmaceutical Health Services Research*. 2021;12(1):69-77. doi:10.1093/jphsr/rmaa022
152. Ikram SZ, Hu Y, Wang F. Disparities in Spatial Accessibility of Pharmacies in Baton Rouge, Louisiana. *Geographical Review*. 2015;105(4):492-510. doi:10.1111/j.1931-0846.2015.12087.x
153. Tharumia Jagadeesan C, Wirtz VJ. Geographical accessibility of medicines: a systematic literature review of pharmacy mapping. *J of Pharm Policy and Pract*. 2021;14(1). doi:10.1186/s40545-020-00291-7
154. Sharareh N, Tang S, Bress A, Mathis WS, Berenbrok LA, Hernandez I. Geographic access to community pharmacies based on walking, driving, and public transportation in the 10 most populated U.S. areas. *J Am Pharm Assoc (2003)*. Published online July 18, 2025:102479. doi:10.1016/j.japh.2025.102479
155. Catalano G, Khan MMM, Chatzipanagiotou OP, Pawlik TM. Pharmacy Accessibility and Social Vulnerability. *JAMA Netw Open*. 2024;7(8):e2429755. doi:10.1001/jamanetworkopen.2024.29755
156. Padeiro M. *Planning for Health: Access to Medications by Elderly People in Pharmaceutical Deprivation Areas*.; 2017. Accessed January 26, 2025. https://www.researchgate.net/profile/Miguel-Padeiro/publication/318362648_Planning_for_health_access_to_medications_by_elderly_people_in_pharmaceutical_deprivation_areas/links/5965ff23a6fdcc85dd3b9d21/Planning-for-health-access-to-medications-by-elderly-people-in-pharmaceutical-deprivation-areas.pdf
157. Do D, Geldsetzer P. Trends in Mail-Order Pharmacy Use in the U.S. From 1996 to 2018: An Analysis of the Medical Expenditure Panel Survey. *American Journal of Preventive Medicine*. 2021;61(2):e63-e72. doi:10.1016/j.amepre.2021.02.017
158. National Community Pharmacists Association. *National Consumer Survey: More than 8 in 10 Adults Prefer Their Local Pharmacist over Mail Order*.; 2021. Accessed January 26, 2025. <https://ncpa.org/newsroom/news-releases/2021/03/04/national-consumer-survey-more-8-10-adults-prefer-their-local>
159. Anderson KE, Mattingly TJ. Measuring Geographic Access to Pharmacies. *JAMA Netw Open*. 2025;8(3):e250725. doi:10.1001/jamanetworkopen.2025.0725

160. Mathis WS, Berenbrok LA, Kahn PA, Appolon G, Tang S, Hernandez I. Vulnerability Index Approach to Identify Pharmacy Deserts and Keystone Pharmacies. *JAMA Netw Open*. 2025;8(3):e250715. doi:10.1001/jamanetworkopen.2025.0715
161. Labban M, Chen CR, Frego N, et al. Disparities in Travel-Related Barriers to Accessing Health Care From the 2017 National Household Travel Survey. *JAMA Netw Open*. 2023;6(7):e2325291. doi:10.1001/jamanetworkopen.2023.25291
162. U.S. Congress. *One Big Beautiful Bill Act, H.R. 1, 119th Cong.*; 2025. Accessed July 18, 2025. <https://www.congress.gov/bill/119th-congress/house-bill/1/text>
163. Agarwal R, Gupta A, Fendrick AM. Value-Based Insurance Design Improves Medication Adherence Without An Increase In Total Health Care Spending. *Health Affairs*. 2018;37(7):1057-1064. doi:10.1377/hlthaff.2017.1633
164. Franken M, Stolk E, Scharringhausen T, De Boer A, Koopmanschap M. A comparative study of the role of disease severity in drug reimbursement decision making in four European countries. *Health Policy*. 2015;119(2):195-202. doi:10.1016/j.healthpol.2014.10.007
165. Hofmann S, Branner J, Misra A, Lintener H. A Review of Current Approaches to Defining and Valuing Innovation in Health Technology Assessment. *Value in Health*. 2021;24(12):1773-1783. doi:10.1016/j.jval.2021.06.006
166. Teerawattananon Y, Vishwanath Dabak S, Culyer A, Mills A, Kingkaew P, Isaranuwatthai W. Fifteen Lessons from Fifteen Years of the Health Intervention and Technology Assessment Program in Thailand. *Health Systems & Reform*. 2023;9(3). doi:10.1080/23288604.2024.2330974
