

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Investigating Disparities in Treatment Quality and Decision-Making in the UC Health System

**Permalink**

<https://escholarship.org/uc/item/6hx2t0bd>

**Author**

Davidson, Jayso'n

**Publication Date**

2025

Peer reviewed|Thesis/dissertation

Investigating Disparities in Treatment Quality and Decision-Making in the UC Health System

by  
Jaysón Davidson

DISSERTATION  
Submitted in partial satisfaction of the requirements for degree of  
DOCTOR OF PHILOSOPHY

in  
Pharmaceutical Sciences and Pharmacogenomics

in the  
GRADUATE DIVISION  
of the  
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Signed by: *John A. (Tony) Capra* John A. (Tony) Capra  
FEE00BDBE74841E... Chair

Signed by: *Atul J. Butte* Atul J. Butte

Signed by: *Urmimala Sarkar* Urmimala Sarkar  
9AC52E3245D6494...

---

---

Committee Members

Copyright 2025

by

Jaysón Davidson

## Dedication and Acknowledgements

It's been a long journey, and words can't fully capture the gratitude I feel toward those who supported me and poured into me from the start of my Ph.D. journey to the end.

First, I thank God for blessing me with wonderful parents and a loving family who instilled in me the core values that guided me through this journey. I am grateful for the strength to pursue my Ph.D. and for the unwavering support system that helped me navigate its challenges.

To my high school Trigonometry teacher, Ms. Holly Kid, and my Pre-Calculus teacher, Mr. Bill, thank you for recognizing my potential in math and science and introducing me to the possibilities within STEM. Your encouragement set me on this path, and my parents pushed me to explore STEM further, ultimately leading me to where I am today.

Thank you to my parents, Stacey and Jaymes Davidson, for encouraging me to push my boundaries and providing me with the tools to use my gifts wisely. To my siblings, Jalesa, Jamea, and Jaymes, thank you for your love, guidance, and protection, and for keeping me grounded along the way. To my grandparents, I am deeply grateful for your time and wisdom, which have shaped my life and this journey. You instilled in me the values of faith and ambition and taught me to stay close to the Lord while dreaming higher than I ever imagined.

I also want to thank my friends, fraternity brothers, and colleagues for their unwavering support throughout this arduous process. Your encouragement lifted me through difficult times, allowing me to stay true to myself. Thank you for checking in on me when things got tough and providing a safe space to express my emotions and frustrations.

To Dr. Oluwatoyin Asojo, thank you for believing in me and giving me my start in research. You stepped in during a challenging time in my life when I struggled with daily migraines and showed me the life of a researcher. You changed my life and connected me with incredible people along the way. Thank you to Dr. Peter Hotez, whose encouragement inspired me to shoot for the stars and apply to UCSF. Without you, I would not have had the chance to meet the amazing friends and mentors that I have at UCSF. To Dr. Don Brunson, thank you for teaching me the cheat codes to networking and making a name for myself in the scientific community.

To my Chemistry and Chemical Biology cohort, I am so grateful to the Pandemic Cohort (GCOAT) for being such an amazing friend group and support system, and to the PSPG family for adopting and supporting me at UCSF. Special thanks to my close Bay Area and UCSF friends, Sydney, Tim, Jas, Zeynab, Jaela, Miguel, Trase, Alex, Phillip, Ben, Arka, Kelechi, and Pierrout, for making these past five years unforgettable.

At UCSF, I owe thanks to Dr. D'Anne Duncan for guiding my decision to choose UCSF and providing me with the tools to stay supported and build my network. To Dr. Jason Gestwicki, thank you for supporting me during my transition from the CCB to PSPG program. Dr. Deanna Kroetz, thank you for your mentorship during my first couple years at UCSF. To Dr. Marina Sirota, Dr. Vivek Rudrapatna, Dr. Thomas Peterson, and Dr. Akinyemi Oni-Orisan, thank you for guiding me through my qualifying exams and coaching me through the toughest phase of the Ph.D. process.

I am especially grateful to my mentor, Dr. Atul Butte. Watching your TED Talk in 2021 inspired me to switch disciplines into big data in health, and I am honored to have been mentored by you. Working in your lab has expanded my thinking and introduced me to some of the most brilliant minds. To Dr. Douglas Arneson, thank you for guiding me through my rotation project and

teaching me how to use EHRs to study health disparities. To Dr. Rohit Vashisht, thank you for stepping in when I needed guidance, coaching me, and helping me with research methodology, SQL, and R coding for my diabetes project. Your support made a huge difference. To Dr. Travis Zack and Dr. Irene Chen, thank you for working with me on my final Ph.D. project on Multiple Myeloma. You both stepped in when I felt lost and helped me design my study, learn new skills, and fulfill my research commitment to my UPSTREAM fellowship.

To my lab colleagues, Dr. Michelle Wang, Dr. Marie Binvignat, Dr. Dana Ludwing, Dr. Sanchita Bhattacharya, Dr. Madhumita Sushil, Dr. Christian Williams, Dr. Travis Zack, Dr. Kendra Radtke, Boris Oskotsky, Ayan Patel, Dr. Zicheng Hu, Albert Lee, and Dr. Jordan Guillot for always lending a helping hand, collaborating, and continuously exposing me to innovative research ideas.

A special thank you to Dr. Brenda Maio and Dr. Harry Shenghuan Sun for being amazing co-students throughout this Ph.D. journey. It has been a privilege to go through this process with you, and I look forward to seeing the impact we each make in the future of healthcare innovation.

Finally, I want to thank Dr. Tony Capra, Dr. Urmimala Sarkar, Dr. Irene Chen, and Dr. Chris Boone for mentoring and guiding me through the final stages of my Ph.D. program. Your mentorship made the completion of this journey truly worthwhile!

My unwavering faith and belief in myself have given me the strength to stay grounded and committed to making an impact. As I continue my journey in science, I hope to change lives, improve care, and reduce health disparities for all. The verse Matthew 6:34 "Therefore do not worry about tomorrow, for tomorrow will worry about its own things. Sufficient for the day is its own trouble" has been a guiding force, keeping me grounded and steadfast throughout my journey.

If there is one message, I want to leave with anyone reading this, it's this: never let anyone say you can't achieve something. Hard things may seem impossible until they're done, so dare to be bold and tackle challenges, because if someone must do it, it might as well be you!

Thank you to everyone who has been a part of this journey!

## AUTHOR CONTRIBUTIONS

Chapter 2 From Genes to Geography, from Cells to Community, from Biomolecules to Behaviors:  
the importance of Social Determinants of Health

Jaysón M. Davidson 1,2, Rohit Vashisht 2, and Atul J. Butte 2

1 Pharmaceutical Science and Pharmacogenomics Graduate Program, University of  
California San Francisco, San Francisco, CA 94143

2 Bakar Computational Health Sciences Institute, University of California San Francisco,  
San Francisco, CA 94143

Conceptualization, J.M.D. and A.J.B.; writing—original draft preparation, J.M.D.; writing—  
review and editing, J.M.D., R.V, and A.J.B.; visualization, J.M.D. All authors have read and  
agreed to the published version of the manuscript.

Chapter 3 Real-World Type 2 Diabetes Second-Line Treatment Allocation Among Patients

Jaysón M. Davidson<sup>1,2</sup>, Rohit Vashisht, Ph.D.<sup>1,2</sup>, Kendra Radtke, PharmD, Ph.D.<sup>1</sup>, Ayan Patel<sup>2</sup>,  
Suneil K. Koliwad, M.D., Ph.D.<sup>3,4</sup>, Atul J. Butte, M.D., Ph.D.<sup>1,2</sup>

1. Bakar Computational Health Sciences Institute, University of California San Francisco, San  
Francisco, CA

2. Center for Data-driven Insights and Innovation, University of California Health, Oakland, CA

3. Diabetes Center, University of California San Francisco, San Francisco, CA

4. Department of Medicine, Division of Endocrinology & Metabolism, University of California San Francisco, San Francisco, CA

Conceptualization, J.M.D., R.V., K.R and A.J.B.; writing—original draft preparation, J.M.D.; writing—review and editing, J.M.D., R.V., S.K.K, and A.J.B, visualization, J.M.D., Area Deprivation Index data, A.P. All authors have read and agreed to the published version of the manuscript.

#### Chapter 4 Investigating CAR-T Treatment Access for Multiple Myeloma Patients Using Real-World Evidence

Jaysón M. Davidson 1,2, Travis Zack, MD, Ph.D. 1,2,3, Kumar, Anupama 3, Ayan Patel. 2, Irene Y. Chen 4, Atul J. Butte, M.D., Ph.D. 1,2

1. Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA

2. Center for Data-driven Insights and Innovation, University of California Health, Oakland, CA

3. Department of Medicine, Division of Hematology and Oncology, University of California San Francisco, San Francisco, CA

4. Department of Computational Precision Health, University of California San Francisco, San Francisco, CA, University of California Berkeley, Berkeley, CA

Conceptualization, J.M.D., T.Z., I.Y.C., and A.J.B.; writing—original draft preparation, J.M.D.; writing—review and editing, J.M.D., A.K., T.Z., I.Y.C., and A.J.B., visualization, J.M.D., Area Deprivation Index data, A.P. All authors have read and agreed to the published version of the manuscript.

## ABSTRACT

### Investigating Disparities in Treatment Quality and Decision-Making in the UC Health System

Jaysón Davidson

Social Determinants of Health (SDOH) are non-medical factors including socioeconomic status, chronic disease burden, and mental health, that significantly influence healthcare access and outcomes. While these factors are important for understanding disease risk and disparities, they remain underrepresented in Electronic Health Records (EHRs). EHRs primarily capture relevant clinical data such as diagnoses, medications, and lab results. However, integrating SDOH offers a transformative approach to understanding health outcomes, disease risk, and healthcare disparities, particularly those shaped by socioeconomic and environmental conditions, through computational research and artificial intelligence.

This dissertation highlights the importance of incorporating SDOH into real-world evidence (RWE) studies to advance research on disease risk and health outcomes. Improvements in EHR infrastructure, including diagnosis codes, wearable technology, and census tract information, allow for deeper insights into health disparities. This work examines the challenges and opportunities related to integrating SDOH into EHRs, explores their growing role in RWE, and identifies pathways for equitable and actionable reporting of study outcomes.

Using de-identified data from the University of California Health Data Warehouse, this dissertation applies statistical modeling, machine learning, and generative artificial intelligence to investigate treatment disparities in two conditions: Type 2 diabetes and Multiple Myeloma. In the diabetes study, patients with lower socioeconomic status were more likely to receive less optimal second-line therapies, highlighting disparities in treatment pathways. In the Multiple Myeloma study, access to CAR-T therapy was significantly associated with treatment location and patient race and ethnicity, highlighting barriers to innovative therapies for diverse population groups.

This dissertation further examines the impact of SDOH on treatment decisions, unmet clinical needs, and barriers to equitable care. By integrating SDOH metrics with EHR data, this work informs data-driven strategies to improve access to advanced therapies and address healthcare disparities. These findings aim to inform policies that ensure all patient populations have fair and consistent access to advanced and innovative medical treatments.

# Table of Contents

CHAPTER 1 INTRODUCTION.....	1
1.1 DISSERTATION OVERVIEW .....	1
1.2 THE FOLLOWING CHAPTERS .....	4
INTRODUCTION REFERENCES .....	6
CHAPTER 2 FROM GENES TO GEOGRAPHY, FROM CELLS TO COMMUNITY, FROM BIOMOLECULES TO BEHAVIORS: THE IMPORTANCE OF SOCIAL DETERMINANTS OF HEALTH .....	11
2.1 ABSTRACT .....	11
2.2 INTRODUCTION.....	12
2.3 SOCIAL DETERMINANTS OF HEALTH.....	12
2.4 SDOH INTEGRATION INTO ELECTRONIC HEALTH RECORDS .....	15
2.5 THE FUTURE OF SDOH IN REAL WORLD EVIDENCE STUDIES .....	17
2.6 CONCLUSION .....	18
2.7 FUNDING.....	19
2.8 CONFLICTS OF INTEREST .....	20
2.9 TABLES AND FIGURES.....	22
REFERENCES .....	24
CHAPTER 3 REAL-WORLD TYPE 2 DIABETES SECOND-LINE TREATMENT ALLOCATION AMONG PATIENTS .....	29
3.1 ABSTRACT .....	29

3.2 INTRODUCTION.....	31
3.3 METHODS .....	32
3.4 RESULTS .....	36
3.5 DISCUSSION .....	38
3.6 FUNDING.....	42
3.7 CONFLICTS OF INTEREST .....	42
3.8 TABLES AND FIGURES.....	44
3.9 SUPPLEMENTAL MATERIALS .....	52
REFERENCES .....	71
CHAPTER 4 INVESTIGATING CAR-T TREATMENT ACCESS FOR MULTIPLE MYELOMA PATIENTS USING REAL-WORLD EVIDENCE .....	78
4.1 ABSTRACT .....	78
4.2 INTRODUCTION.....	80
4.3 METHODS .....	81
4.4 RESULTS .....	85
4.5 DISCUSSION .....	88
4.6 FUNDING.....	92
4.7 CONFLICTS OF INTEREST .....	92
4.8 TABLES AND FIGURES.....	94
4.9 SUPPLEMENTAL MATERIALS .....	106
REFERENCES .....	119
CHAPTER 5 CONCLUSION .....	127
CHAPTER 6 FUTURE WORK .....	130

## List of Tables

TABLE 2-1: INTERNATIONAL CLASSIFICATION OF DISEASE (ICD) Z-CODES FOR SDOH.....	22
TABLE 3-1: BASELINE CHARACTERISTICS OF TYPE 2 DIABETES PATIENTS WITH SECOND-LINE TREATMENTS.....	44
TABLE 3-2: RESULTS OF MULTINOMIAL REGRESSION MODEL.....	47
TABLE 3-3: LIKELIHOOD OF SECOND-LINE TREATMENT OPTIONS FOR RANDOMLY SELECTED PATIENTS .....	56
TABLE 3-4: GPT-4 ACCURACY EVALUATION FOR SYNTHETIC SDOH DATASET .....	62
TABLE 3-5: GPT-4 ACCURACY EVALUATION FOR MIMIC-III TRANSPORTATION ATTRIBUTES IDENTIFICATION .....	65
TABLE 3-6: GPT-4 ACCURACY EVALUATION FOR MIMIC-III HOUSING ATTRIBUTES IDENTIFICATION.....	66
TABLE 3-7: GPT-4 ACCURACY EVALUATION FOR MIMIC-III RELATIONSHIP ATTRIBUTES IDENTIFICATION.....	67
TABLE 3-8: GPT-4 ACCURACY EVALUATION FOR MIMIC-III PARENT AND SUPPORT ATTRIBUTES IDENTIFICATION .....	68

TABLE 3-9: GPT-4 ACCURACY EVALUATION FOR MIMIC-III EMPLOYMENT	
ATTRIBUTES IDENTIFICATION.....	68
TABLE 4-1: BASELINE CHARACTERISTICS OF MULTIPLE MYELOMA	
PATIENTS .....	94
TABLE 4-2: LOGISTIC REGRESSION MODEL RESULTS .....	97
TABLE 4-3: DISTRIBUTION OF RACE BY UC-LOCATION .....	108
TABLE 4-4: DISTRIBUTION OF INTERNATIONAL STAGING SYSTEM (ISS)	
DISEASE STAGE BY UC-LOCATION.....	108
TABLE 4-5: DISTRIBUTION OF NUMBER OF CRAB FEATURES BY UC-	
LOCATION.....	109
TABLE 4-6: ELIGIBILITY FOR UCSF PATIENTS WHO DISCUSSED CAR-T	
THERAPY .....	113
TABLE 4-7: ELIGIBILITY FOR UCSF PATIENTS WHO DID NOT DISCUSS CAR-T	
THERAPY .....	113
TABLE 4-8: SELECTED ANNOTATED GPT-4 PROMPT OUTPUTS FOR UCSF	
DEIDENTIFIED CLINICAL NOTE DATASET .....	114

## List of Figures

FIGURE 2-1: GROUPED SDOH FACTORS CATEGORIZED AT THE POPULATION AND INDIVIDUAL LEVEL .....	23
FIGURE 3-1: SECOND-LINE TREATMENT COHORT DESIGN.....	49
FIGURE 3-2: SECOND-LINE TREATMENT OPTIONS FOR TYPE 2 DIABETES PATIENTS .....	50
FIGURE 3-3: GPT-4 EXTRACTION AND IDENTIFICATION OF INDIVIDUAL-LEVEL SDOH FACTORS.....	51
FIGURE 3-4: COHORT SELECTION GLUCAGON-LIKE PEPTIDE 1 RECEPTOR AGONIST.....	53
FIGURE 3-5: COHORT SELECTION OF SGLT2-INHIBITORS.....	54
FIGURE 3-6: COHORT SELECTION OF DPP4-INHIBITORS.....	55
FIGURE 3-7: UCSF COHORT SELECTION FOR GLUCAGON-LIKE PEPTIDE 1 RECEPTOR AGONIST .....	57
FIGURE 3-8: UCSF COHORT SELECTION FOR SGLT2-INHIBITORS .....	58
FIGURE 3-9: UCSF COHORT SELECTION FOR DPP4-INHIBITORS .....	59
FIGURE 3-10: UCSF COHORT SELECTION FOR SULFONYLUREAS .....	60
FIGURE 3-11: GPT-4 PROMPT FOR SYNTHETIC SDOH DATASET .....	61
FIGURE 3-12: GPT-4 PROMPT FOR MIMIC-III SDOH DATASET.....	63

FIGURE 3-13: GPT-4 PROMPT FOR UCSF DEIDENTIFIED CLINICAL NOTE DATASET .....69

FIGURE 4-1: BASELINE DEMOGRAPHICS DISTRIBUTION FOR PATIENTS.....99

FIGURE 4-2: DISTRIBUTION OF UCSF MULTIPLE MYELOMA PATIENTS FOR CAR-T ELIGIBILITY .....102

FIGURE 4-2: DISTRIBUTION OF UCSF MULTIPLE MYELOMA PATIENTS FOR CAR-T ELIGIBILITY .....103

FIGURE 4-2: DISTRIBUTION OF UCSF MULTIPLE MYELOMA PATIENTS FOR CAR-T ELIGIBILITY .....104

FIGURE 4-3: DISTRIBUTION OF TOPIC MODELING FOR CAR-T DISCUSSIONS.....105

FIGURE 4-4: UCHDW COHORT SELECTION FOR MULTIPLE MYELOMA PATIENTS .....107

FIGURE 4-5: UCSF COHORT SELECTION FOR MULTIPLE MYELOMA PATIENTS .....110

FIGURE 4-6: GPT-4 PROMPT FOR UCSF DEIDENTIFIED CLINICAL NOTE DATASET .....111

# Chapter 1 Introduction

This chapter provides an overview of the dissertation. The primary motivation and research need of this dissertation will be summarized, followed by a concise description of each following chapter.

## 1.1 DISSERTATION OVERVIEW

The United States spent \$4.9 trillion on healthcare in 2023 and performed worse in life expectancy than any other high-income country in terms of access to care, care processes, administrative efficiency, equity, and health outcomes<sup>1,2</sup>. Life expectancy in the U.S. varies across population groups due to the adverse effects of geographic location, socioeconomic status, and racism against specific racial and ethnic groups<sup>3</sup>. In addition, within each racial and ethnic group, a patient's location affects their life expectancy; specifically, patients who live in rural or low socioeconomic zip codes may experience worsened life expectancies<sup>4</sup>. The differences in life expectancy are a downstream effect of access to healthcare, racial and ethnic background, socioeconomic status, and geographic location, which lead to disparate health outcomes for patients from diverse population groups<sup>5</sup>. These disparities have been widely studied over the past decade, primarily through the lens of race and ethnicity to understand differences in care<sup>6</sup>. However, this conventional thinking has shifted as clinicians recognize that broader social factors significantly impact health outcomes<sup>7</sup>. As a result, patient-centered care increasingly emphasizes addressing Social Determinants of Health (SDOH)<sup>7</sup>.

SDOH are the “conditions in which people are born, live, work, play, worship, and age, which affect a wide range of health, functioning, quality of life outcomes, and risks”<sup>8</sup>. SDOH influences

a patient's access to healthcare and treatments, health outcomes, and comorbidities by incorporating information related to socioeconomic status, insurance coverage, living conditions, access to healthy food, health literacy, and quality of care<sup>9</sup>. This information is fundamental to capturing, representing, and assessing the impact of social and environmental factors on differences in care and complex health challenges to improve the quality of care for patients<sup>9</sup>.