167. Committee on Strategies to Better Align Investments in Innovations for Therapeutic Development with Disease Burden and Unmet Needs, Board of Health Care Services, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. *Aligning Investments in Therapeutic Development with Therapeutic Need: Closing the Gap*. (Berwick D, MacKenzie E, Helman A, Schumm S, eds.). National Academies Press; 2025. doi:10.17226/29157
168. Wens I, Claessens Z, Vanneste A, Barbier L, Janssens R, Huys I. Unmet medical needs definition and incentives: stakeholders perspectives on the reform of the EU pharmaceutical legislation. *Front Med*. 2025;11. doi:10.3389/fmed.2024.1506243
169. U.S. Congress. *Ryan White Comprehensive AIDS Resources Emergency (CARE) Act, Public Law 101-381*; 1990. <https://www.congress.gov/bill/101st-congress/senate-bill/2240>
170. Barnabas RV, Szpiro AA, van Rooyen H, et al. Community-based antiretroviral therapy versus standard clinic-based services for HIV in South Africa and Uganda (DO ART): a randomised trial. *The Lancet Global Health*. 2020;8(10):e1305-e1315. doi:10.1016/S2214-109X(20)30313-2
171. Barnabas RV, Szpiro AA, Ntinga X, et al. Fee for home delivery and monitoring of antiretroviral therapy for HIV infection compared with standard clinic-based services in South Africa: a randomised controlled trial. *The Lancet HIV*. 2022;9(12):e848-e856. doi:10.1016/S2352-3018(22)00254-5
172. Ásbjörnsdóttir KH, Ajjampur SSR, Anderson RM, et al. Assessing the feasibility of interrupting the transmission of soil-transmitted helminths through mass drug administration: The DeWorm3

- cluster randomized trial protocol. Garba A, ed. *PLoS Negl Trop Dis*. 2018;12(1):e0006166. doi:10.1371/journal.pntd.0006166
173. Anderson LA. National Drug Codes Explained. October 1, 2020. Accessed February 15, 2023. <https://www.drugs.com/ndc.html>
 174. US Food and Drug Administration. Generic Drugs: Questions & Answers. March 16, 2021. Accessed February 16, 2023. <https://www.fda.gov/drugs/frequently-asked-questions-popular-topics/generic-drugs-questions-answers>
 175. Cartwright DJ. ICD-9-CM to ICD-10-CM Codes: What? Why? How? *Advances in Wound Care*. 2013;2(10):588-592. doi:10.1089/wound.2013.0478
 176. Filzmoser P, Eisl A, Endel F. ATC → ICD: Determination of the reliability for predicting the ICD code from the ATC code. Published online December 10, 2009.
 177. Weisser A, Endel G, Filzmoser P, Gyimesi M. ATC -> ICD – evaluating the reliability of prognoses for ICD-10 diagnoses derived from the ATC-Code of prescriptions. *BMC Health Serv Res*. 2008;8(S1):A10, 1472-6963-8-S1-A10. doi:10.1186/1472-6963-8-S1-A10
 178. Weisser A, Endel F, Endel G, Filzmoser P. Results of the project ATC-ICD. Accessed February 16, 2023. <https://www.sozialversicherung.at/cdscontent/load?contentid=10008.715143&version=1391184571>
 179. Weibrecht N, Endel F, Zechmeister M. Evaluating ATC-ICD: Assessing the relationship between selected medication and diseases with machine learning. *IJPDS*. 2019;4(3). doi:10.23889/ijpds.v4i3.1292
 180. Watson Health. IBM Micromedex RED BOOK. <https://www.ibm.com/downloads/cas/OZKXWE1A>. <https://www.ibm.com/downloads/cas/OZKXWE1A>
 181. US Food and Drug Administration. openFDA National Drug Code Directory. Accessed February 16, 2023. <https://open.fda.gov/data/downloads/>
 182. Dieleman JL, Beauchamp M, Crosby SW, et al. Tracking US Health Care Spending by Health Condition and County. *JAMA*. Published online February 14, 2025. doi:10.1001/jama.2024.26790
 183. Agency for Health care Research and Quality. Medical Expenditure Panel Survey (MEPS). Accessed November 12, 2020. <https://www.meps.ahrq.gov/mepsweb/>
 184. Centers for Medicare & Medicaid Services. *Part D Event (PDE) File*. ResDAC Research Data Assistance Center Accessed April 4, 2025. <https://resdac.org/cms-data/files/pde>
 185. Centers for Medicare & Medicaid Services. *TAF Technical Documentation: Claims Files*. MACBIS Medicaid and CHIP Business Information Solutions; 2021. Accessed April 4, 2025. https://resdac.org/sites/datadocumentation.resdac.org/files/2021-08/TAF_TechGuide_Claims_Files.pdf