Adverse SDOH, such as lack of transportation, housing instability, and low income, contribute to health disparities among patients from diverse racial and ethnic backgrounds<sup>10-14</sup>. For example, a patient with greater disease severity may require more effective treatments, such as newly approved drugs but may not receive them due to cost or limited availability<sup>15,16</sup>. As a result, high-quality care is not always accessible. However, optimizing treatment decision-making can help reduce disparities in care.

To understand the drivers of treatment disparities and health outcomes, this study relies on the utilization of Electronic Health Records (EHRs), which are essential in real-world research on healthcare access. EHRs provide critical patient data, including treatments, conditions, drug responses, insurance, and demographics, enabling real-world evidence (RWE) studies. This data is pivotal to investigating treatment decisions, disease risk, care quality, and health outcomes, uncovering disparities in care<sup>17</sup>. In this study, EHRs serve as a key tool for exploring how social determinants of health (SDOH) influence treatment decisions and overall health outcomes.

RWE studies have been used to understand treatment decision-making, comparative effectiveness of treatments, and outcomes in chronic illnesses, diseases, and cancers<sup>18</sup>. These studies have addressed critical questions regarding patient outcomes and variations in care by race and ethnicity<sup>19,20</sup>. They have demonstrated the effectiveness of advanced treatments in diabetes and

cancer, highlighted treatment utilization patterns across various diseases, and examined patient outcomes, including mortality rates. By leveraging EHRs, RWE studies have also contributed to improving care for chronic illnesses, oncology, cardiovascular, and metabolic diseases<sup>21-26</sup>. While RWE has enhanced our understanding of treatment efficacy through observational studies, treatment decision-making, and comparative effectiveness studies, it remains limited in scope. Most studies assess differences by race and ethnicity but fail to fully capture disparities driven by social determinants of health (SDOH), largely due to limited diagnosis codes and insufficient SDOH-related data.

Previous RWE studies in diabetes and cancer were limited by the lack of comprehensive SDOH data and poor integration into EHRs<sup>27,28</sup>. As a result, there remains a significant lack of insight into how SDOH affects the quality of care for patients, especially for patients from diverse backgrounds. However, population-level metrics derived from health indices that incorporate SDOH factors such as income, education, housing, and neighborhood quality offer an opportunity to assess how social and environmental factors influence health outcomes across geographic areas.

The objective of this dissertation is to leverage both individual-level and population-level SDOH metrics to understand how SDOH impact the quality of care for patients from diverse population groups. To accomplish this, we will use de-identified data from the University of California Data Warehouse (UCDW) to associations between SDOH and treatment decisions. This dissertation explores two use cases: Type 2 diabetes, a prevalent chronic condition affecting 14.7% of U.S. adults, and Multiple Myeloma, a hematologic cancer that disproportionately impacts Black and Latinx patients.

These studies aim to bridge the knowledge gap for clinicians and researchers interested in leveraging SDOH to improve health equity for diverse populations. This dissertation presents practical approaches to integrating and utilizing SDOH data in EHRs. It identifies disparities in treatment decisions among patients diagnosed with both Type 2 diabetes, and Multiple Myeloma, two disease states that significantly impact diverse population groups. It also addresses real-world evidence gaps related to SDOH and differences in care between patients from low socioeconomic and high socioeconomic neighborhoods. This dissertation offers insights into improving access to care and advancing health equity across California.

## 1.2 THE FOLLOWING CHAPTERS

Chapter 2 “From Genes to Geography, from Cells to Community, from Biomolecules to Behaviors: The Importance of Social Determinants of Health” describes key opportunities and challenges associated with integration of Social Determinants of Health into Electronic Health Records. J.M.D contributions were conceptualization, writing—original draft preparation, writing—review and editing, and visualization.

Chapter 3 “Real-World Type 2 Diabetes Second-Line Treatment Allocation Among Patients” describes disparities associated with neighborhood location and second-line treatment allocation for Type 2 Diabetes patients in the University of California Health System. J.M.D contributions were conceptualization, writing—original draft preparation, writing—review and editing, and visualization.

Chapter 4 “Investigating CAR-T Treatment Access for Multiple Myeloma Patients Using Real-World Evidence” describes key barriers associated with CAR-T therapy access for Multiple Myeloma patients in the University of California Health System. J.M.D contributions were

conceptualization, writing—original draft preparation, writing—review and editing, and visualization.

Chapter 5 “Conclusions” provides a summary of the dissertation.

Chapter 6 “Future Work” provides an outlook for potential future work.

## INTRODUCTION REFERENCES

1. Office of the Actuary. (2023, December 13). National Health Expenditures 2022 highlights. CMS.gov. [https://www.cms.gov/newsroom/fact-sheets/national-health-expenditures-2022-highlights#\\_ednref1](https://www.cms.gov/newsroom/fact-sheets/national-health-expenditures-2022-highlights#_ednref1)
2. U.S. Census Bureau Quickfacts: United States. U.S. Census Bureau. (n.d.). <https://www.census.gov/quickfacts/fact/table/US/PST045223>
3. GBD US Health Disparities Collaborators (2022). Life expectancy by county, race, and ethnicity in the USA, 2000-19: a systematic analysis of health disparities. *Lancet* (London, England), 400(10345), 25–38. [https://doi.org/10.1016/S0140-6736\(22\)00876-5](https://doi.org/10.1016/S0140-6736(22)00876-5)
4. Dwyer-Lindgren, L., Baumann, M. M., Li, Z., Kelly, Y. O., Schmidt, C., Searchinger, C., La Motte-Kerr, W., Bollyky, T. J., Mokdad, A. H., & Murray, C. J. (2024). Ten Americas: a systematic analysis of life expectancy disparities in the USA. *Lancet* (London, England), 404(10469), 2299–2313. [https://doi.org/10.1016/S0140-6736\(24\)01495-8](https://doi.org/10.1016/S0140-6736(24)01495-8)
5. LaVeist TA, Pérez-Stable EJ, Richard P, et al. The Economic Burden of Racial, Ethnic, and Educational Health Inequities in the US. *JAMA*. 2023;329(19):1682–1692. doi:10.1001/jama.2023.5965
6. Munira Z. Gunja, Evan D. Gumas, and Reginald D. Williams II, U.S. Health Care from a Global Perspective, 2022: Accelerating Spending, Worsening Outcomes (Commonwealth Fund, Jan. 2023). <https://doi.org/10.26099/8ejy-yc74>
7. Braveman, P., & Gottlieb, L. (2014). The social determinants of health: it's time to consider the causes of the causes. *Public health reports* (Washington, D.C. : 1974), 129 Suppl 2(Suppl 2), 19–31. <https://doi.org/10.1177/00333549141291S206>

8. Healthy People 2030, U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. 2018. Available online: <https://health.gov/healthypeople/objectives-and-data/social-determinants-health> (accessed on 11 October 2024).
9. Davidson, J., Vashisht, R., & Butte, A. J. (2022). From Genes to Geography, from Cells to Community, from Biomolecules to Behaviors: The Importance of Social Determinants of Health. *Biomolecules*, 12(10), 1449. <https://doi.org/10.3390/biom12101449>
10. Braveman, P., & Gottlieb, L. (2014). The social determinants of health: it's time to consider the causes of the causes. *Public health reports (Washington, D.C. : 1974)*, 129 Suppl 2(Suppl 2), 19–31. <https://doi.org/10.1177/00333549141291S206>
11. Riley W. J. (2012). Health disparities: gaps in access, quality and affordability of medical care. *Transactions of the American Clinical and Climatological Association*, 123, 167–174.
12. Akinyelure, O. P., Jaeger, B. C., Safford, M. M., Oparil, S., Carson, A. P., Sims, A., Hannon, L., 3rd, Howard, G., Muntner, P., & Hardy, S. T. (2024). Social Determinants of Health and Incident Apparent Treatment-Resistant Hypertension Among White and Black US Adults: The REGARDS Study. *Journal of the American Heart Association*, 13(10), e031695. <https://doi.org/10.1161/JAHA.123.031695>
13. Llamocca, E. N., Yeh, H. H., Miller-Matero, L. R., Westphal, J., Frank, C. B., Simon, G. E., Owen-Smith, A. A., Rossom, R. C., Lynch, F. L., Beck, A. L., Waring, S. C., Lu, C. Y., Daida, Y. G., Fontanella, C. A., & Ahmedani, B. K. (2023). Association Between Adverse Social Determinants of Health and Suicide Death. *Medical care*, 61(11), 744–749. <https://doi.org/10.1097/MLR.0000000000001918>

14. Chelak, K., & Chakole, S. (2023). The Role of Social Determinants of Health in Promoting Health Equality: A Narrative Review. *Cureus*, 15(1), e33425. <https://doi.org/10.7759/cureus.33425>
15. Karanth, S., Osazuwa-Peters, O. L., Wilson, L. E., Previs, R. A., Rahman, F., Huang, B., Pisu, M., Liang, M., Ward, K. C., Schymura, M. J., Berchuck, A., & Akinyemiju, T. F. (2024). Health Care Access Dimensions and Racial Disparities in End-of-Life Care Quality among Patients with Ovarian Cancer. *Cancer research communications*, 4(3), 811–821. <https://doi.org/10.1158/2767-9764.CRC-23-0283>
16. Olateju, O. A., Shen, C., & Thornton, J. D. (2024). The Affordable Care Act and income-based disparities in health care coverage and spending among nonelderly adults with cancer. *Health affairs scholar*, 2(5), qxae050. <https://doi.org/10.1093/haschl/qxae050>
17. Nelson C.A., Butte A.J., Baranzini S.E. Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. *Nat. Commun.* 2019;10:1–7. doi: 10.1038/s41467-019-11069-0.
18. Dang A. (2023). Real-World Evidence: A Primer. *Pharmaceutical medicine*, 37(1), 25–36. <https://doi.org/10.1007/s40290-022-00456-6>
19. Giordano S. H. (2015). Comparative effectiveness research in cancer with observational data. *American Society of Clinical Oncology educational book*. American Society of Clinical Oncology. Annual Meeting, e330–e335. [https://doi.org/10.14694/EdBook\\_AM.2015.35.e330](https://doi.org/10.14694/EdBook_AM.2015.35.e330)
20. Palmer, S. C., Tendal, B., Mustafa, R. A., Vandvik, P. O., Li, S., Hao, Q., Tunnicliffe, D., Ruospo, M., Natale, P., Saglimbene, V., Nicolucci, A., Johnson, D. W., Tonelli, M., Rossi, M. C., Badve, S. V., Cho, Y., Nadeau-Fredette, A. C., Burke, M., Faruque, L. I.,

- Lloyd, A., ... Strippoli, G. F. M. (2021). Sodium-glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ (Clinical research ed.)*, 372, m4573. <https://doi.org/10.1136/bmj.m4573>
21. Patorno, E., Schneeweiss, S., & Wang, S. V. (2020). Transparency in real-world evidence (RWE) studies to build confidence for decision-making: Reporting RWE research in diabetes. *Diabetes, obesity & metabolism*, 22 Suppl 3(Suppl 3), 45–59. <https://doi.org/10.1111/dom.13918>
  22. Khunti, K., Almalki, M., Chan, J. C. N., & Amod, A. (2023). The Role of Real-World Evidence in Treatment Decision-Making, Regulatory Assessment, and Understanding the Perspectives of People with Type 2 Diabetes: Examples with Gliclazide MR. *Diabetes therapy : research, treatment and education of diabetes and related disorders*, 14(10), 1609–1625. <https://doi.org/10.1007/s13300-023-01458-6>
  23. Ryan, D. H., Lingvay, I., Deanfield, J., Kahn, S. E., Barros, E., Burguera, B., Colhoun, H. M., Cercato, C., Dicker, D., Horn, D. B., Hovingh, G. K., Jeppesen, O. K., Kokkinos, A., Lincoff, A. M., Meyhöfer, S. M., Oral, T. K., Plutzky, J., van Beek, A. P., Wilding, J. P. H., & Kushner, R. F. (2024). Long-term weight loss effects of semaglutide in obesity without diabetes in the SELECT trial. *Nature medicine*, 30(7), 2049–2057. <https://doi.org/10.1038/s41591-024-02996-7>
  24. Wang, C. P., Zylberberg, H. M., Borman, Z. A., Engelman, S., Yanes, R., Hirten, R. P., Sands, B. E., Cohen, B. L., Ungaro, R. C., & Rao, B. B. (2023). Impact of Care in an Interdisciplinary Inflammatory Bowel Disease Specialty Clinic on Outcomes in Patients

- Insured with Medicaid. *Journal of clinical gastroenterology*, 57(9), 908–912.  
<https://doi.org/10.1097/MCG.0000000000001769>
25. Lingvay, I., Deanfield, J., Kahn, S. E., Weeke, P. E., Toplak, H., Scirica, B. M., Rydén, L., Rathor, N., Plutzky, J., Morales, C., Lincoff, A. M., Lehrke, M., Jeppesen, O. K., Gajos, G., Colhoun, H. M., Cariou, B., Ryan, D., & SELECT Trial Investigators (2024). Semaglutide and Cardiovascular Outcomes by Baseline HbA1c and Change in HbA1c in People With Overweight or Obesity but Without Diabetes in SELECT. *Diabetes care*, 47(8), 1360–1369. <https://doi.org/10.2337/dc24-0764>
  26. Madhumita Sushil, Travis Zack, Divneet Mandair, Zhiwei Zheng, Ahmed Wali, Yan-Ning Yu, Yuwei Quan, Dmytro Lituiev, Atul J Butte, A comparative study of large language model-based zero-shot inference and task-specific supervised classification of breast cancer pathology reports, *Journal of the American Medical Informatics Association*, Volume 31, Issue 10, October 2024, Pages 2315–2327, <https://doi.org/10.1093/jamia/ocae146>
  27. Cook, L. A., Sachs, J., & Weiskopf, N. G. (2021). The quality of social determinants data in the electronic health record: a systematic review. *Journal of the American Medical Informatics Association : JAMIA*, 29(1), 187–196. <https://doi.org/10.1093/jamia/ocab199>
  28. Li, C., Mowery, D. L., Ma, X., Yang, R., Vurgun, U., Hwang, S., Donnelly, H. K., Bandhey, H., Senathirajah, Y., Visweswaran, S., Sadhu, E. M., Akhtar, Z., Getzen, E., Freda, P. J., Long, Q., & Becich, M. J. (2024). Realizing the potential of social determinants data in EHR systems: A scoping review of approaches for screening, linkage, extraction, analysis, and interventions. *Journal of clinical and translational science*, 8(1), e147. <https://doi.org/10.1017/cts.2024.571>

# **Chapter 2 From Genes to Geography, from Cells to Community, from Biomolecules to Behaviors: the importance of Social Determinants of Health**

## 2.1 ABSTRACT

Much scientific work over the past few decades has linked health outcomes and disease risk to genomics, to derive a better understanding of disease mechanisms at the genetic and molecular level. However, genomics alone does not quite capture the full picture of one's overall health. Modern computational biomedical research is moving in the direction of including social/environmental factors that ultimately affect quality of life and health outcomes at both the population and individual level. The future of studying disease now lies at the hands of the social determinants of health (SDOH) to answer pressing clinical questions and address healthcare disparities across population groups through its integration into electronic health records (EHRs). In this perspective article, we argue that the SDOH are the future of disease risk and health outcomes studies due to their vast coverage of a patient's overall health. SDOH data availability in EHRs has improved tremendously over the years with EHR toolkits, diagnosis codes, wearable devices, and census tract information to study disease risk. We discuss the availability of SDOH data, challenges in SDOH implementation, its future in real-world evidence studies, and the next steps to report study outcomes in an equitable and actionable way.

## 2.2 INTRODUCTION

Understanding disease at a molecular level has really pushed into the field of genetics, which has been the major basis for studying disease risk over the past two decades. Researchers have strongly believed that disease risk was generally conferred to genetic inheritance, and now more recently to specific rare and common mutations<sup>1,2</sup>. Using tools in molecular and cellular biology, researchers, and medical providers can investigate many diseases and conditions. But the results of previous investigations have shown that disease risk is too complex to model using genetics or molecules alone. Indeed, genetics, social and environmental factors including socioeconomic status, geolocation, age, and racial and ethnic background plays a role in disease risk across different population groups<sup>3</sup>. The growing evidence increasingly indicates the importance of accounting for the social and environmental factors that are likely to affect health outcomes. While Dr. Phil Bourne, whom this special issue honors, is certainly known for his work in computational methodologies and structural biology, he also understood the importance of external influences on health and called for better methods to measure and “describe individuals' activity spaces and exposure to the built, natural, social, and economic environments that influence behaviors and health outcomes”<sup>4</sup>.

## 2.3 SOCIAL DETERMINANTS OF HEALTH

Social Determinants of Health (SDOH) are one of the ways to capture, represent, and assess the impact of social and environmental factors in clinical research and to improve patient care. “SDOH are the conditions in which people are born, live, and work, play, worship, and age, that affect a wide range of health, functioning, and quality-of-life out-comes and risks”<sup>5</sup>. A patient's SDOH can be used to estimate their access to healthcare, treatments, positive or negative health outcomes and

comorbidities by using information related to an individual's health including alcohol and tobacco usage, socio-economic status, insurance status, living situation, access to healthy foods, and access to health literacy and quality<sup>5</sup>. The main components of SDOH commonly gathered in medicine are grouped into 5 domains: economic stability, education access and quality, health care access and equity, social and community context, and neighborhood and built environment<sup>5</sup>. Though noted separately, each domain is interconnected to match the complexity of SDOH variables to represent SDOH at both a population and individual level<sup>6,7</sup>.

Population-level SDOH measures are heavily reliant on census tract information derived from the United States Census Bureau (U.S Census). Census tracts is an indicative of geographical area, which is small, relatively permanent statistical subdivisions of a county providing information on demographic and housing estimates, occupation codes, industry codes, product and service codes, and material/fuel codes<sup>8</sup>. Census tracts have surveys such as the American community survey, decennial census, economic surveys, population estimates, public sector, and economic census that can be leveraged to assess the overall impact of socioeconomic parameters on the health and wellbeing of patients in a given healthcare system at a given geographical location. Census tract information is gathered by assigning each person, household, housing unit, institution, farm, business establishment, or other responding entity to a specific location, and then assigning that location to a zip code tabulation area appropriate to the census or sample survey by way of geocoding<sup>8,9</sup>. “The geocoding process assures that the Census Bureau can provide correct counts for small geographic entities, and that both the Census Bureau and data users can accumulate the data for small entities to provide totals for larger geographic entities such as zip code areas”<sup>8</sup>. Census tract information has been used to develop indices that directly explain SDOH of people by using their zip code location to develop the Area Deprivation Index, Social Vulnerability Index,

and Modified Retail Food Index<sup>10-13</sup>. Indices that use census tract information often categorize data by socioeconomic status, location, and education to calculate the deprivation or vulnerability of people residing in a location.

SDOH is utilized in clinical care and research studies by way of Electronic Health Records (EHR's) which is the primary way to capture real-world data from provider to patient encounters in a health system<sup>14</sup>. EHR's provide a unique opportunity to study the relationship between SDOH and the management and outcomes of clinical diseases through real-world data (RWD). RWD captured in EHR's is used to develop real-world evidence (RWE) studies that analyze data and inform providers about causes of different treatment strategies, disease risk, quality of life, and outcomes for different patients and populations. RWE studies often contain diverse patient populations that are representative of real patients' health where common SDOH are collected. Prior to EHR's, SDOH was primarily captured by population level questionnaires administered by the U.S. census, or through direct questionnaires administered in clinical trials. However, innovation of EHR's has provided us with patient derived data to help us understand social & lifestyle factors of patients. SDOH data coupled with questionnaires, and clinical data in EHR's could be used to enable precision medicine studies on healthcare access and health outcomes, by linking with data about treatments, disease conditions, drug response, insurance status and demographics.

Although classification of SDOH at the individual or patient level is becoming increasingly standardized for operational and clinical research purposes, a current challenge in the wide adoption of SDOH in real-world evidence studies is due to missing data, HIPAA regulations, and quality control issues that severely limits the amount of available data to answer clinical questions

with higher precision<sup>15-18</sup>. Therefore, the role SDOH plays in various chronic illnesses and diseases is ill defined but has the potential to address population and individual specific questions in the future. Research shows that “public health goals cannot be realized without addressing underlying SDOH” that contribute to disparities and outcomes<sup>19,20</sup>. Therefore, healthcare focus should also include SDOH in addition to race/ethnicity due to inconsistencies of clinical data. Plethora of research reveals numerous socioeconomic parameters potentially accelerating disease risk, especially among minorities<sup>20</sup>. We must improve our understanding of the impact of SDOH on disease risk by investigating the different roles that SDOH plays on patients, population groups, healthcare providers, healthcare access, and health outcomes (**Figure-1**).

#### 2.4 SDOH INTEGRATION INTO ELECTRONIC HEALTH RECORDS

In EHR databases across the country there is incompleteness in SDOH data that have led previous real world evidence studies to use only race/ethnicity, gender, and age as measures of SDOH. In theory, those demographics can provide context but cannot capture the full picture of one’s overall health. In an attempt to capture SDOH effectively in EHR’s, efforts to map de-identified patients’ information to census tracts has been extremely important in providing researchers the ability to use evidence based SDOH to answer clinical questions. However, SDOH is often just captured in clinical notes, without structured coding and we need better methods to get SDOH data trapped in notes. Currently, social aspects found in clinical notes vary across EHR databases in the country, however the Institute of Medicine (IOM) has worked aggressively on identifying SDOH domains to be suggested for use in EHR’s for academic research purposes<sup>21</sup>. The type of data suggested includes sociodemographic domains, psychological domains, behavioral domains, and individual-level social relationships and living conditions domains. Our field is still in its early days of

extracting specific SDOH information and mapping such data to and from EHR's, but alongside social indices, other structured data elements, such as insurance status, can be used to understand a patient's socioeconomic status. The indices can be used to answer questions related to SDOH and in cooperation with EHR's to understand surgical outcomes, drug distribution, health outcomes and hospital re-admissions. As time progresses, we will effectively utilize more SDOH data in EHR's and real-world evidence studies.

EHR toolkits offer precise categorization of SDOH captured from census tracts for use in real world evidence studies and SDOH standardized vocabularies are offered through these toolkits to map data to census tracts in EHR databases but are not widely adopted yet. A popular toolkit called the PhenX toolkit offers ontologies such as Health Insurance Coverage, food insecurity, air quality index, wealth, job insecurity, food swamp, and more to increase SDOH measures used in studies<sup>23</sup>. The SDOH ontologies were precisely picked to enable high qualitative measures that will increase statistical power of studies<sup>23</sup>. The International Classification of Diseases (ICD) coding system has incorporated specific ICD-Z codes for SDOH that refer to problems related to education & literacy, housing and economic circumstances, social environment, upbringing, primary support group, psychosocial circumstances, and occupational exposure to risk factors as listed in **(Table 1)**. Although important, the Z-codes are listed as non-diagnosis codes rather than disease specific codes, they all play a major part in understanding the causes of health disparities in communities. At this point, there are several SDOH standardized vocabulary and ontologies to use, but challenges occur when the utilization of these codes and with capturing these codes and related data from patients in an effective way to use for research. Until EHR databases increase their SDOH data availability for SDOH-specific studies, navigating this field will remain to be tricky.

When SDOH is made available for research, one can start to model and predict occurrences across a wide range of diseases, relating SDOH elements to both numerical measurement tests, diagnostics, and other health outcomes. It will be important to choose diseases that are easy to define, in terms of severity, prevalence, and incidence of the condition, as well as diseases that have been known to be prominent in lower socioeconomic areas. With SDOH mapped to and from EHR's we can understand what causes differences in prevalence and incidence of disease between different population groups.

## 2.5 THE FUTURE OF SDOH IN REAL WORLD EVIDENCE STUDIES

The next step in biomedical informatics is to have a streamlined method of mapping census tract information to EHR's that can be followed by each institution. In structured data, there needs to be a unified goal to create more diverse databases by providing healthcare access to underserved populations to ensure that SDOH studies are well balanced to produce answers backed by realistic outcomes. In the future we could increase the specific versatility of SDOH for medical studies by developing a universal “Medical Grade Social Index” that could contain the specific SDOH information that is most directly related to clinical outcomes, to best answer research questions. The power of this data will create knowledge for gaps in healthcare that will implicate the need to incorporate brick and mortar need-based healthcare services in deprived communities.

The future looks bright for even newer sources of data in SDOH. Although new, patient derived data has allowed us to obtain SDOH data beyond EHR's by using health measures including wearable fitness monitors, smart phones, step trackers, food trackers and derived data from telemedicine. Tracking information from patients allows us to understand patients' daily activity that lead to chronic illnesses and diseases and can potentially lead to eliminating them. But

currently patient derived data offers very little information across a large population. Luckily, the adoption of mHealth has increased SDOH measures by utilizing mobile technology to improve health goals across communities. mHealth is medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices as defined by the World Health Organization<sup>24</sup>. mHealth is not widely adopted yet but we can use patient derived data in the future to estimate a patient's health/risk of clinical diseases and provide better patient centered health approaches to improve overall individual health. How mHealth and digital devices will feed into SDOH is still an active area of research. Other newer sources of SDOH related data will also include genomic information for large populations uploaded into EHR databases, such as being done at various institutions around the country<sup>25</sup>.

Thus far, questionnaires, population surveys, and patient reported outcomes are heavily used in SDOH and epidemiological research studies because it gives researchers the opportunity to answer clinical research questions when the EHR database system does not itself contain a full picture of SDOH data. Because of inconsistencies in SDOH data availability, the field has been using questionnaire-based data to answer in-depth questions related to SDOH with EHR data to understand surgical outcomes, drug distribution, health outcomes and hospital readmissions<sup>26-29</sup>.

## 2.6 CONCLUSION

Communication of real-world evidence studies, especially related to disparities in health, has the potential to create political conflict, medical mistrust, and harm to marginalized communities. Therefore, it is important to effectively communicate results in a manner that is both understood by clinicians and patients from all paths of life so that the knowledge of health outcomes are more

accessible. More importantly, including a diverse set of researchers to harness the power of this data is important for leveraging different mindsets to understand SDOH clinical questions in a different context. Since we know that a large part of health disparities occurs in communities under-represented in medicine, the best way to relate to these communities is by having people who understand the vernacular, culture, and the patience needed to communicate risk, health literacy, and health accessibility to the most affected communities.

All of this is an acknowledgement, and perhaps a frustration, that studying SDOH is likely to be harder than studying genes and molecules and cells. But the more we fine tune the idea and execution of SDOH in EHR databases across the country we will increase the opportunities to use precision medicine to target clinical diseases. Precision medicine will be used to accurately prescribe patients based on real world evidence of health outcomes and treatment patterns respective to different population groups. More importantly, it will give clinicians the ability to prescribe patients using patient centered approach methods derived from research. But throughout our work here, it is critical to remember that precision medicine means more than just genes and molecules and cells. The future implementation of SDOH will provide greater precision of treatments based on an array of demographics, lifestyle factors, and environmental factors, all of which are likely to make more of a difference for our patients than any given measured cell or base-pair in their genome.

## 2.7 FUNDING

Research reported in this publication was supported by the UCSF Bakar Computational Health Sciences Institute, and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1 TR001872, and the Food and Drug

Administration, through U01 FD005978 to the UCSF–Stanford Center of Excellence in Regulatory Sciences and Innovation (CERSI). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or FDA. None of the study sponsors had any influence over the data interpretation or conclusions from this study.

## 2.8 CONFLICTS OF INTEREST

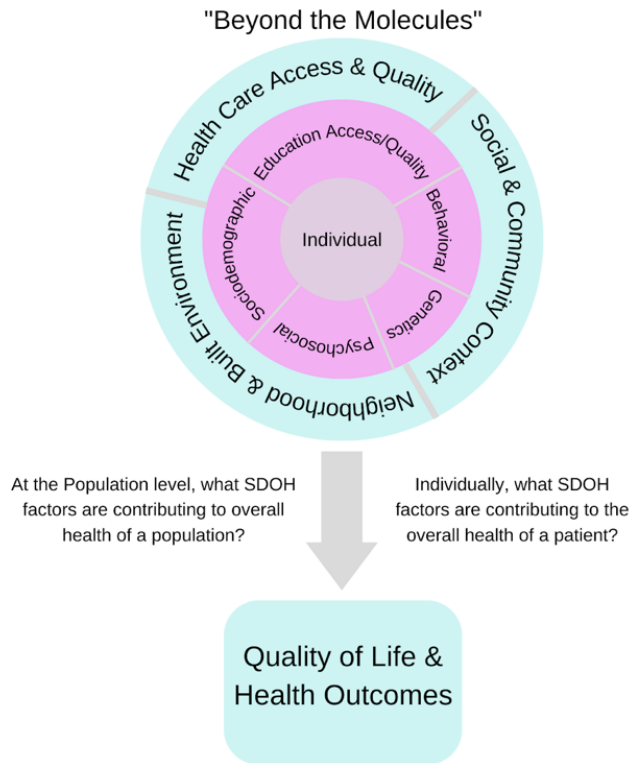
Atul Butte is a co-founder and consultant to Personalis and NuMedii; con-sultant to Mango Tree Corporation, and in the recent past, Samsung, 10x Genomics, Helix, Path-way Genomics, and Verinata (Illumina); has served on paid advisory panels or boards for Geisinger Health, Regenstrief Institute, Gerson Lehman Group, AlphaSights, Covance, Novartis, Genentech, and Merck, and Roche; is a shareholder in Personalis and NuMedii; is a minor share-holder in Apple, Meta (Facebook), Alphabet (Google), Microsoft, Amazon, Snap, 10x Genomics, Illumina, Regeneron, Sanofi, Pfizer, Royalty Pharma, Moderna, Sutro, Doximity, BioNtech, Invi-tae, Pacific Biosciences, Editas Medicine, Nuna Health, Assay Depot, and Vet24seven, and several other non-health related companies and mutual funds; and has received honoraria and travel reimbursement for invited talks from Johnson and Johnson, Roche, Genentech, Pfizer, Merck, Lilly, Takeda, Varian, Mars, Siemens, Optum, Abbott, Celgene, AstraZeneca, AbbVie, Westat, and many academic institutions, medical or disease specific foundations and associations, and health systems. Atul Butte receives royalty payments through Stanford University, for several patents and other disclosures licensed to NuMedii and Personalis. Atul Butte’s research has been funded by NIH, Peraton (as the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson Foundation, Leon Lowenstein Foundation, Intervalien Foundation, Priscilla Chan and Mark Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent past, the March

of Dimes, Juvenile Diabetes Research Foundation, California Governor's Office of Planning and Research, California Institute for Regenerative Medicine, L'Oreal, and Progenity.

## 2.9 TABLES AND FIGURES

**Table 2-1: International Classification of Disease (ICD) Z-codes for SDOH**

<b>ICD-10 CM Code</b>	<b>SDoH Categories</b>
Z55	Problems related to education and literacy
Z56	Problems related to employment and unemployment
Z57	Occupational exposure to risk factors
Z58	Problems related to physical environment
Z59	Problems related to housing and economic circumstances
Z60	Problems related to social environment
Z62	Problems related to upbringing
Z63	Other problems related to primary support group, including family circumstances
Z64	Problems related to certain psychosocial circumstances
Z65	Problems related to other psychosocial circumstances



**Figure 2-1: Grouped SDOH factors categorized at the population and individual level**

At each level we are defining the SDOH factors that contribute to the overall health of a population, and overall health of a patient that causes the difference between a good outcome and a poor outcome.

## REFERENCES

1. Baptista P. V. (2005). Principles in genetic risk assessment. *Therapeutics and clinical risk management*, 1(1), 15–20. <https://doi.org/10.2147/tcrm.1.1.15.53606>
2. Hall, W. D., Morley, K. I., & Lucke, J. C. (2004). The prediction of disease risk in genomic medicine. *EMBO Reports*, 5 Spec No, S22–S26. <https://doi.org/10.1038/sj.embor.7400224>
3. Fraiman, Y. S., & Wojcik, M. H. (2021, January 1). The influence of social determinants of health on the genetic diagnostic odyssey: who remains undiagnosed, why, and to what effect? *Pediatric Research*. Springer Nature. <https://doi.org/10.1038/s41390-020-01151-5>
4. Breen, N., Berrigan, D., Jackson, J. S., Wong, D., Wood, F. B., Denny, J. C., Zhang, X., & Bourne, P. E. (2019). Translational Health Disparities Research in a Data-Rich World. *Health equity*, 3(1), 588–600. <https://doi.org/10.1089/heq.2019.0042>
5. US Department and Human Services, Secretary’s Advisory Committee on National Health Promotion & Disease Prevention Objectives for 2020 (2010)
6. Haire-Joshu, D., & Hill-Briggs, F. (2019, April 1). The Next Generation of Diabetes Translation: A Path to Health Equity. *Annual Review of Public Health*. Annual Reviews Inc. <https://doi.org/10.1146/annurev-publhealth-040218-044158>
7. Kind, A. J. H., & Buckingham, W. R. (2018). Making Neighborhood-Disadvantage Metrics Accessible — The Neighborhood Atlas. *New England Journal of Medicine*, 378(26), 2456–2458. <https://doi.org/10.1056/nejmp1802313>
8. Census Bureau, U. S. (1994). Census tracts and block numbering areas. Retrieved from <https://www2.census.gov/geo/pdfs/reference/GARM/Ch10GARM.pdf>

9. Kolak, M., Bhatt, J., Park, Y. H., Padrón, N. A., & Molefe, A. (2020). Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States. *JAMA Network Open*, 3(1). <https://doi.org/10.1001/jamanetworkopen.2019.19928>
10. Singh, G. K. (2003). Area Deprivation and Widening Inequalities in US Mortality, 1969-1998. *American Journal of Public Health*, 93(7), 1137–1143. <https://doi.org/10.2105/AJPH.93.7.1137>
11. Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2020). A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management*, 8(1). <https://doi.org/10.2202/1547-7355.1792>
12. Nagano, I., Wu, Z., Asano, K., & Takahashi, Y. (2011). Census Tract Level State Maps of the Modified Retail Food Environment Index (mRFEI). *National Center for Chronic Disease Prevention and Health Promotion Division*, 178(1–2), 134–142.
13. Knighton, A. J., Savitz, L., Belnap, T., Stephenson, B., & VanDerslice, J. (2016). Introduction of an Area Deprivation Index Measuring Patient Socio-economic Status in an Integrated Health System: Implications for Population Health. *EGEMs (Generating Evidence & Methods to Improve Patient Outcomes)*, 4(3), 9. <https://doi.org/10.13063/2327-9214.1238>
14. Nelson, C. A., Butte, A. J., & Baranzini, S. E. (2019). Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-11069-0>
15. The Office for Civil Rights (OCR), & Malin, B. (2012). *Guidance Regarding Methods for de-identification of protected health information in accordance with the Health Insurance*

- Portability and Accountability Act (HIPAA) Privacy Rule. *Health Information Privacy*, 1–32.
16. Shortreed, S. M., Cook, A. J., Coley, R. Y., Bobb, J. F., & Nelson, J. C. (2019, May 1). Challenges and opportunities for using big health care data to advance medical science and public health. *American Journal of Epidemiology*. Oxford University Press. <https://doi.org/10.1093/aje/kwy292>
  17. Rudrapatna, V. A., & Butte, A. J. (2020, February 3). Opportunities and challenges in using real-world data for health care. *Journal of Clinical Investigation*. American Society for Clinical Investigation. <https://doi.org/10.1172/JCI129197>
  18. Cantor, M. N., & Thorpe, L. (2018). Integrating data on social determinants of health into electronic health records. *Health Affairs*, 37(4), 585-590. <https://doi.org/10.1377/hlthaff.2017.1252>
  19. Marmot, M. (2005). Social determinants of health inequalities. *Lancet*, 365(9464), 1099–1104. [https://doi.org/10.1016/S0140-6736\(05\)71146-6](https://doi.org/10.1016/S0140-6736(05)71146-6)
  20. Thornton, P. L., Kumanyika, S. K., Gregg, E. W., Araneta, M. R., Baskin, M. L., Chin, M. H., ... Mangione, C. M. (2020). New research directions on disparities in obesity and type 2 diabetes. *Annals of the New York Academy of Sciences*, 1461(1), 5–24. <https://doi.org/10.1111/nyas.14270>
  21. Capturing social and behavioral domains and measures in electronic health records: Phase 2. (2015). *Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2* (pp. 1–351). National Academies Press. <https://doi.org/10.17226/18951>

22. World Health Organization. (2004). ICD-10 : international statistical classification of diseases and related health problems : tenth revision, 2nd ed. World Health Organization. *Radiology Management*, 33(3).
23. Hamilton, et al. (2011) The PhenX Toolkit: Get the Most From Your Measures. *American Journal of Epidemiology*, 174(3), 253-60.
24. eHealth, W.H.O.G.O.f., mHealth: new horizons for health through mobile technologies: second global survey on eHealth. 2011, World Health Organization: Geneva.
25. Akbari, P., Sosina, O.A., Bovijn, J. et al. Multiancestry exome sequencing reveals INHBE mutations associated with favorable fat distribution and protection from diabetes. *Nat Commun* 13, 4844 (2022). <https://doi.org/10.1038/s41467-022-32398-7>
26. Hughes, M. M., Wang, A., Grossman, M. K., Pun, E., Whiteman, A., Deng, L., ... Toblin, R. L. (2021). County-Level COVID-19 Vaccination Coverage and Social Vulnerability — United States, December 14, 2020-March 1, 2021. *MMWR Surveillance Summaries*, 70(12), 431–436. <https://doi.org/10.15585/mmwr.mm7012e1>
27. Azap, R. A., Paredes, A. Z., Diaz, A., Hyer, J. M., & Pawlik, T. M. (2020). The association of neighborhood social vulnerability with surgical textbook outcomes among patients undergoing hepatopancreatic surgery. *Surgery (United States)*, 168(5), 868–875. <https://doi.org/10.1016/j.surg.2020.06.032>
28. Ghirimoldi, F. M., Schmidt, S., Simon, R. C., Wang, C. P., Wang, Z., Brimhall, B. B., ... Shireman, P. K. (2021). Association of Socioeconomic Area Deprivation Index with Hospital Readmissions After Colon and Rectal Surgery. *Journal of Gastrointestinal Surgery*, 25(3), 795–808. <https://doi.org/10.1007/s11605-020-04754-9>

29. Rosenzweig, M. Q., Althouse, A. D., Sabik, L., Arnold, R., Chu, E., Smith, T. J., ... Schenker, Y. (2021). The Association between Area Deprivation Index and Patient-Reported Outcomes in Patients with Advanced Cancer. *Health Equity*, 5(1), 8–16. <https://doi.org/10.1089/hecq.2020.0037>

# Chapter 3 Real-World Type 2 Diabetes Second-Line Treatment Allocation Among Patients

## 3.1 ABSTRACT

**Objective:** This study aimed to evaluate the impact of socioeconomic disparities on the allocation of second-line treatments among patients with type 2 diabetes (T2D).

**Materials and Methods:** We conducted an observational study using real-world data from over 9 million patients across five University of California Health centers. The study included patients who initiated a second-line T2D medication after metformin, with hemoglobin A1c (HbA1c) measurements within  $\pm 7$  days of treatment initiation from 2012 through September 2024. Multinomial regression models assessed the association between socioeconomic status and second-line treatment choices. Additionally, we used the GPT-4 large language model with a zero-shot learning approach to analyze 270 clinical notes from 105 UCSF T2D patients using the inclusion criteria above. GPT-4 identified adverse social determinants of health (SDOH) within the notes and categorized them into six domains: transportation, housing, relationships, patients with children, support, and employment.

**Results:** Among 15,090 patients (56.7% male, 43.3% female; mean age 59.3 years; mean HbA1c 8.91%), second-line treatments included sulfonylureas (SUs;  $n = 6,732$ ), DPP4 inhibitors ( $n = 2,918$ ), GLP-1 receptor agonists ( $n = 2,736$ ), and SGLT2 inhibitors ( $n = 2,704$ ). Patients from lower socioeconomic neighborhoods were more likely to receive SUs over other medications:

DPP4i (OR = 0.96, [95% CI, 0.95-0.98]), GLP-1RA (OR = 0.94, [95% CI, 0.92-0.96]), SGLT2i (OR = 0.95, [95% CI, 0.93-0.97]). Using UCSF clinical notes for T2D patients, we identified adverse SDOH including housing (n=8), transportation (n=1), relationships (n=22), employment (n=12), support (n=1), and patients with children (n=25).

Conclusions: Socioeconomic factors influence second-line T2D treatment choices. Addressing these disparities is essential to ensuring equitable access to advanced T2D therapies.