 186. Kythera Labs, Inc. *Deciphering Healthcare*.; 2020. Accessed February 25, 2023. <https://www.kytheralabs.com/>

187. Health Care Cost Institute. *HCCI 2.0 Commercial Claims*. Accessed February 25, 2023. <https://healthcostinstitute.org/data-access-hub>
188. US Office of Management and Budget (OMB). 1977 Statistical Policy Directive No. 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting. Published online May 1978. Accessed April 4, 2025. <https://www2.census.gov/about/ombraceethnicityitwg/1978-statistical-policy-handbook.pdf>
189. US Census Bureau. Census Bureau Releases 2020 Census Population for More Than 200 New Detailed Race and Ethnicity Groups. Published online September 21, 2023. Accessed April 4, 2025. <https://www.census.gov/library/stories/2023/09/2020-census-dhc-a-race-overview.html>
190. Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey 2018 Full Year Consolidated Data File. Published online August 2020. Accessed March 22, 2023. https://www.meps.ahrq.gov/data_stats/download_data/pufs/h209/h209doc.shtml#Demographic253
191. Jarrín OF, Nyandegge AN, Grafova IB, Dong X, Lin H. Validity of Race and Ethnicity Codes in Medicare Administrative Data Compared With Gold-standard Self-reported Race Collected During Routine Home Health Care Visits. *Medical Care*. 2020;58(1):e1-e8.
192. Filice CE, Joynt KE. Examining Race and Ethnicity Information in Medicare Administrative Data. *Medical Care*. 2017;55(12):e170-e176. doi:10.1097/MLR.0000000000000608
193. Grafova IB, Jarrín OF. Beyond Black and White: Mapping Misclassification of Medicare Beneficiaries Race and Ethnicity. *Med Care Res Rev*. 2021;78(5):616-626. doi:10.1177/1077558720935733
194. Medicaid and CHIP Payment and Access Commission (MACPAC). *Availability of Race and Ethnicity Data for Medicaid Beneficiaries*.; 2022. Accessed February 25, 2023. https://www.macpac.gov/wp-content/uploads/2022/03/MACPAC-brief_Race-and-Ethnicity-Data-Availability.pdf
195. Mokdad AH, Bisignano C, Hsu JM, et al. The burden of diseases, injuries, and risk factors by state in the USA, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*. 2024;404(10469):2314-2340. doi:10.1016/S0140-6736(24)01446-6
196. US Census Bureau. American Community Survey, 2009–19 American Community Survey 1-Year Estimates, Variable: Health insurance coverage status by age for race and ethnicity groups. Tables B27001B-B27001E, B27001H, B27001I. Accessed using tidycensus package. Accessed June 13, 2024. <https://CRAN.R-project.org/package=tidycensus>
197. State Health Compare, State Health Access Data Assistance Center (SHADAC). *SHADAC Analysis of American Community Survey (ACS) Public Use Microdata Sample (PUMS) Files*. University of Minnesota Accessed April 24, 2024. statehealthcompare.shadac.org
198. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. National Academies Press; 2009:12696. doi:10.17226/12696
199. US Census Bureau. American Community Survey, 2009–2017 American Community Survey 5-Year Estimates. Tables AJ6XE001-AJ6XE031 using Census data portal. Accessed October 21, 2021. <https://data.census.gov/cedsci/>

200. US Census Bureau. Decennial Census, 2000, 2010. Tables H9X001-H9X049; using Census data portal. Accessed February 21, 2021. <https://data.census.gov/cedsci/>
201. Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM. **TMB** : Automatic Differentiation and Laplace Approximation. *J Stat Soft.* 2016;70(5). doi:10.18637/jss.v070.i05
202. Health Expenditures by State of Residence, 1991-2020. Accessed July 20, 2023. <https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountsstatehealthaccountsresidence>
203. Manson S, Schroeder J, Van Riper D, et al. National Historical Geographic Information System: Version 19.0. Published online 2024. doi:10.18128/D050.V19.0