## 3.2 INTRODUCTION

The annual prevalence of Type 2 diabetes (T2D) remains elevated, affecting approximately 11.3% of the U.S. population and 9.3% of individuals globally, including minority groups<sup>1,2</sup>. Present clinical treatment guidelines target achievement of glycated hemoglobin A1c (HbA1c) levels below 7% for most adults<sup>3,4,5</sup>. Metformin (MF) remains the first-line of therapy, however, newer second-line treatments are often needed when metformin alone fails to achieve glycemic control<sup>6-9</sup>. Factors, particularly social determinants of health (SDOH), which are the conditions in the environments where people are born, live, learn, play, worship, and age are known to impact a range of health, functioning, and quality of life outcomes<sup>9,10</sup>. Specific SDOH<sup>10</sup> such as neighborhood disadvantage and socioeconomic status, along with other adverse factors, have been previously studied and shown to significantly impact diabetes management<sup>11-14</sup>, medication adherence<sup>15</sup>, and behaviors such as physical activity and dietary patterns<sup>16-23</sup>. Real-World characterization studies show wide-spread heterogeneity in the choice of second-line treatments<sup>24-26</sup>. However, the knowledge about the possible influence of patient-level SDOH on the allocation of second-line treatments among patients with T2D remains limited.

We sought to systematically analyze the association between SDOH, specifically socioeconomic status (SES) as estimated by mapping of home location to the Area Deprivation Index (ADI)<sup>27</sup> and the allocation of second-line treatments among patients with T2D after MF monotherapy. The study leveraged clinical data from electronic health records (EHRs) of over 9 million patients across five academic medical centers within the University of California Health (UC Health) system.

### 3.3 METHODS

#### *Data*

Data for this study was extracted from the University of California (UC) Health System, which includes 20 health professional schools (6 medical schools), 5 academic health centers (UC San Francisco, UC Los Angeles, UC Davis, UC Irvine, and UC San Diego), and 12 hospitals. It is the eighth largest non-profit health system in the United States by revenue and has built a secure central data warehouse (UCHDW) for operational improvement, promotion of quality patient care, and enabling the next generation of clinical research<sup>28</sup>. The repository currently holds data securely on over 9 million patients seen since 2012. Every quarter, EHR data is extracted from each site and transformed into vendor neutral Observational Medical Outcomes Partnership (OMOP) common data model used widely across the United States and the world<sup>29</sup>. De-identification of the data has been completed to enable clinical research projects, under guidance from UC campus institutional review boards, privacy, and compliance. Research use of UCHDW is considered non-human subject's research.

#### *Study Population*

UCHDW data was extracted from patients diagnosed with T2D and had a documented prescription of second-line treatment in addition to existing MF therapy. Patients with T2D were identified using ICD-10 diagnosis code E11.\*, HbA1c lab measurements (**Supplemental Materials**), and T2D related drug prescriptions for four second-line treatment classes (GLP-1RA, SGLT2i, DPP4i, and SU) (**Supplemental Materials**), from EHRs spanning from year 2012 up to November 2023. To select the cohort of patients who received a second-line SU prescription, for example, we

excluded individuals who had a previous order of another second-line treatment prior to the prescription for an SU and those with documented type 1 diabetes or an atypical form of diabetes such as gestational diabetes. We also excluded patients who did not have an order for MF before receiving an SU, and those who lacked T2D diagnosis code prior to receiving the SU. We also required patients to have documented HbA1c measurements within (+7/-7d) of SU treatment initiation (index date), where the index date was the first prescription of SU (**Figure-1**). Finally, we excluded all patients whose HbA1c was < 6.5% and were under 17 years of age at the time of SU treatment initiation and those who did not have an ADI value reported at the index date. Similar exclusion and inclusion criteria were applied to select individuals with other second-line T2D medications. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines to ensure the quality of our observational study<sup>30</sup>.

#### *Area Deprivation Index (ADI)*

We utilized the ADI as a metric to estimate patient-level SDOH<sup>27</sup>. ADI serves as a valuable resource for identifying and assessing healthcare needs for clinical treatments among patients by providing a general estimate of SDOH based on home location census tract characteristics. This index integrates SDOH data from 17 different census tract variables sourced from the U.S. Census, with factors related to poverty, housing, employment, and education. ADI as a metric has been used in previous observational studies<sup>31-35</sup> to understand its SDOH role on health outcomes in other disease states. For this study, census tract-based SDOH data was calculated to create the ADI, which ranks areas by percentiles, distinguishing the most deprived areas with higher percentile scores from the least deprived areas with lower percentile scores<sup>27,36,37</sup>.

### *Study Variables*

To describe patient baseline characteristics, we used ADI, HbA1c, Age (in years), patient reported Gender (Male, Female), presence of primary insurance coverage, patient-reported Race, and Ethnicity. T2D severity was inferred as the average of HbA1c measurements taken within the (+7/-7d) window relative to the index date. Age (in years) was calculated relative to the index date from date of birth. Patients with primary insurance healthcare coverage were utilized as Medicare, Medicaid, or private insurance types. Patient reported race was categorized as White, Asian, Black or African American category, Native Hawaiian or Pacific Islander, American Indian or Alaska Native category, Multi-Race, Other Race, or Unknown. Patient reported ethnicity was categorized as Hispanic or Latino, Not Hispanic or Latino or Unknown.

### *Statistical Analysis*

Descriptive statistics were calculated first for each covariate and then for the overall patient population. Multinomial regression was used to estimate the association of ADI with second-line treatment allocation, adjusting for other covariates considered in this study<sup>38,39</sup>. For the multinomial regression, the four second-line treatment classes (GLP-1RA, SGLT2i, DPP4i, and SU) were considered as outcomes, with SU serving as the reference class. We adjusted for HbA1c, age (in years), gender (Male, Female), primary insurance coverage, race, and ethnicity by using a multinomial regression model<sup>40,41</sup>. Association and its strength was reported as odds ratio with 95% confidence Intervals (CI) including p-values indicating the statistical significance of how likely a patient was to receive a second-line treatment compared with SU when adjusted for study covariates. An Odds Ratio (OR) was considered significant if it's 95% CI did not span 1, and p-value < 0.05. An OR = 1 indicates no difference between SU and other second-line treatments, OR

> 1 indicates an increased likelihood of allocating other second-line treatments compared with SU, and OR < 1 indicates decreased likelihood of allocating other second-line treatments compared with SU. All calculations were performed using R statistical software version 3.6.3 (R Project for Statistical Computing).

### *SDOH Identification through Large Language Models*

To identify individual-level SDOH factors, we extracted T2D patients from the UCSF Deidentified clinical database using the inclusion and exclusion criteria outlined in the study population section of the methods. We excluded patients who did not have clinical notes associated with their diagnoses. We further refined the dataset by isolating sections of the clinical notes related to Social History and SDOH, excluding patients whose clinical notes did not contain relevant information. This resulted in a final cohort of 105 patients with 270 clinical notes (**Supplemental Materials**). To identify individual-level SDOH factors, we utilized the GPT-4 inference model with a zero-shot learning approach. GPT-4 was prompted to analyze the Social History and SDOH sections of the clinical notes written by physicians. The model's accuracy was first tested and validated using an annotated synthetic dataset and further evaluated with an annotated MIMIC-III dataset curated by Guevara et al<sup>42</sup>. Outputs from these sample datasets were compared against the ground truth to check for accuracy in identifying SDOH factors (**Supplemental Materials**). GPT-4 was then prompted with the task to identify SDOH factors within UCSF clinical notes, categorized into the following domains as outlined in the annotation guidelines by Guevara et al.: Transportation (Distance, Resource, Other), Housing (Poor, Undomiciled, Other), Relationship (Married, Partnered, Divorced, Widowed, Single), Parent, Support (Minus, Plus), Employment (Employed, Underemployed, Unemployed, Disability, Retired, Student)<sup>42</sup>. Adverse SDOH factors were

consolidated into broader categories: Adverse Transportation, Adverse Housing, Adverse Relationship, Patients with Children, Adverse Support, and Adverse Employment. The identified SDOH factors were visualized and plotted.

### 3.4 RESULTS

Our study included 15,090 patients (mean age 59.3 years, SD 13.4 years). **Table 1** presents a summary of treatment allocation and demographic information. Of these patients, 6,732 were prescribed SU, 2,704 were prescribed DPP4i, 2,918 were prescribed GLP-1RA, and 2,736 were prescribed SGLT2i after final inclusion and exclusion criteria. 56.7% of the patients identified as male, 43.3% female, and 46.6% identified as White. Primary Insurance coverage was distributed among patients as private (43.1%), Medicaid (30.1%), Medicare (25.0%), and veterans affairs (1.8%). The mean [SD] HbA1c values, reflecting T2D disease severity across second-line treatments, were: DPP4i (8.59% [1.85]), GLP-1RA (8.83% [2.06]), SGLT2i (8.62% [1.88]), and SU (9.19% [2.02]). Overall, the ADI for each second-line treatment allocation was skewed toward the least economically disadvantaged neighborhoods ( $ADI < 5$ ), totaling 52.4% of patients in the study cohort. Older patients ( $OR = 1.02$  [95% CI, 1.01-1.02]), as well those self-identifying as Asian ( $OR = 1.19$ , [95% CI, 1.05-1.35]), American Indian or Alaskan Native ( $OR = 1.33$ , [95% CI, 0.74-2.41]) were more likely to be prescribed a DPP4i as compared to individuals who identify as White adjusting for ADI (**Table 2**). Females ( $OR = 1.44$ , [95% CI, 1.31-1.58]) were more likely to be prescribed GLP-1RA compared to Males. Patients who identified as Unknown Race ( $OR = 1.81$ , [95% CI, 1.49-2.20]) were more likely to be prescribed SGLT2i when compared to Whites. While patients identifying as “Other” race ( $OR = 1.21$ , [95% CI, 1.06-1.38]) were more likely to be prescribed DPP4i when compared to White patients. Hispanic patients were less likely to be

prescribed DPP4i (OR = 0.73, [95% CI, 0.64-0.83]), GLP-1RA (OR = 0.78, [95% CI, 0.69-0.88]), and SGLT2i (OR = 0.90, [95% CI, 0.79-1.02]) when compared to non-Hispanic patients.

Our results also highlighted that HbA1c was a significant determinant of second-line treatment allocation, specifically preferring an order of SU after metformin initiation, aligning with known treatment guidelines<sup>6</sup>. With every unit increase in HbA1c percentage indicative of increased disease severity, patients were more likely to be prescribed an SU versus any of the other three second-line treatments DPP4i (OR = 0.90, [95% CI, 0.88-0.92]), GLP-1RA (OR = 0.90, [95% CI, 0.87-0.92]), SGLT2i (OR = 0.88, [95% CI, 0.85-0.90]). Additionally, our results highlighted that primary insurance coverage was significant in determining second-line treatment allocation. Patients with Medicaid as their primary insurance coverage were less likely to receive to GLP-1RA when compared to patients with private (OR = 2.39, [95% CI, 2.12-2.69]), Medicare (OR = 1.30, [95% CI, 1.12-1.51]), or Veterans Affairs insurance (OR = 2.26, [95% CI, 1.60-3.21]).

Interestingly, we found ADI to be a significant independent variable underlying allocation of four second-line treatments. As ADI increased from a relatively low value reflecting a least economic disadvantage to progressively higher values reflecting low socioeconomic neighborhoods, patients were more likely to be prescribed an SU versus the other three second-line treatments DPP4i (OR = 0.96, [95% CI, 0.95-0.98]), GLP-1RA (OR = 0.94, [95% CI, 0.92-0.96]), SGLT2i (OR = 0.95, [95% CI, 0.93-0.97]). As illustrated with randomly selected patients that followed the similar inclusion and exclusion criteria in our study but were not part of the model building (**Figure-2**), a Hispanic female patient aged 46 years with an HbA1c of 11.8% and Private insurance as their primary insurance living in a least disadvantaged neighborhood (ADI 1) is most likely to be prescribed a GLP-1RA as a second-line medication in addition to MF (38.5% likelihood). Whereas

a non-Hispanic White male patient aged 49 years with an HbA1c of 11.9% and Medicaid as their primary insurance who lives in an economically disadvantaged neighborhood (ADI 10) is most likely to be prescribed SU as a second-line treatment (67.2% likelihood). In another example, a Hispanic female patient aged 53 years with a 9.2% HbA1c and Medicaid insurance living in ADI 10 is most likely to be prescribed SU as a second-line treatment (61.0% likelihood).

Our results from prompt engineering using GPT-4 LLMs showed the identification of individual-level SDOH factors for 105 patients whose clinical notes contained social history. Among UCSF patients with clinical notes, 54 were prescribed SU, 21 were prescribed DPP4i, 15 were prescribed GLP-1RA, and 15 were prescribed SGLT2i. We identified the following present SDOH factors among these patients: adverse housing (n=8), adverse transportation (n=1), adverse relationships (n=22), adverse employment (n=12), adverse support (n=1), and patients with children (n=25) **(Figure-3)**.

### 3.5 DISCUSSION

In this retrospective cohort study, we evaluated the association of social determinants of health such as ADI on the allocation of second-line treatments among patients with T2D. We found that as socioeconomic disadvantage metrics ranged from relatively affluent to progressively socioeconomically disadvantaged neighborhoods, patients were more likely to receive an SU, a less clinically effective, but more cost-effective treatment<sup>3,6,8</sup>. Conversely, patients who lived in relatively least socioeconomic disadvantaged neighborhoods were more likely to receive newer, more expensive, and potentially more effective second-line treatments such as SGLT2i, DPP4i, and GLP-1RA. We also found that patients with Medicaid as their primary insurance, often associated with low-cost healthcare coverage, were less likely to receive these advanced second-

line therapies. This disparity underlying treatment allocation is important, given that SGLT2i and GLP-1RAs have clearly shown benefits in limiting development of serious comorbid conditions including CKD, cardiovascular events, heart failure, and obesity<sup>3,6</sup>. Our findings thus highlight a gap in the fair allocation of effective second-line treatments among patients with T2D.

Beyond underlying socioeconomic disparities in second-line treatment allocation, we also found a reliance favoring SU when the urgency for glucose lowering was highest. Indeed, in each comparative analysis of specific agents as second-line treatments for T2D, we found that the worse a patient's glycemic control, as measured by HbA1c, the more likely that patient was to receive an SU. This reliance is in line with the traditional prescribing patterns and older ADA guidelines that emphasize SU treatment as the first adjunct to MF for T2D. It also highlights the notion that older more established medication classes might be favored when A1c lowering is the top priority, regardless of SES. This fits with classic literature and ongoing supportive data indicating that optimal HbA1c lowering remains the most important factor in limiting the development of T2D-associated complications<sup>6</sup>. This finding is important because it highlights that patients with a high ADI are more likely to have a more severe A1c at second-line treatment initiation and are more likely to receive SU.

Our analysis also revealed disparities in second-line treatment allocation based on patient gender and age. First, we found that males were more likely to receive a DPP4i, or a GLP-1RA over SU compared to females, potentially consistent with the fact that adult males have higher T2D rates than females<sup>43</sup>. Advancing age was associated with a higher likelihood of receiving an SU over other second-line treatment choices. This finding may reflect a preference for SU among aged

patients to minimize risk of hypoglycemia, cognitive impairment, and autonomic symptoms during management of diabetes<sup>6</sup>.

Our study showed that patients prescribed SU as a second-line treatment had a higher prevalence of documented adverse SDOH factors compared to those prescribed GLP-1RA, SGLT2i, and DPP4i. More specifically, these patients faced challenges such as adverse housing, relationships, transportation issues, with many also having caregiving responsibilities for children. This suggests they may have been prescribed a cost-effective treatment rather than a newer second-line option due to non-medical factors. Interestingly, patients prescribed SGLT2i had a higher prevalence of documented adverse employment factors than those prescribed other treatments, indicating that some patients facing employment challenges are still being treated with newer drugs. However, the relatively low number of individual-level SDOH factors captured in our data provides only a partial view of the adverse contributors influencing a patient's second-line treatment prescription.

There are several limitations in the study. The health disparities associated with SES in the UCHDW, as highlighted in our study, were estimated using the ADI at the neighborhood level, which only provides an approximate measure of a patient's SES. Additionally, the inability to capture a significant amount of individual-level SES data and other SDOH parameters due to the absence of such data elements, limited a comprehensive evaluation of the patient population. We acknowledge that our understanding of SES at the individual level remains incomplete and could potentially be improved in future studies. The sample size of our data was also constrained, primarily due to the stringent cohort selection criteria employed in this study. Furthermore, our study does not account for the approval dates of newer drugs such as SGLT2i, DPP4i, and GLP-1RA which were introduced later during the study period. Lastly, the distribution of patients in

this study was slightly skewed towards least economically disadvantaged neighborhoods due to the high number of patients living in higher socioeconomic areas that are treated within the UC-Health System.

Our findings serve as a preliminary step toward a deeper exploration of how individual SES and SDOH can contribute to health disparities when data is collected in large quantities. We believe that differences in disease severity at second-line treatment allocation may stem from factors such as limited access to healthcare, financial constraints, low health literacy, and geographic location of a patient. Moreover, our findings suggest that disparities in treatment allocation may be influenced by treatment cost and unequal access to healthcare providers who prescribe newer, more effective medications (e.g., GLP-1RA). To further understand the disparities highlighted in our study, a comprehensive examination of high-quality individual-level SDOH factors is needed to effectively address the underlying issues pertaining to treatment selection and disease severity disparities.

In conclusion, our analysis highlights the association of SDOH with treatment choices, significantly impacting disease state and health outcomes. Our study shows growing associations of SDOH on clinical treatments, with and without adjusting for other covariates. Furthermore, our study shows that SDOH can be utilized to understand differences in care among patients from distinct population groups based on socioeconomic status. The recognition of SDOH disparities within the UC-Health system for T2D extends to healthcare systems across the United States. Understanding the role of SDOH, alongside considerations of disease severity, age, and gender, is crucial in addressing disparities in care among diverse population groups. A comprehensive understanding of treatment disparities will lead to the fair allocation of medical resources to

population groups in which disparities exist. Most importantly utilizing SDOH to understand differences in care for multiple disease could drive institutions to improve their delivery of care to disparate population groups. Therefore, we must encourage the generation of high-quality SDOH data at both the neighborhood and individual level to increase the viability of clinical-related questions across disease states.

### 3.6 FUNDING

Research reported in this publication was supported by the UCSF Bakar Computational Health Sciences Institute, and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI grant number UL1 TR001872, along with the Food and Drug Administration, through U01 FD005978 to the UCSF–Stanford Center of Excellence in Regulatory Sciences and Innovation (CERSI). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or FDA. None of the study sponsors had any influence over the data interpretation or conclusions of this study.

### 3.7 CONFLICTS OF INTEREST

Kendra Radtke is a current employee and shareholder of Genentech/Roche. Atul Butte is a co-founder and consultant to Personalis and NuMedii; consultant to Mango Tree Corporation, and in the recent past, Samsung, 10x Genomics, Helix, Pathway Genomics, and Verinata (Illumina); has served on paid advisory panels or boards for Geisinger Health, Regenstrief Institute, Gerson Lehman Group, AlphaSights, Covance, Novartis, Genentech, Merck, and Roche; is a shareholder in Personalis and NuMedii; is a minor shareholder in Apple, Meta (Facebook), Alphabet (Google), Microsoft, Amazon, Snap, 10x Genomics, Illumina, Regeneron, Sanofi, Pfizer, Royalty Pharma,

Moderna, Sutro, Doximity, BioNtech, Invitae, Pacific Biosciences, Editas Medicine, Nuna Health, Assay Depot, and Vet24seven, and several other non-health-related companies and mutual funds; and has received honoraria and travel reimbursement for invited talks from Johnson and Johnson, Roche, Genentech, Pfizer, Merck, Lilly, Takeda, Varian, Mars, Siemens, Optum, Abbott, Celgene, AstraZeneca, AbbVie, Westat, and many academic institutions, medical or disease-specific foundations and associations, and health systems. Atul Butte receives royalty payments through Stanford University for several patents and other disclosures licensed to NuMedii and Personalis. Atul Butte's research has been funded by NIH, Peraton (as the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson Foundation, Leon Lowenstein Foundation, Intervalien Foundation, Priscilla Chan and Mark Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent past, the March of Dimes, Juvenile Diabetes Research Foundation, California Governor's Office of Planning and Research, California Institute for Regenerative Medicine, L'Oreal, and Progenity.

### 3.8 TABLES AND FIGURES

**Table 3-1: Baseline Characteristics of Type 2 Diabetes patients with Second-Line Treatments**

Baseline Characteristics of type 2 diabetes patients with second-line treatment initiation between January 2012 - November 2023. \*Percentages within each ADI variable group add to 100% in each column.

Variable	No. (%) of patients					P-Value
	DPP4i (N=2918)	GLP-1RA (N=2736)	SGLT2i (N=2704)	Sulfonylurea (N=6732)	Total (N=15090)	
<b>Age</b>						
Mean (SD)	62.7 (13.4)	55.5 (13.7)	60.3 (13.4)	59.0 (12.9)	59.3 (13.4)	< 0.001
<b>HbA1c</b>						
Mean (SD)	8.59 (1.85)	8.83 (2.06)	8.62 (1.88)	9.19 (2.02)	8.91 (1.99)	< 0.001
<b>Gender</b>						
Female	1306 (44.8%)	1363 (49.8%)	996 (36.8%)	2870 (42.6%)	6535 (43.3%)	< 0.001
Male	1612 (55.2%)	1373 (50.2%)	1708 (63.2%)	3862 (57.4%)	8555 (56.7%)	
<b>Race</b>						
American Indian or Alaska Native	17 (0.6%)	22 (0.8%)	10 (0.4%)	34 (0.5%)	83 (0.6%)	< 0.001
Asian	550 (18.8%)	315 (11.5%)	498 (18.4%)	916 (13.6%)	2279 (15.1%)	
Black or African American	174 (6.0%)	238 (8.7%)	195 (7.2%)	539 (8.0%)	1146 (7.6%)	

Variable	No. (%) of patients					
	DPP4i (N=2918)	GLP-1RA (N=2736)	SGLT2i (N=2704)	Sulfonylurea (N=6732)	Total (N=15090)	P-Value
<b>Native Hawaiian or Other Pacific Islander</b>	29 (1.0%)	20 (0.7%)	29 (1.1%)	81 (1.2%)	159 (1.1%)	< 0.001
<b>Other Race</b>	560 (19.2%)	502 (18.3%)	480 (17.8%)	1535 (22.8%)	3077 (20.4%)	
<b>Unknown</b>	247 (8.5%)	332 (12.1%)	300 (11.1%)	430 (6.4%)	1309 (8.7%)	
<b>White</b>	1341 (46.0%)	1307 (47.8%)	1192 (44.1%)	3197 (47.5%)	7037 (46.6%)	
<b>Ethnicity</b>						
<b>Hispanic or Latino</b>	708 (24.3%)	794 (29.0%)	742 (27.4%)	2389 (35.5%)	4633 (30.7%)	< 0.001
<b>Not Hispanic or Latino</b>	1977 (67.8%)	1697 (62.0%)	1740 (64.3%)	3946 (58.6%)	9360 (62.0%)	
<b>Unknown</b>	233 (8.0%)	245 (9.0%)	222 (8.2%)	397 (5.9%)	1097 (7.3%)	
<b>Area Deprivation Index (ADI)</b>						
<b>1</b>	368 (12.6%)	308 (11.3%)	337 (12.5%)	519 (7.7%)	1532 (10.2%)	< 0.001
<b>2</b>	293 (10.0%)	262 (9.6%)	266 (9.8%)	551 (8.2%)	1372 (9.1%)	
<b>3</b>	323 (11.1%)	302 (11.0%)	298 (11.0%)	529 (7.9%)	1452 (9.6%)	
<b>4</b>	353 (12.1%)	360 (13.2%)	315 (11.6%)	771 (11.5%)	1799 (11.9%)	
<b>5</b>	317 (10.9%)	328 (12.0%)	336 (12.4%)	773 (11.5%)	1754 (11.6%)	
<b>6</b>	293 (10.0%)	336 (12.3%)	326 (12.1%)	947 (14.1%)	1902 (12.6%)	
<b>7</b>	295 (10.1%)	283 (10.3%)	250 (9.2%)	772 (11.5%)	1600 (10.6%)	

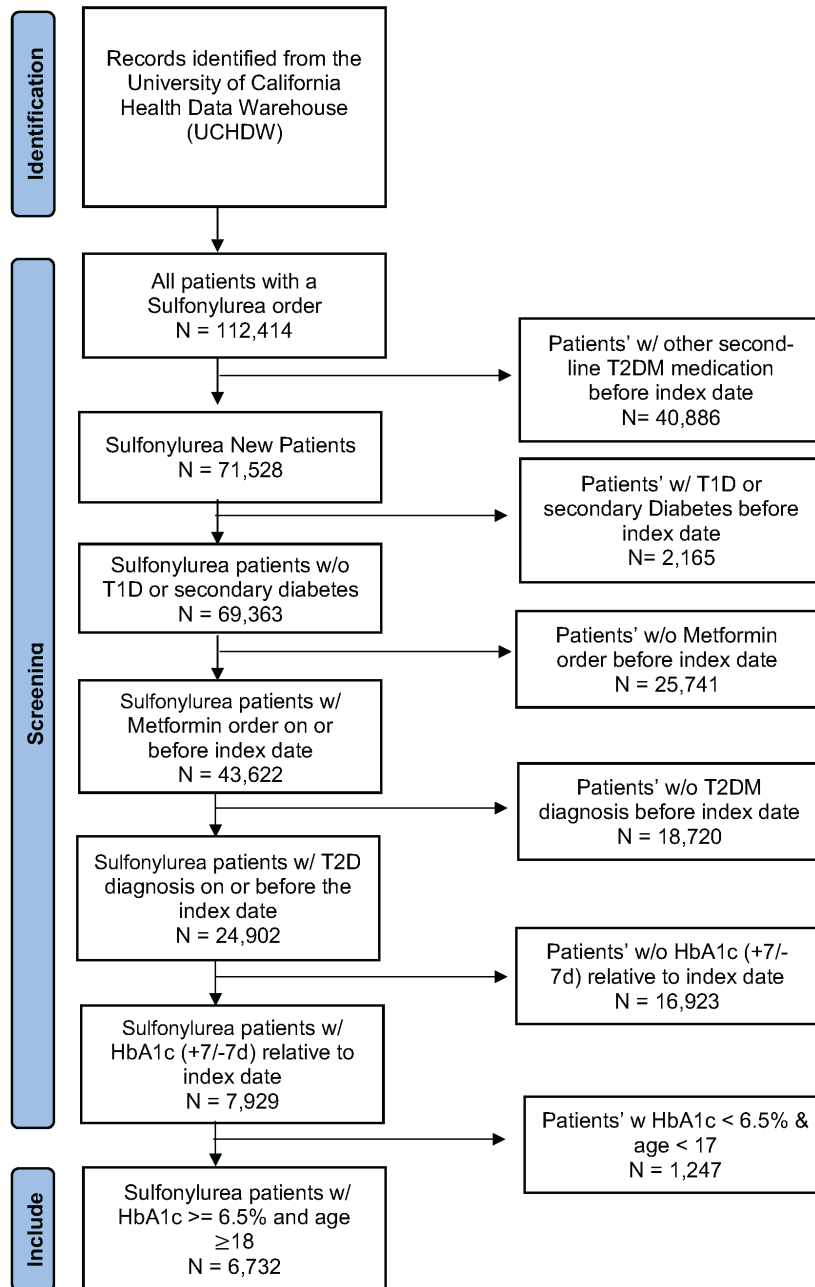
Variable	No. (%) of patients					
	DPP4i (N=2918)	GLP-1RA (N=2736)	SGLT2i (N=2704)	Sulfonylurea (N=6732)	Total (N=15090)	P-Value
<b>8</b>	236 (8.1%)	245 (9.0%)	213 (7.9%)	676 (10.0%)	1370 (9.1%)	< 0.001
<b>9</b>	231 (7.9%)	176 (6.4%)	208 (7.7%)	659 (9.8%)	1274 (8.4%)	
<b>10</b>	209 (7.2%)	136 (5.0%)	155 (5.7%)	535 (7.9%)	1035 (6.9%)	
<b>Primary Insurance</b>						
<b>Medicaid</b>	700 (24.0%)	607 (22.2%)	720 (26.6%)	2515 (37.4%)	4542 (30.1%)	< 0.001
<b>Medicare</b>	880 (30.2%)	495 (18.1%)	634 (23.4%)	1762 (26.2%)	3771 (25.0%)	
<b>Private</b>	1273 (43.6%)	1580 (57.7%)	1301 (48.1%)	2349 (34.9%)	6503 (43.1%)	
<b>Veterans Affairs</b>	65 (2.2%)	54 (2.0%)	49 (1.8%)	106 (1.6%)	274 (1.8%)	

**Table 3-2: Results of Multinomial Regression Model**

Odds Ratio table of patients receiving second-line treatment across patient characteristics with SU as the reference. An OR = 1 indicates no difference between SU and other drugs, OR > 1 indicates increased likelihood of other treatments being used as a second line therapeutic compared with SU and OR < 1 indicates decreased likelihood of other treatments compared with SU.

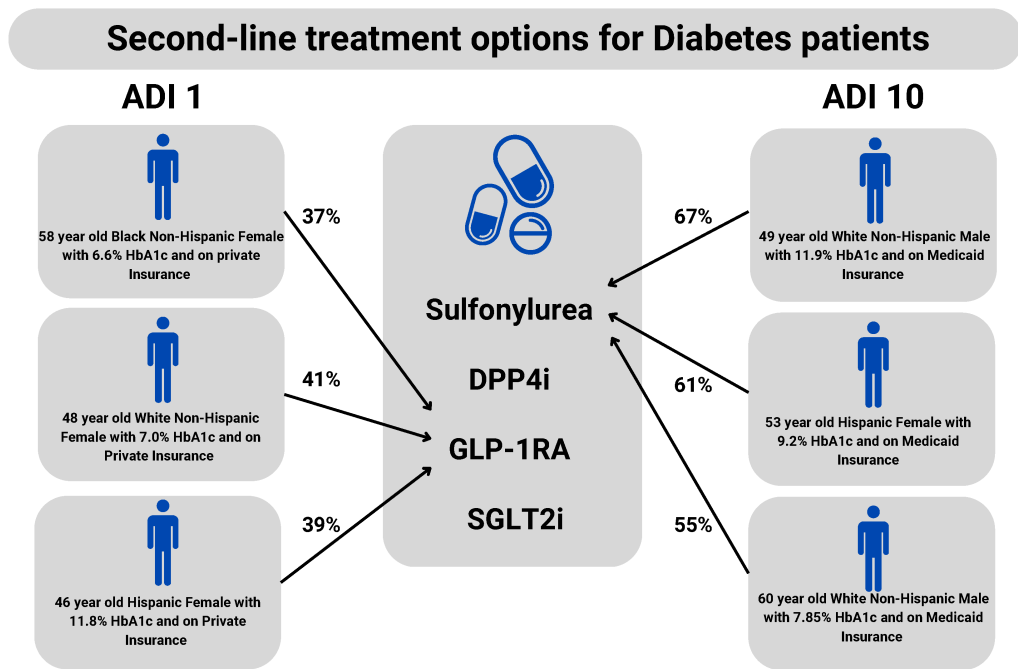
Variable	DPP4i		GLP-1RA		SGLT2i	
	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
<b>Area Deprivation Index (ADI)</b>	0.96 (0.95-0.98)	< 0.001	0.94 (0.92-0.96)	< 0.001	0.95 (0.93-0.97)	< 0.001
<b>HbA1c</b>	0.90 (0.88-0.92)	< 0.001	0.90 (0.87-0.92)	< 0.001	0.88 (0.85-0.90)	< 0.001
<b>Age</b>	1.02 (1.01-1.02)	< 0.001	0.98 (0.97-0.98)	< 0.001	1.01 (1.00-1.01)	0.001
<b>Gender</b>						
<b>Female</b>	1.10 (1.00-1.20)	0.043	1.44 (1.31-1.58)	< 0.001	0.79 (0.72-0.86)	< 0.001
<b>Race</b>						
<b>American Indian or Alaska Native</b>	1.33 (0.74-2.41)	0.345	1.52 (0.87-2.64)	0.138	0.85 (0.42-1.74)	0.662
<b>Asian</b>	1.19 (1.05-1.35)	0.008	0.62 (0.53-0.72)	< 0.001	1.25 (1.09-1.43)	0.001
<b>Black or African American</b>	0.81 (0.67-0.98)	0.031	1.03 (0.86-1.22)	0.759	1.06 (0.88-1.27)	0.533
<b>Native Hawaiian or Other Pacific Islander</b>	0.86 (0.56-1.33)	0.507	0.53 (0.32-0.87)	0.012	1.01 (0.65-1.56)	0.969
<b>Other Race</b>	1.21 (1.06-1.38)	0.004	1.04 (0.91-1.19)	0.552	1.04 (0.90-1.19)	0.604
<b>Unknown</b>	1.35 (1.10-1.65)	0.004	1.54 (1.27-1.86)	< 0.001	1.81 (1.49-2.20)	< 0.001
<b>Ethnicity</b>						
<b>Hispanic or Latino</b>	0.73 (0.64-0.83)	< 0.001	0.78 (0.69-0.88)	< 0.001	0.90 (0.79-1.02)	0.104
<b>Unknown</b>	1.02 (0.83-1.25)	0.859	0.92 (0.75-1.14)	0.459	0.90 (0.73-1.12)	0.346

Variable	DPP4i		GLP-1RA		SGLT2i	
	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
<b>Primary Insurance</b>						
<b>Medicare</b>	1.19 (1.05-1.36)	0.007	1.30 (1.12-1.51)	< 0.001	0.93 (0.81-1.07)	0.297
<b>Private Health Insurance</b>	1.67 (1.49-1.87)	< 0.001	2.39 (2.13-2.69)	< 0.001	1.56 (1.40-1.75)	< 0.001
<b>Veterans Affairs</b>	1.62 (1.17-2.24)	0.004	2.26 (1.60-3.21)	< 0.001	1.29 (0.91-1.84)	0.158



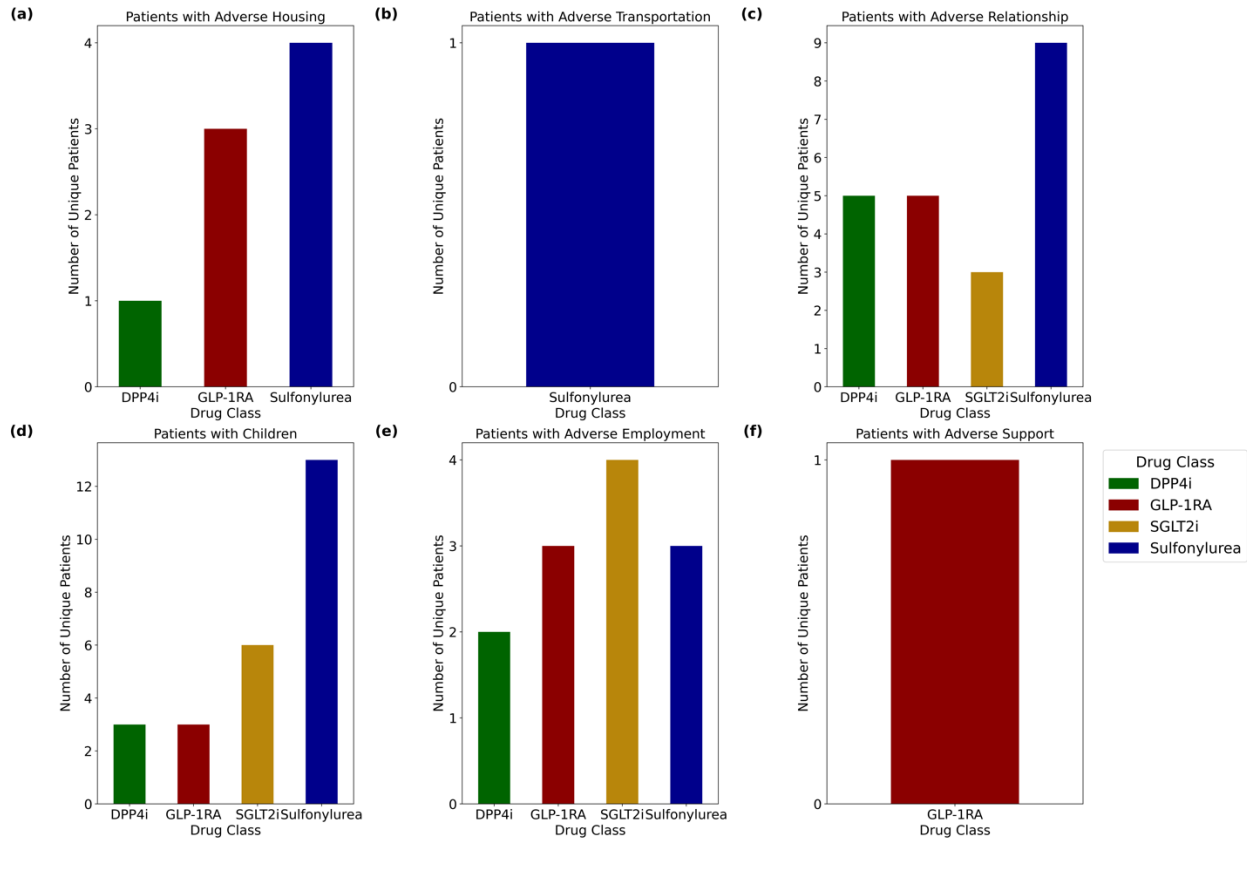
**Figure 3-1: Second-Line Treatment Cohort Design**

Flow chart illustrating cohort selection for patients being started on second-line sulfonylurea (SU) across five academic health systems in University of California Health. A similar cohort selection process was repeated for each of the other 3 categories of second-line treatments.



**Figure 3-2: Second-line treatment options for Type 2 diabetes patients**

Patients randomly selected from our cohort with varying characteristics (age, race, ethnicity, HbA1c, and insurance) are shown here. When each of these patients lives in an economically advantaged neighborhood (ADI 1, on the left) the model determines that they are most likely to receive a GLP-1RA as a second line treatment (percent indicates the model likelihood). When those same patients live in an economically disadvantaged neighborhood (ADI 10, on the right), they are most likely to receive a sulfonylurea.



**Figure 3-3: GPT-4 extraction and identification of Individual-Level SDOH Factors**

GPT-4 was provided with 522 clinical notes with Social History and SDOH sections and was prompted to extract and identify individual-level SDOH factors from patients with a second-line T2D treatment. Identified Individual-Level SDOH Factors for UCSF patients with 3a. Adverse Housing 3b. Transportation, 3c. Relationship, 3d. patients with children, 3e. Employment, and 3f. Support stratified by their Drug Class.

### 3.9 SUPPLEMENTAL MATERIALS

Figure 3-4: Cohort selection of Glucagon-Like Peptide 1 Receptor Agonist

Figure 3-5: Cohort selection of SGLT2-Inhibitors

Figure 3-6: Cohort selection of DPP4-Inhibitors

Table 3-3: Likelihood of Second-Line Treatment Options for Randomly Selected patients

Figure 3-7: UCSF Cohort Selection of Glucagon-Like Peptide 1 Receptor Agonist

Figure 3-8: UCSF Cohort Selection of SGLT2-Inhibitors

Figure 3-9: UCSF Cohort Selection of DPP4-Inhibitors

Figure 3-10: UCSF Cohort Selection of Sulfonylureas

Figure 3-11: GPT-4 Prompt for Synthetic SDOH Dataset

Table 3-4: GPT-4 Accuracy Evaluation for Synthetic SDOH Dataset

Figure 3-12: GPT-4 Prompt for MIMIC-III SDOH Dataset

Table 3-5: GPT-4 Accuracy Evaluation for MIMIC-III Transportation Attributes Identification

Table 3-6: GPT-4 Accuracy Evaluation for MIMIC-III Housing Attributes Identification

Table 3-7: GPT-4 Accuracy Evaluation for MIMIC-III Relationship Attributes Identification

Table 3-8: GPT-4 Accuracy Evaluation for MIMIC-III Parent and Support Attributes Identification

Table 3-9: GPT-4 Accuracy Evaluation for MIMIC-III Employment Attributes Identification

Figure 3-13: GPT-4 Prompt for UCSF Deidentified Clinical Note Dataset

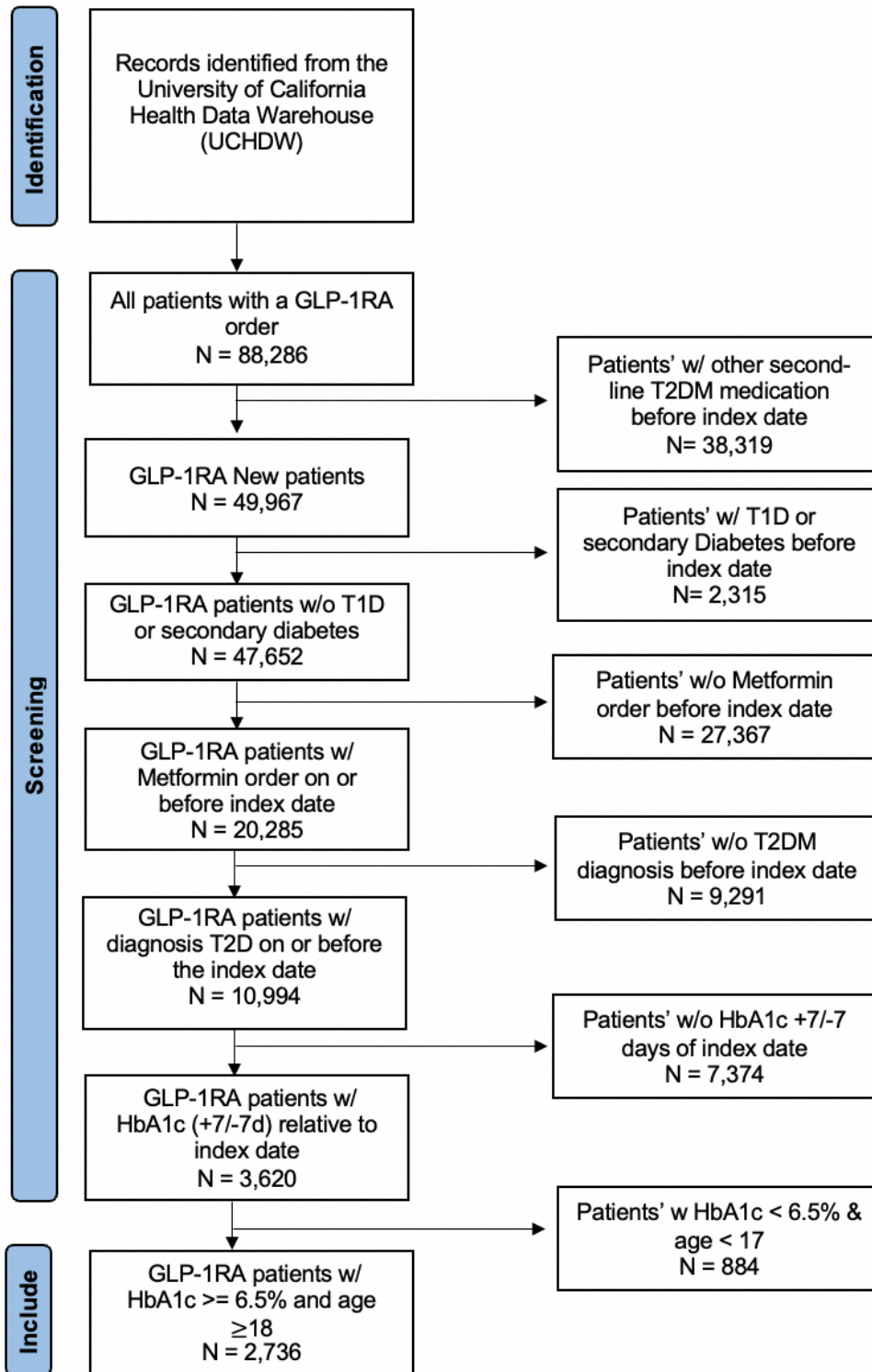
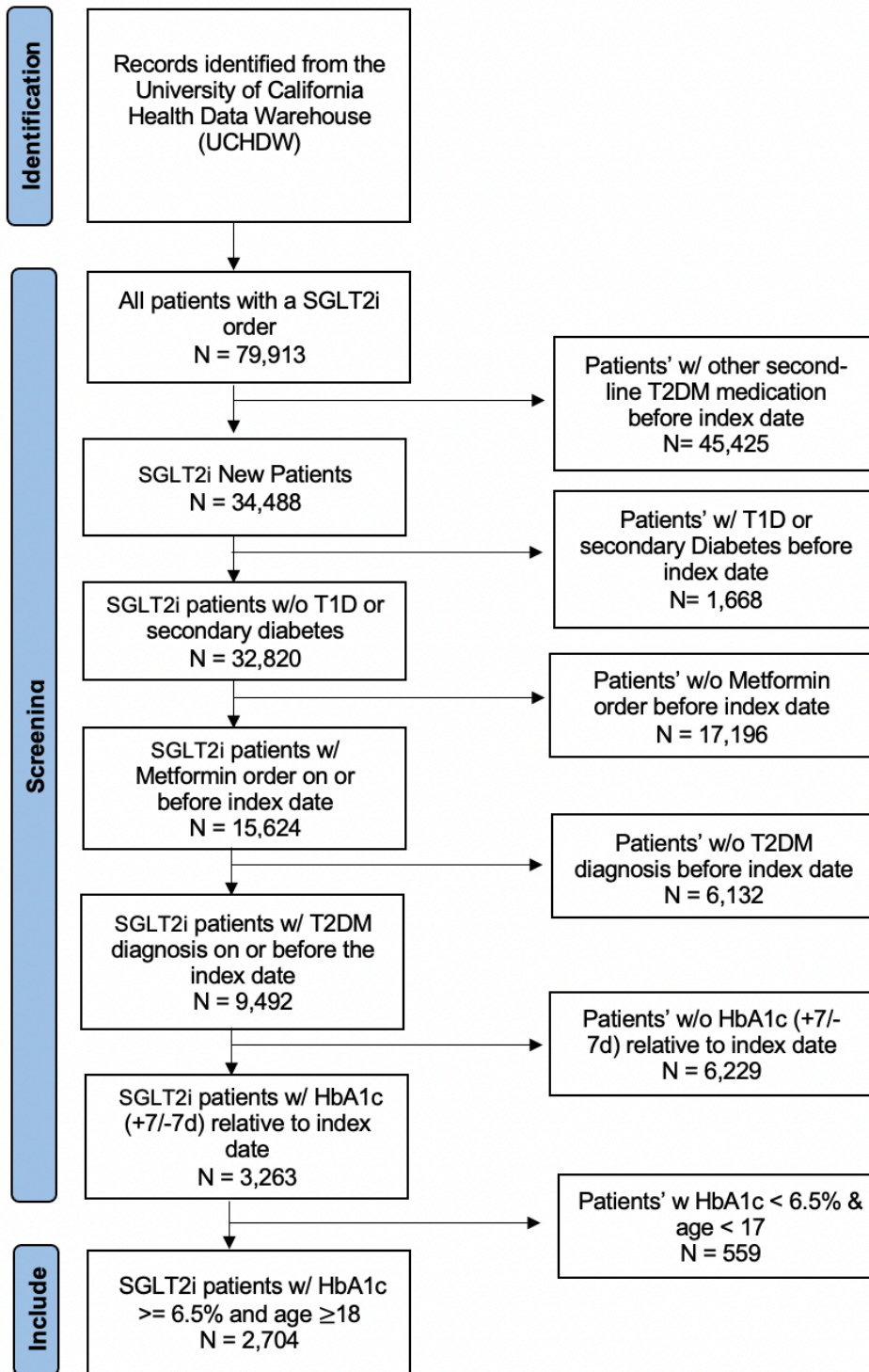


Figure 3-4: Cohort selection Glucagon-Like Peptide 1 Receptor Agonist



**Figure 3-5: Cohort selection of SGLT2-Inhibitors**

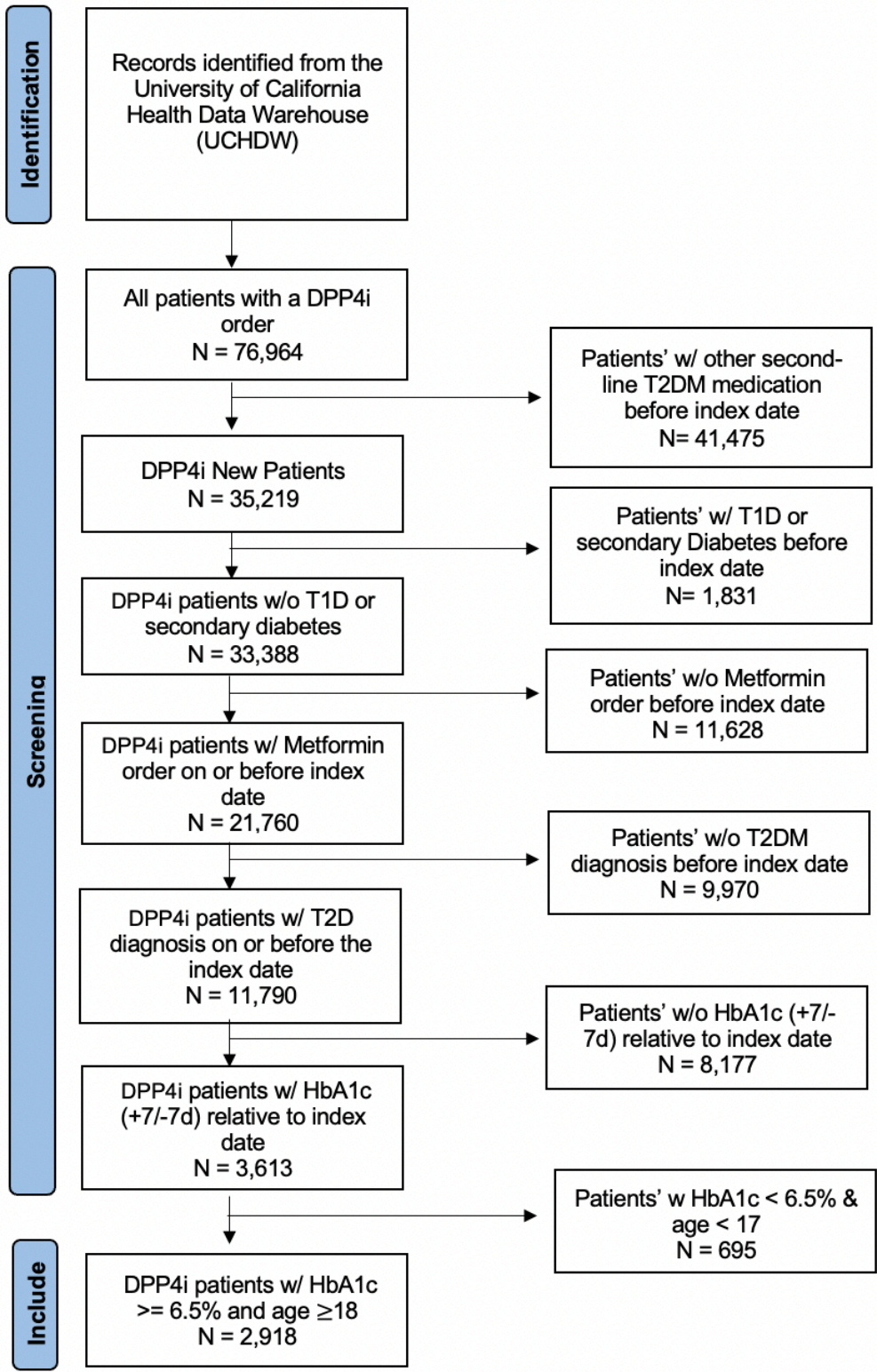


Figure 3-6: Cohort selection of DPP4-Inhibitors

**Table 3-3: Likelihood of Second-Line Treatment Options for Randomly Selected patients**

Patient Description	Likelihood of Second-line Treatment Options			
	Sulfonylurea	DPP4i	GLP-1RA	SGLT2i
48 year old white non hispanic female with HbA1c 7.0% private insurance and lives in ADI 1	0.25	0.17	0.41	0.16
49 year old white non hispanic male with HbA1c 11.9% Medicaid insurance and lives in ADI 10	0.67	0.11	0.10	0.12
44 year old black non hispanic female with HbA1c 6.9% with Medicaid insurance and lives in ADI 1	0.40	0.12	0.31	0.17
42 year old white hispanic female with HbA1c 11.5% with Medicaid insurance and lives in ADI 10	0.69	0.08	0.14	0.08
37 year old asian non hispanic male with HbA1c 11.3% private insurance and lives in ADI 10	0.52	0.14	0.16	0.17
60 year old white non hispanic male with HbA1c 7.85% medicaid insurance and lives in ADI 10	0.55	0.17	0.10	0.18
55 year old white non hispanic male with HbA1c 8.3% medicaid insurance and lives in ADI 10	0.57	0.15	0.11	0.16
60 year old white hispanic female with HbA1c 8.2% medicare insurance and A lives in DI 1	0.46	0.18	0.22	0.15
53 year old hispanic female with HbA1c 9.2% medicaid insurance and lives in ADI 10	0.61	0.14	0.13	0.11
44 year old black female with HbA1c 8.6% medicaid insurance and lives in ADI 10	0.57	0.11	0.21	0.12
51 year old white non hispanic male with HbA1c 10.1% private insurance and lives in ADI 1	0.36	0.17	0.27	0.20
55 year old white non hispanic female with HbA1c 7.9% private insurance and lives in ADI 1	0.28	0.20	0.36	0.17
58 year old black non hispanic female with HbA1c 6.6% private insurance and lives in ADI 1	0.26	0.18	0.37	0.20
56 year old white non hispanic female with HbA1c 8.1% private insurance and lives in ADI 1	0.29	0.20	0.35	0.17
40 year old asian non hispanic male with HbA1c 7.8% medicaid insurance and lives in ADI 1	0.45	0.16	0.14	0.25
46 year old hispanic female with HbA1c 11.8% private insurance and lives in ADI 1	0.32	0.12	0.39	0.17



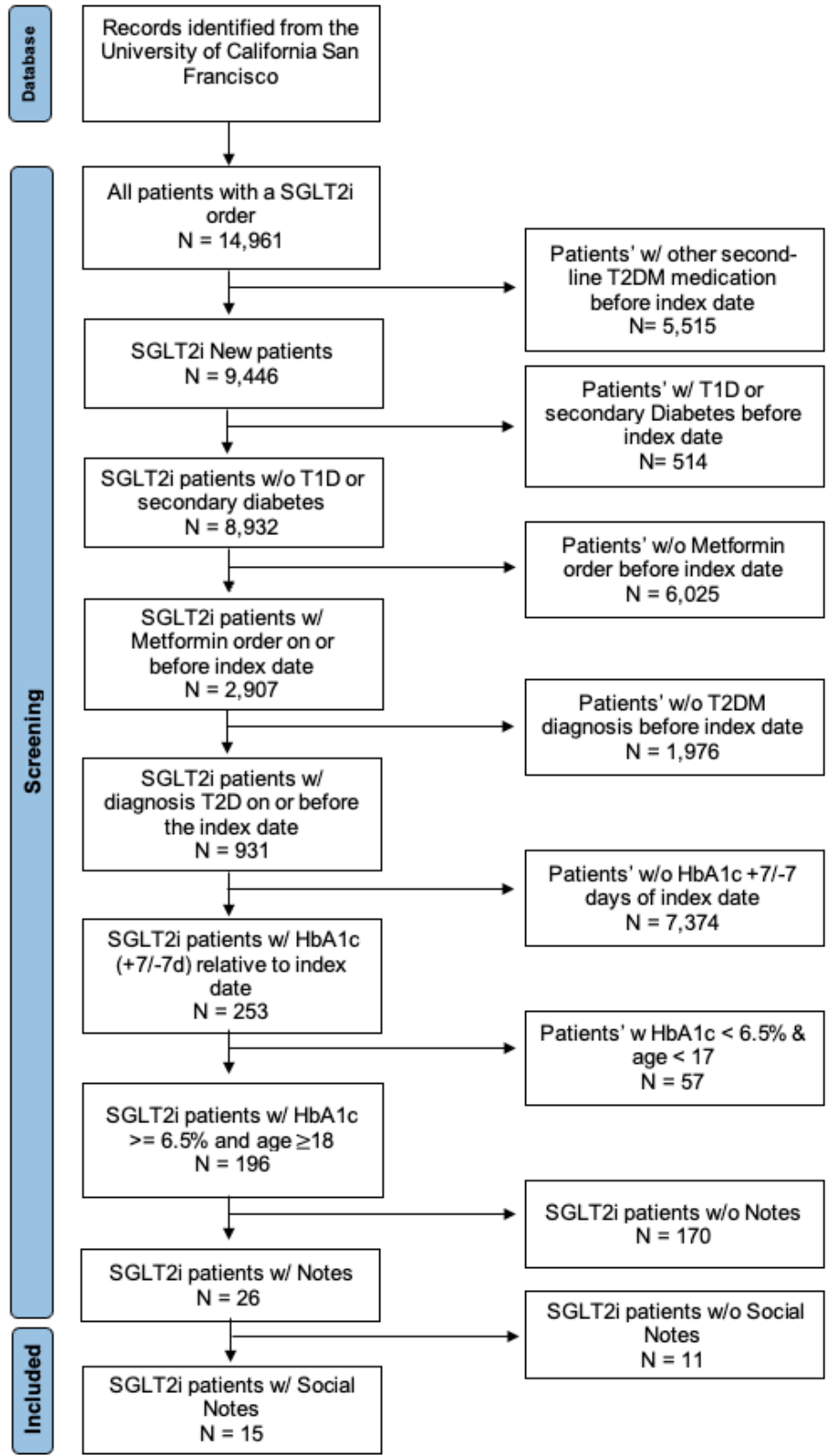


Figure 3-8: UCSF Cohort selection for SGLT2-Inhibitors

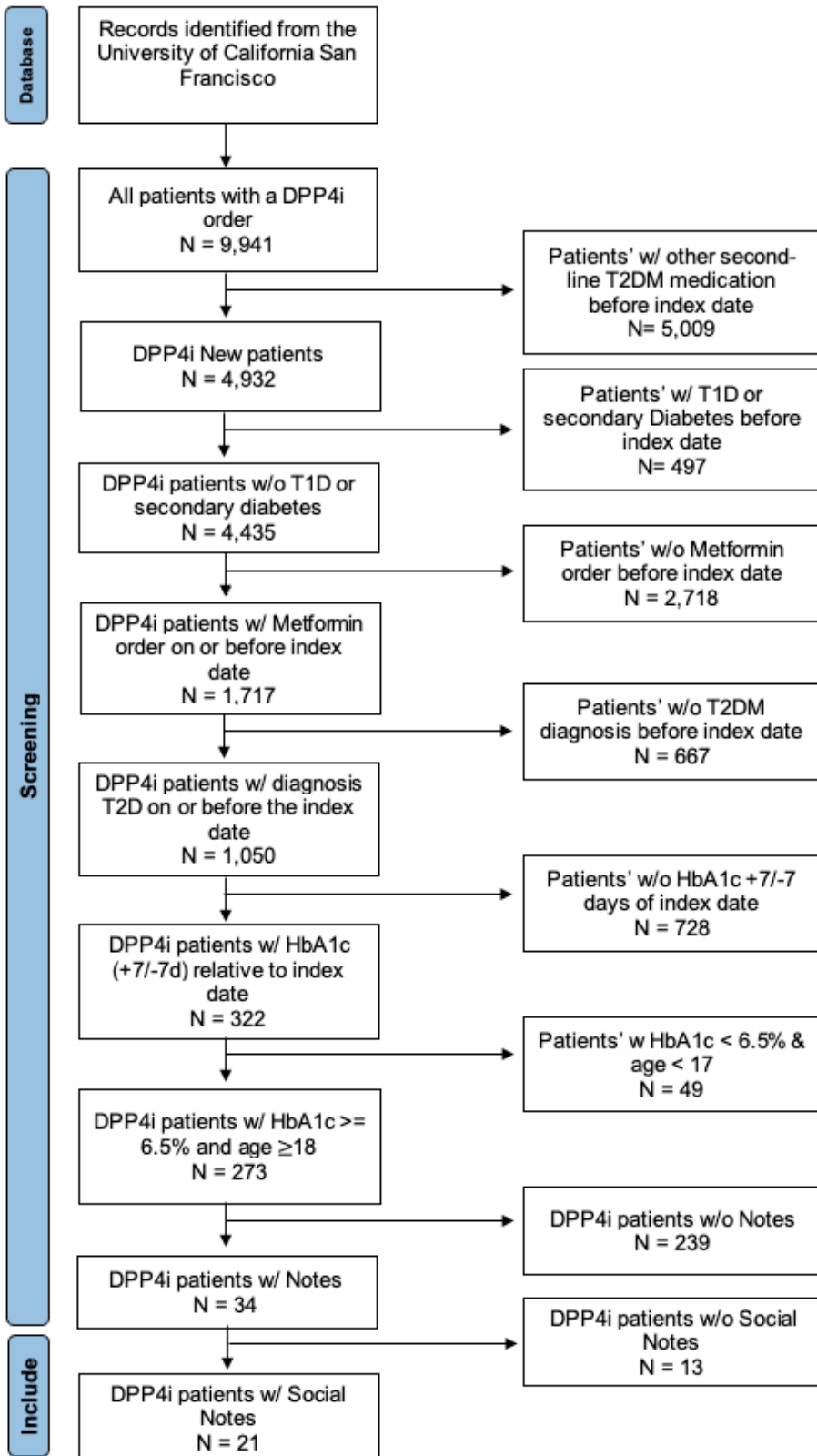


Figure 3-9: UCSF Cohort selection for DPP4-Inhibitors

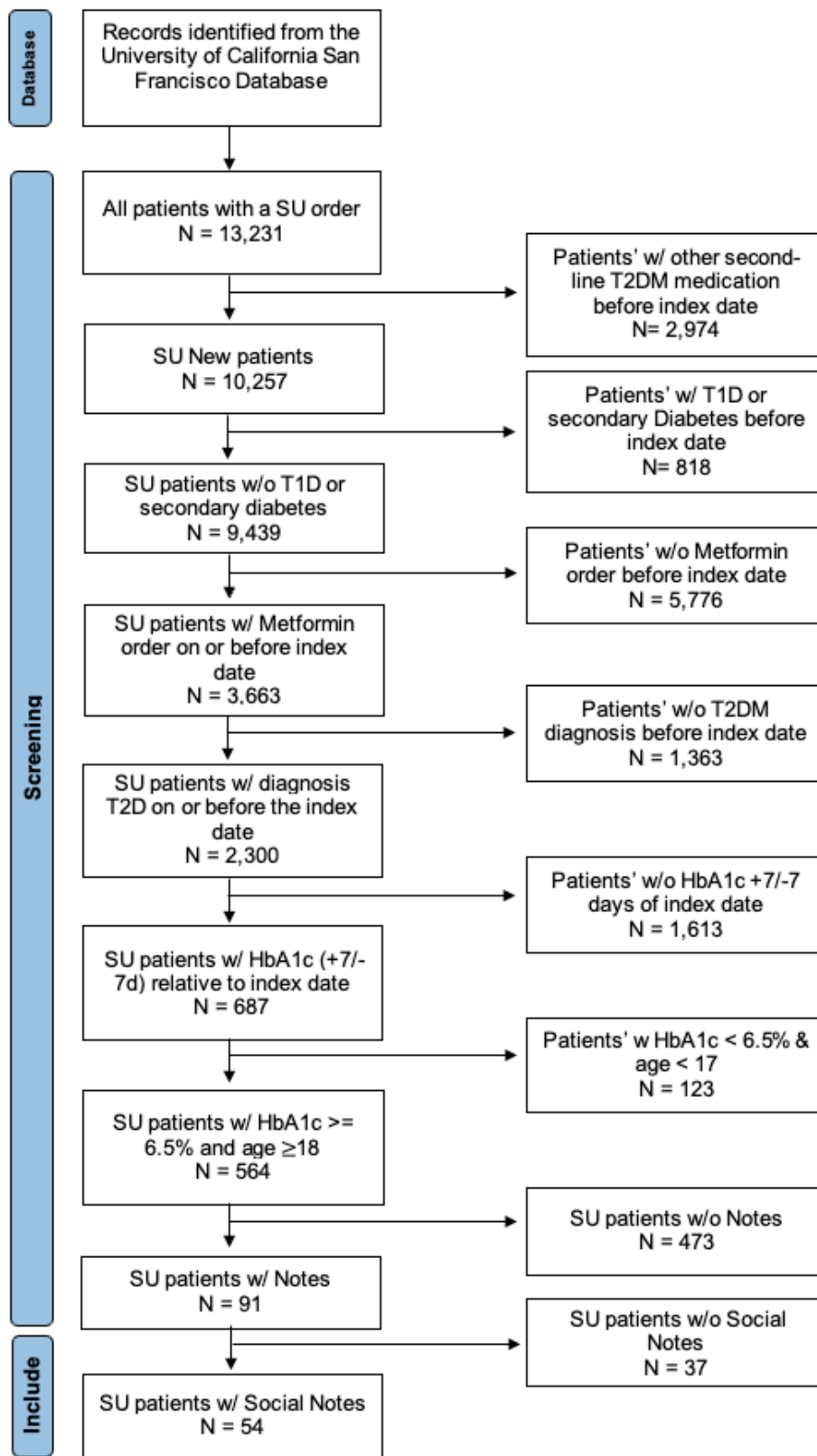


Figure 3-10: UCSF Cohort selection for Sulfonylureas

```

def get_system_prompt(self):
    sysprompt = """
    Pretend you are an endocrinologist reading the given type 2 diabetes report. \
    Extract Social information based on the clinical note categorized by Social Determinants of Health
    Factors
    within the Social History section of the notes. \
    Do not use patient history to answer, only provide the current patient social information as answer. \
    If the note says that patient "histories were reviewed and updated as appropriate" please omit the
    note.\
    Please answer as concisely as possible in the given format. \
    Please give one output for each Unnamed value.\
    """
    return sysprompt

def get_user_prompt(self):
    prompt = """Please identify and label Social Determinants of Health information into Factors. \
    Categorize and label each Social Determinants of Health factor as 'Adverse', 'Non-Adverse', or
    'Unknown'. \
    Adverse means that the patients is experiencing negative effects of social factors.
    Non-Adverse means that the patient does not experience negative effects of social factors.
    Unknown means that the patient does not have social factor information.
    Notes that do not provide social information are irrelevant here. \
    Please do not provide answers with an explanation.\
    Provide the answers as a json in the following format with the following Social Determinants of Health
    factor keys:
    {
    "Transportation": str,
    "Housing": str,
    "Relationship": str,
    "Parent": str,
    "Support":str,
    "Employment":str
    }
    """
    return prompt

```



**Figure 3-11: GPT-4 Prompt for Synthetic SDOH Dataset**

**Table 3-4: GPT-4 Accuracy Evaluation for Synthetic SDOH Dataset**

	Housing	Transportation	Relationship	Parent	Support	Employment
<b>Acc:</b>	1781.0	1763.0	1729.0	1654.0	1779.0	1780.0
<b>Acc: Norm</b>	0.990	0.980	0.961	0.919	0.989	0.989
<b>Micro F1</b>	0.990	0.980	0.961	0.952	0.989	0.989
<b>Micro Precision</b>	0.990	0.980	0.961	0.919	0.989	0.989
<b>Micro Recall</b>	0.990	0.980	0.961	0.919	0.989	0.989

```

def get_system_prompt(self):
    sysprompt = """
    Pretend you are an endocrinologist reading the given type 2 diabetes report. \
    Extract Social information based on the clinical note categorized by Social Determinants of Health Factors \
    within the Social History section of the notes. \
    Do not use patient history to answer, only provide the current patient social information as answer. \
    If the note says that patient "histories were reviewed and updated as appropriate" please omit the note. \
    Please answer as concisely as possible in the given format. \
    Please give one output for each Unnamed value. \
    """
    return sysprompt

def get_user_prompt(self):
    prompt = """Please identify and label Social Determinants of Health information into Factors. \
    Categorize and label each Social Determinants of Health factors as the attributes below'. \
    Notes that do not provide social information are irrelevant here. \
    Notes without social information can be returned with an output of Nan. \
    If the note says that patient "histories were reviewed and updated as appropriate" please omit the note. \
    Please provide outputs in binary of [0,1]. \
    0 means factor not present. \
    1 means factor is present. \
    Please do not provide answers with an explanation. \

    Housing Attributes: \
    Poor: The patient is housed with governmental assistance, or there is evidence that they are struggling to \
    afford their housing due to resource limitations. \
    Undomiciled: There is evidence that the patient does not have stable housing and/or is living in a shelter. \
    Other: The sentence expresses a housing issue that is not necessarily financial/resource-limited in nature. \
    If Housing status is not mentioned, DO NOT categorize it. \

    Transportation Attributes: \
    Distance: The patient has issues getting to and from the treatment site due to the distance from their \
    residence. \
    Resource: The patient has issues (financial or material) to get to and from the treatment site. \
    Other: Sentence represents a transportation adequacy issue that is not described by the other two \
    attributes, including mentions of time. \
    If Transportation status is not mentioned, DO NOT categorize it. \

    Relationship Attributes: \
    Married: Patient is married. \
    Partnered: In a non-marital relationship with at least one other person. \
    Divorced: Patient is divorced or separated. \
    Widowed: Patient is widowed. \
    Single: Patient is not in a relationship and does not meet the above criteria. \
    If Relationship status is not mentioned, DO NOT categorize it. \
    """

```



**Figure 3-12: GPT-4 Prompt for MIMIC-III SDOH Dataset**

```
Parent Attributes: \
This SDOH should be used for descriptions of a patient being a parent to at least one child who is a minor
(under the age of 18 years old). \
If Parent status is not mentioned, DO NOT categorize it. \

Employment Attributes: \
Employed: The patient is currently employed, and there is no evidence that the position is financially
insecure. Financial security does not need to be explicitly stated. \
Under-Employed: The patient is employed in a financially insecure position. \
Unemployed: The patient is not currently employed in a paid position. \
Disability: The patient is currently on disability. \
Retired: The patient is retired and living off of income other than a paid position. \
Student: The patient is a student as their primary occupation. \
If Employment status is not mentioned, DO NOT categorize it. \

Social Support Attributes: \
+ Support: The patient is receiving social support from people they know personally. \
- Support: The patient is not receiving social support from people they know personally. \
Sentences that only mention someone who is in the life of the patient with no other information about
support provided, should NOT be considered for a Social Support tag. \
If Social Support status is not mentioned, DO NOT categorize it. \

Provide the answers as a json in the following format with the following Social Determinants of Health
factor keys:
{
  "TRANSPORTATION_attribute": str,
  "HOUSING_attribute": str,
  "RELATIONSHIP_attribute": str,
  "PARENT": str,
  "SUPPORT_attribute ": str,
  "EMPLOYMENT_attribute": str,
}
"""

return prompt
```



**Figure 3-12: GPT-4 Prompt for MIMIC-III SDOH Dataset (Continued from Previous)**

**Table 3-5: GPT-4 Accuracy Evaluation for MIMIC-III Transportation Attributes Identification**

	<b>TRANSPORTATION: Distance</b>	<b>TRANSPORTATION: Resource</b>	<b>TRANSPORTATION: Other</b>
<b>Acc:</b>	804.0	796.0	800.0
<b>Acc: Norm</b>	1.000	0.990	0.995
<b>Micro F1</b>	1.000	0.995	0.995
<b>Micro Precision</b>	1.000	0.990	0.995
<b>Micro Recall</b>	1.000	0.990	0.995

**Table 3-6: GPT-4 Accuracy Evaluation for MIMIC-III Housing Attributes Identification**

	<b>HOUSING: Poor</b>	<b>HOUSING: Undomiciled</b>	<b>HOUSING: Other</b>
<b>Acc:</b>	804.0	803.0	800.0
<b>Acc: Norm</b>	1.000	0.999	0.995
<b>Micro F1</b>	1.000	0.999	0.998
<b>Micro Precision</b>	1.000	0.999	0.995
<b>Micro Recall</b>	1.000	0.999	0.995

**Table 3-7: GPT-4 Accuracy Evaluation for MIMIC-III Relationship Attributes Identification**

	<b>RELATIONSHIP : Married</b>	<b>RELATIONSHIP : Partnered</b>	<b>RELATIONSHIP : Divorced</b>	<b>RELATIONSHIP : Widowed</b>	<b>RELATIONSHIP : Single</b>
<b>Acc:</b>	771.0	795.0	800.0	804.0	804.0
<b>Acc: Norm</b>	0.959	0.989	0.995	1.000	1.000
<b>Micro F1</b>	0.959	0.989	0.995	1.000	1.000
<b>Micro Precision</b>	0.959	0.989	0.995	1.000	1.000
<b>Micro Recall</b>	0.959	0.989	0.995	1.000	1.000

**Table 3-8: GPT-4 Accuracy Evaluation for MIMIC-III Parent and Support Attributes Identification**

	<b>PARENT</b>	<b>SUPPORT: Minus</b>	<b>SUPPORT: Plus</b>
<b>Acc:</b>	773.0	790.0	674.0
<b>Acc: Norm</b>	0.961	0.983	0.838
<b>Micro F1</b>	0.961	0.983	0.912
<b>Micro Precision</b>	0.961	0.983	0.838
<b>Micro Recall</b>	0.961	0.983	0.838

**Table 3-9: GPT-4 Accuracy Evaluation for MIMIC-III Employment Attributes Identification**

	<b>EMPLOYMENT: Underemployed</b>	<b>EMPLOYMENT: Unemployed</b>	<b>EMPLOYMENT: Disability</b>	<b>EMPLOYMENT: Retired</b>	<b>EMPLOYMENT: Student</b>
<b>Acc:</b>	802.0	803.0	797.0	788.0	794.0
<b>Acc: Norm</b>	0.998	0.999	0.991	0.980	0.988
<b>Micro F1</b>	0.998	0.999	0.991	0.980	0.988
<b>Micro Precision</b>	0.998	0.999	0.991	0.980	0.988
<b>Micro Recall</b>	0.998	0.999	0.991	0.980	0.988

```

def get_system_prompt(self):
    sysprompt = """
    Pretend you are an endocrinologist reading the given type 2 diabetes report. \
    Extract Social information based on the clinical note categorized by Social Determinants of Health Factors \
    within the Social History section of the notes. \
    Do not use patient history to answer, only provide the current patient social information as answer. \
    If the note says that patient "histories were reviewed and updated as appropriate" please omit the note.\
    Please answer as concisely as possible in the given format. \
    Please give one output for each Unnamed value.\
    """
    return sysprompt

def get_user_prompt(self):
    prompt = """Please identify and label Social Determinants of Health information into Factors. \
    Categorize and label each Social Determinants of Health factors as the attributes below'. \
    Notes that do not provide social information are irrelevant here. \
    Notes without social information can be returned with an output of Nan.\
    If the note says that patient "histories were reviewed and updated as appropriate" please omit the note.\
    Please provide outputs in binary of [0,1].\
    0 means factor not present.\
    1 means factor is present.\
    Please do not provide answers with an explanation.\

    Housing Attributes: \
    Poor: The patient is housed with governmental assistance, or there is evidence that they are struggling to \
    afford their housing due to resource limitations. \
    Undomiciled: There is evidence that the patient does not have stable housing and/or is living in a shelter. \
    Other: The sentence expresses a housing issue that is not necessarily financial/resource-limited in nature. \
    If Housing status is not mentioned, DO NOT categorize it. \

    Transportation Attributes: \
    Distance: The patient has issues getting to and from the treatment site due to the distance from their \
    residence. \
    Resource: The patient has issues (financial or material) to get to and from the treatment site. \
    Other: Sentence represents a transportation adequacy issue that is not described by the other two \
    attributes, including mentions of time. \
    If Transportation status is not mentioned, DO NOT categorize it. \

    Relationship Attributes: \
    Married: Patient is married. \
    Partnered: In a non-marital relationship with at least one other person. \
    Divorced: Patient is divorced or separated. \
    Widowed: Patient is widowed. \
    Single: Patient is not in a relationship and does not meet the above criteria. \
    If Relationship status is not mentioned, DO NOT categorize it. \

```



**Figure 3-13: GPT-4 Prompt for UCSF Deidentified Clinical Note Dataset**

```
Parent Attributes: \
This SDOH should be used for descriptions of a patient being a parent to at least one child who is a minor
(under the age of 18 years old). \
If Parent status is not mentioned, DO NOT categorize it. \

Employment Attributes: \
Employed: The patient is currently employed, and there is no evidence that the position is financially
insecure. Financial security does not need to be explicitly stated. \
Under-Employed: The patient is employed in a financially insecure position. \
Unemployed: The patient is not currently employed in a paid position. \
Disability: The patient is currently on disability. \
Retired: The patient is retired and living off of income other than a paid position. \
Student: The patient is a student as their primary occupation. \
If Employment status is not mentioned, DO NOT categorize it. \

Social Support Attributes: \
+ Support: The patient is receiving social support from people they know personally. \
- Support: The patient is not receiving social support from people they know personally. \
Sentences that only mention someone who is in the life of the patient with no other information about
support provided, should NOT be considered for a Social Support tag. \
If Social Support status is not mentioned, DO NOT categorize it. \

Provide the answers as a json in the following format with the following Social Determinants of Health
factor keys:
{
  "TRANSPORTATION_attribute": str,
  "HOUSING_attribute": str,
  "RELATIONSHIP_attribute": str,
  "PARENT": str,
  "SUPPORT_attribute ": str,
  "EMPLOYMENT_attribute": str,
}
"""

return prompt
```



**Figure 3-13: GPT-4 Prompt for UCSF Deidentified Clinical Note Dataset (Continued from Previous)**

## REFERENCES

1. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract.* 2019;157:107843. doi:10.1016/j.diabres.2019.107843
2. Centers for Disease Control and Prevention. National Diabetes Statistics Report website. <https://www.cdc.gov/diabetes/data/statistics-report/index.html>. Accessed November 2023.
3. American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in Diabetes-2020. *Diabetes Care.* 2020;43(Suppl 1):S66-S76. doi:10.2337/dc20-S006
4. Davies MJ, Aroda VR, Collins BS, et al. Management of hyperglycaemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia.* 2022;65(12):1925-1966. doi:10.1007/s00125-022-05787-2
5. International Diabetes Federation. Recommendations for managing type 2 diabetes in primary care. Accessed April 5, 2023. <http://www.idf.org/managing-type2-diabetes>
6. American Diabetes Association Professional Practice Committee. 9. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes-2022. *Diabetes Care.* 2022;45(Suppl 1):S125-S143. doi:10.2337/dc22-S009
7. Hemmingsen B, Lund SS, Gluud C, et al. Intensive glycemic control for patients with type 2 diabetes: systematic review with meta-analysis and trial sequential analysis of randomised clinical trials. *BMJ.* 2011;343:d6898. Published 2011 Nov 24. doi:10.1136/bmj.d6898

8. Fang M, Wang D, Coresh J, Selvin E. Trends in Diabetes Treatment and Control in U.S. Adults, 1999-2018. *N Engl J Med.* 2021;384(23):2219-2228. doi:10.1056/NEJMsa2032271
9. Emanuel EJ, Dellgren JL, McCoy MS, Persad G. Fair Allocation of GLP-1 and Dual GLP-1-GIP Receptor Agonists. *N Engl J Med.* 2024;390(20):1839-1842. doi:10.1056/NEJMp2400978
10. Hasbrouck L. Healthy People 2030: An Improved Framework. *Health Educ Behav.* 2021;48(2):113-114. doi:10.1177/1090198121997812
11. Schmidt-Busby J, Wiles J, Exeter D, Kenealy T. Understanding 'context' in the self-management of type 2 diabetes with comorbidities: A systematic review and realist evaluation. *Diabetes Res Clin Pract.* 2018;142:321-334. doi:10.1016/j.diabres.2018.06.003
12. Ackermann RT, Holmes AM, Saha C. Designing a natural experiment to evaluate a national health care-community partnership to prevent type 2 diabetes. *Prev Chronic Dis.* 2013;10:E12. doi:10.5888/pcd10.120149
13. de Vries McClintock HF, Wiebe DJ, O'Donnell AJ, Morales KH, Small DS, Bogner HR. Neighborhood social environment and patterns of adherence to oral hypoglycemic agents among patients with type 2 diabetes mellitus. *Fam Community Health.* 2015;38(2):169-179. doi:10.1097/FCH.0000000000000069
14. Laraia BA, Karter AJ, Warton EM, Schillinger D, Moffet HH, Adler N. Place matters: neighborhood deprivation and cardiometabolic risk factors in the Diabetes Study of Northern California (DISTANCE). *Soc Sci Med.* 2012;74(7):1082-1090. doi:10.1016/j.socscimed.2011.11.036

15. Shaha KC, Sultana S, Saha SK, Shahidullah SM, Jyoti BK. Patient Characteristics Associated With Medication Adherence to Anti-Diabetic Drugs. *Mymensingh Med J.* 2019;28(2):423-428.
16. McDoom MM, Cooper LA, Hsu YJ, Singh A, Perin J, Thornton RLJ. Neighborhood Environment Characteristics and Control of Hypertension and Diabetes in a Primary Care Patient Sample. *J Gen Intern Med.* 2020;35(4):1189-1198. doi:10.1007/s11606-020-05671-y
17. Myers CA, Slack T, Broyles ST, Heymsfield SB, Church TS, Martin CK. Diabetes prevalence is associated with different community factors in the diabetes belt versus the rest of the United States. *Obesity (Silver Spring).* 2017;25(2):452-459. doi:10.1002/oby.21725
18. Vijayaraghavan M, Jacobs EA, Seligman H, Fernandez A. The association between housing instability, food insecurity, and diabetes self-efficacy in low-income adults. *J Health Care Poor Underserved.* 2011;22(4):1279-1291. doi:10.1353/hpu.2011.0131
19. Truong HP, Luke AA, Hammond G, Wadhwa RK, Reidhead M, Joynt Maddox KE. Utilization of Social Determinants of Health ICD-10 Z-Codes Among Hospitalized Patients in the United States, 2016-2017. *Med Care.* 2020;58(12):1037-1043. doi:10.1097/MLR.0000000000001418
20. Walker RJ, Strom Williams J, Egede LE. Influence of Race, Ethnicity and Social Determinants of Health on Diabetes Outcomes. *Am J Med Sci.* 2016;351(4):366-373. doi:10.1016/j.amjms.2016.01.008

21. Hill-Briggs F, Adler NE, Berkowitz SA, et al. Social Determinants of Health and Diabetes: A Scientific Review. *Diabetes Care*. Published online November 2, 2020. doi:10.2337/dci20-0053
22. Haire-Joshu D, Hill-Briggs F. The Next Generation of Diabetes Translation: A Path to Health Equity. *Annu Rev Public Health*. 2019;40:391-410. doi:10.1146/annurev-publhealth-040218-044158
23. Davidson J, Vashisht R, Butte AJ. From Genes to Geography, from Cells to Community, from Biomolecules to Behaviors: The Importance of Social Determinants of Health. *Biomolecules*. 2022;12(10):1449. Published 2022 Oct 9. doi:10.3390/biom12101449
24. Vashisht R, Patel A, Dahm L, et al. Second-Line Pharmaceutical Treatments for Patients with Type 2 Diabetes [published correction appears in *JAMA Netw Open*. 2023 Nov 1;6(11):e2344113]. *JAMA Netw Open*. 2023;6(10):e2336613. Published 2023 Oct 2. doi:10.1001/jamanetworkopen.2023.36613
25. Hripesak G, Ryan PB, Duke JD, et al. Characterizing treatment pathways at scale using the OHDSI network. *Proc Natl Acad Sci U S A*. 2016;113(27):7329-7336. doi:10.1073/pnas.1510502113
26. Peterson TA, Fontil V, Koliwad SK, Patel A, Butte AJ. Quantifying Variation in Treatment Utilization for Type 2 Diabetes Across Five Major University of California Health Systems. *Diabetes Care*. 2021;44(4):908-914. doi:10.2337/dc20-0344
27. Singh GK. Area deprivation and widening inequalities in US mortality, 1969-1998. *Am J Public Health*. 2003;93(7):1137-1143. doi:10.2105/ajph.93.7.1137

28. Nelson CA, Butte AJ, Baranzini SE. Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. *Nat Commun.* 2019;10(1):3045. Published 2019 Jul 10. doi:10.1038/s41467-019-11069-0
29. Biedermann, P., Ong, R., Davydov, A. et al. Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases. *BMC Med Res Methodol* 21, 238 (2021). <https://doi.org/10.1186/s12874-021-01434-3>
30. Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007;4(10):e297. doi:10.1371/journal.pmed.0040297
31. Fairfield KM, Black AW, Ziller EC, et al. Area Deprivation Index and Rurality in Relation to Lung Cancer Prevalence and Mortality in a Rural State. *JNCI Cancer Spectr.* 2020;4(4):pkaa011. Published 2020 Mar 7. doi:10.1093/jncics/pkaa011
32. Ghirimoldi FM, Schmidt S, Simon RC, et al. Association of Socioeconomic Area Deprivation Index with Hospital Readmissions After Colon and Rectal Surgery. *J Gastrointest Surg.* 2021;25(3):795-808. doi:10.1007/s11605-020-04754-9
33. Vince RA Jr, Jiang R, Bank M, et al. Evaluation of Social Determinants of Health and Prostate Cancer Outcomes Among Black and White Patients: A Systematic Review and Meta-analysis [published correction appears in *JAMA Netw Open.* 2023 Feb 1;6(2):e231256]. *JAMA Netw Open.* 2023;6(1):e2250416. Published 2023 Jan 3. doi:10.1001/jamanetworkopen.2022.50416
34. Hu J, Kind AJH, Nerenz D. Area Deprivation Index Predicts Readmission Risk at an Urban Teaching Hospital. *Am J Med Qual.* 2018;33(5):493-501. doi:10.1177/1062860617753063

35. Luningham JM, Seth G, Saini G, et al. Association of Race and Area Deprivation With Breast Cancer Survival Among Black and White Women in the State of Georgia. *JAMA Netw Open*. 2022;5(10):e2238183. Published 2022 Oct 3. doi:10.1001/jamanetworkopen.2022.38183
36. Knighton AJ, Savitz L, Belnap T, Stephenson B, VanDerslice J. Introduction of an Area Deprivation Index Measuring Patient Socioeconomic Status in an Integrated Health System: Implications for Population Health. *EGEMS (Wash DC)*. 2016;4(3):1238. Published 2016 Aug 11. doi:10.13063/2327-9214.1238
37. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible – The Neighborhood Atlas. *N Engl J Med*. 2018;378(26):2456-2458. doi:10.1056/NEJMp1802313
38. Agresti, A. (2002) *Categorical Data Analysis, Second Edition*. Hoboken, New Jersey: John Wiley & Sons, Inc.
39. Harrell, F. E. (2001) *Regression Modeling Strategies*. New York: Springer-Verlag.
40. Kwak C, Clayton-Matthews A. Multinomial logistic regression. *Nurs Res*. 2002;51(6):404-410. doi:10.1097/00006199-200211000-00009
41. Liang J, Bi G, Zhan C. Multinomial and ordinal Logistic regression analyses with multi-categorical variables using R. *Ann Transl Med*. 2020;8(16):982. doi:10.21037/atm-2020-57
42. Guevara, M., Chen, S., Thomas, S., Chaunzwa, T. L., Franco, I., Kann, B. H., Moningi, S., Qian, J. M., Goldstein, M., Harper, S., Aerts, H. J. W. L., Catalano, P. J., Savova, G. K., Mak, R. H., & Bitterman, D. S. (2024). Large language models to identify social

determinants of health in electronic health records. NPJ digital medicine, 7(1), 6.  
<https://doi.org/10.1038/s41746-023-00970-0>

43. DHHS. (2024). National Diabetes Statistics Report, 2024. National Diabetes Statistics Report, 2

# Chapter 4 Investigating CAR-T Treatment Access for Multiple Myeloma Patients Using Real-World Evidence

## 4.1 ABSTRACT

**Importance:** Multiple myeloma (MM) is the second most common hematologic malignancy in the U.S., with Black patients being diagnosed at twice the rate of White patients. Treatment options are limited and ineffective, but CAR-T therapies show promise. However, their limited availability results in disparities in access.

**Objective:** This study aimed to explore disparities in Multiple Myeloma disease risk and CAR-T therapy access.

**Design:** Our study included a cohort of patients extracted from the University of California Health Data Warehouse (UCHDW) between January 2021 and November 2024.

**Setting:** Data for this study was drawn from 3 academic health centers within the UCHDW.

**Participants:** Our study included a population-based cohort of 12,360 patients diagnosed with Multiple Myeloma who received more than one cancer therapy. Data was analyzed from March 2024 to December 2024.

**Main Outcomes and Measures:** Regression models were used to compute odds ratio (OR) and 95% confidence intervals (CI) associating disease severity, UC-Location, and baseline demographics with CAR-T therapy access. The GPT-4 inference model was prompted with a zero-shot learning

approach to analyze UCSF clinical notes with the following objectives: (1) Was CAR-T discussed? [yes/no], (2) Is the patient eligible for CAR-T? [yes/no/unclear], and (3) Provide the rationale for the eligibility determination.

Results: Our study included 12,360 patients (mean age 68.5 years, SD 12.8 years) treated for multiple myeloma across the University of California Health System, 320 of which received CAR-T (**Table-1**). Overall, 51.6% of MM patients identified as Male, and 48.4% as Female. 61.2% of patients identified as White, with 57.5% Male and 60.9% as White among CAR-T patients. Disease Severity was measured by the International Staging System (ISS) and was distributed by ISS Stage: I (65.3%), II (24.4%), III (2.8%), and None (7.5%). We also used the number of CRAB features to assess disease severity, and the Mean (SD) was 1.86 (0.98), with the most common feature being Anemia. Patients treated at UC-1 (49.3%), and UC-2 (50.0%) were primarily diagnosed with Stage II, while patients at UC-3 (55.5%) were primarily diagnosed with Stage I. Our model indicated that Black patients (OR= 0.33, [95% CI, 0.17-0.62) were less likely to receive CAR-T therapy when compared to White patients. Patients treated at UC-3 with predominantly Black and Stage I diagnosis patients (OR = 0.42, [95% CI, 0.30-0.59]) were less likely to receive CAR-T therapy when compared to UC-1 with predominantly Stage II patients. We identified CAR-T eligibility for 270 UCSF patients and found patients who identified as other Pacific Islander race had the highest rate of eligibility without discussions at 50%, followed by Black or African American (4.2%), Asian (3.2%), and White (0.6%).

Conclusion and Relevance: This study emphasizes the influence of race and UC-Location on disparities in CAR-T therapy access.

## 4.2 INTRODUCTION

Multiple myeloma (MM) is a hematologic malignancy characterized by the clonal proliferation of malignant plasma cells, which can result in end-organ damage, morbidity, and mortality (with a five-year overall survival of 58%)<sup>1</sup>. Black patients are diagnosed at twice the rate and, on average, 5-10 years earlier than White patients<sup>2</sup> and are more likely to present with anemia, hypercalcemia, and kidney dysfunction<sup>3,4</sup>. They are underrepresented in clinical trials investigating novel therapies<sup>5,6</sup>; despite recent real-world studies showing they achieve better survival rates than White patients when given equal access to care<sup>6-9</sup>. Hispanic or Latino patients with myeloma similarly have a higher incidence and younger age of diagnosis. These patients are known to have longer time intervals from diagnosis to therapy initiation and lower utilization of essential therapies such as autologous stem cell transplant (ASCT)<sup>10</sup>.

Standard of care front-line therapy consists of triplet or quadruplet induction therapy, followed by ASCT consolidation and maintenance in eligible patients; transplant-ineligible patients receive attenuated ongoing systemic therapy.<sup>11-16</sup> Despite advances, multiple myeloma invariably relapses. A recent breakthrough in the treatment of relapsed/refractory multiple myeloma (RRMM) is the use of chimeric antigen receptor T-cell (CAR-T) therapy, which are engineered T-cells designed to target B-cell maturation antigen (BCMA) on plasma cells<sup>17,18</sup>. Two FDA-approved BCMA CAR-T therapies in MM, idecabtagene vicleucel (idecel) and ciltacabtagene autoleucel (ciltacel), are now approved for both early and late relapse. These therapies have demonstrated unprecedented outcomes, including a 98% overall response rate (ORR) and nearly 36-month progression-free survival (PFS) with ciltacel<sup>19</sup>, and offer the appeal of a treatment-free interval (TFI), but they require access to a limited number of academic centers with expertise in

CAR-T management and related toxicities. Patients lacking access to these centers, the ability to relocate, or adequate social support for a caregiver may face barriers to receiving CAR-T therapy. While these barriers may disproportionately affect Black, and Hispanic or Latino patients, the extent to which they influence CAR-T therapy access remains poorly understood.

Therefore, we hypothesize that a patient's location and race are strongly associated with CAR-T therapy access for MM patients, considering both disease risk and baseline demographics. To test this hypothesis, our study systematically analyzed the association between access to CAR-T and disease risk, Social Determinants of Health (SDOH), and Race for MM patients. To do so, we leveraged a clinical data warehouse containing over 9 million individuals across the three main academic medical centers that offer CAR-T administration for cancer patients within the University of California (UC) Health system: UCLA, UCSD, and UCSF.

#### 4.3 METHODS

##### *Data Supporting Study: University of California Health Data Warehouse (UCHDW)*

Data for this study was drawn from the University of California Health System, which includes 20 health professional schools (6 medical schools), 5 academic health centers (UC San Francisco, UC Los Angeles, UC Davis, UC Irvine, and UC San Diego), and 12 hospitals. It has built a secure central data warehouse (UCHDW) for operational improvement, promotion of quality patient care, enabling the next generation of clinical research<sup>20</sup>. The repository currently holds data securely on over 9 million patients seen since 2012. EHR data is extracted from each site and transformed into vendor neutral Observational Medical Outcomes Partnership (OMOP) common data model<sup>21</sup>. De-identification of the data has already been completed to enable clinical research projects, under

guidance from UC campus institutional review boards, privacy, and compliance. Research use of UCHDW is considered non-human subject's research.

### *Study Population*

UCHDW data was extracted from patients diagnosed with MM ICD-10 code (C90.0) and had documentation for multiple cancer therapies and procedures from EHRs spanning from the year 2012 up to August 2024 (**Supplemental Materials**). To select a cohort of patients with potential CAR-T therapy, we used the following inclusion/exclusion criteria to create a final study cohort of 12,360 patients. If patients diagnosed with MM did not have primary insurance information, they were excluded from our study to include only patients with insurance. We also excluded patients who didn't have and those who didn't have an Area Deprivation Index (ADI) value. We further excluded patients without a visit on or after January 1, 2021. Because eligibility for non-investigational CAR-T currently requires at least one previous line of therapy, patients who did not have more than one cancer treatment were excluded from our study. We considered all therapies patients received after MM diagnosis, with at least one year of treatment. Patients who were treated at a UC-facility after January 2021 that offered CAR-T therapies were included in our study. Finally, all patients under 18 years of age at the time of diagnosis were excluded from the cohort. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines to ensure the quality of our observational study<sup>22</sup>.

### *Area Deprivation Index (ADI)*

To estimate patient-level SDOH, we employed ADI as a proxy for patients socioeconomic status<sup>23</sup>. ADI does not account for specific individual or family characteristics, but it is a general estimate

of SDOH based on home location census tract characteristics by integrating data from 17 different variables sourced from the U.S. Census, with factors related to poverty, housing, employment, and education. ADI as a metric has been used in previous observational studies<sup>24-28</sup> to understand its SDOH role on health outcomes in other disease states, distinguishing the most deprived areas with higher percentile scores from the least deprived areas with lower percentile scores<sup>23,29-31</sup>.

### *Study Variables*

Patients were classified as receiving CAR-T or not through data collected by UC Health as above. Additional variables included ADI, Age (in years), patient-reported Gender (Male, Female), presence of primary insurance coverage, patient-reported Race, Ethnicity, UC-Location, and International Staging System (ISS) Disease Stage. Number of (CRAB) features were identified for patients with C: Hypercalcemia, R: Renal failure, A: Anemia, or B: Bone disease, or bone pain. To identify cancer procedures and therapies for patients, we categorized all treatments (surgery, chemotherapy, radiation, ASCT, monoclonal antibodies, corticosteroids, alkylators, bispecific antibodies, CAR-T, immunomodulatory agents, nuclear export inhibitors, proteasome inhibitors to ensure that we correctly accounted for cancer therapeutic classes before CAR-T treatments. The International Staging System (ISS) for MM disease severity was calculated for patients by measuring the patient's serum albumin, and serum  $\beta$ 2 microglobulin (S $\beta$ 2M) levels<sup>32</sup>. Patients were categorized as ISS Stage I, II, III, or Unknown. Age (in years) was calculated relative to the index date from the patients' date of birth. Patients with primary insurance healthcare coverage were utilized as Medicare, Medicaid, Veterans Affairs, or private insurance types. Patient reported race was captured as White, Asian, Black or African American, Native Hawaiian or Pacific Islander, American Indian or Alaska Native, Multi Race, Other Race, or Unknown. Patients reported

ethnicity was utilized as Hispanic or Latino, Not Hispanic -or Latino or Unknown. UC-Location was identified as patients treated at UC-1, UC-2, UC-3.

### *Statistical Analysis*

We conducted our data analysis from March 2024 to December 2024. Descriptive statistics were calculated first for each covariate and secondly, for the overall patient cohort, by using Analysis of Variance (ANOVA), and chi-squared tests<sup>33,34</sup>. We employed a Generalized Linear Model (GLM) to estimate the association of primary insurance coverage, race, and ethnicity and our covariates on our key outcome: CAR-T therapy access<sup>33,34</sup>. We adjusted for age (in years), gender (male, female), ADI, Primary Insurance coverage, UC-Location, number of CRAB features, ISS Disease Stage, race, and ethnicity. Association and its strength were reported as odds ratio (OR) with 95% confidence Intervals (CI) including p-values indicating the statistical significance of how likely a patient was to receive CAR-T therapy. An OR was considered significant if its 95% CI did not span 1, and p-value < 0.05. An OR = 1 indicates no difference between receiving CAR-T or not, OR > 1 indicates increased likelihood of a patient receiving CAR-T, and OR < 1 indicates decreased likelihood of a patient receiving CAR-T. All the calculations were performed using R statistical software version 3.6.3 (R Project for Statistical Computing).

### *CAR-T Eligibility Identification through Large Language Models*

To identify CAR-T eligibility for patients from the UCSF de-identified clinical database, we applied the inclusion and exclusion criteria outlined in the study population section of the methods. We excluded patients who did not have clinical notes associated with their diagnoses. Additionally, patients whose clinical notes had an index date on or before January 1, 2021, were excluded. We

also excluded patients with clinical notes that were not within 365 days of the index date of their last cancer therapy or procedure, resulting in a final cohort of 270 patients with 417 clinical notes **(Supplemental Materials)**. To identify CAR-T eligibility, we utilized the GPT-4 inference model with a zero-shot learning approach. GPT-4 was prompted to analyze the “Assessment & Plan” section of the clinical notes written by physicians. The model was tasked with four objectives: (1) Was CAR-T discussed? [yes/no], (2) Is the patient eligible for CAR-T? [yes/no/unclear], (3) Provide the rationale for the eligibility determination, and (4) Classify the eligibility rationale for CAR-T discussions into predefined groups: ("Ineligible due to disease criteria", "Ineligible due to frailty or comorbidities", "Ineligible due to social determinants of health", "Ineligible due to not enough prior lines of therapy", "Eligible now or potentially eligible in future", "Not enough information to determine eligibility", and "Other"). The model’s accuracy was tested and validated on a subsample of clinical notes by comparing its outputs against the ground truth to assess its performance in identifying CAR-T eligibility. GPT-4 model outputs were analyzed and checked for accuracy by clinicians. The identified CAR-T eligibility outputs were then visualized and presented in a tabular format.

#### 4.4 RESULTS

Our study included 12,360 patients (mean age 68.5 years, SD 12.8 years) treated for multiple myeloma across the University of California, 320 of which received CAR-T **(Table-1)**. Of these patients, 51.6% of MM patients identified as Male, and 48.4% as Female. 61.2% of patients identified as White, while 7.8% identified as Black or African American and 14.9% as Hispanic or Latino. Among patients receiving CAR-T, 57.5% were Male and 60.9% identified as White. Patients treated at each UC-Location were primarily White, but UC-3 had a higher distribution of

patients who identified as Black, and Unknown, while UC-1 had a higher distribution of patients who identified as Unknown Race, and Hispanic or Latino.

Primary insurance coverage was distributed as Medicaid (13.5%), Medicare (39.5%), Private (44.9%), and Veterans Affairs Insurance (2.1%). Disease Severity was measured by the International Staging System (ISS), with patients distributed by ISS Stage: I (65.3%), II (24.4%), III (2.8%), and None (7.5%). Patients treated at each UC-Location were primarily diagnosed with Stage I but patients at UC-1 (21.4%), and UC-2 (39.2%) had a higher distribution of patients diagnosed with Stage II. We used the number of CRAB features per patient to assess disease severity, the Mean (SD) was 1.86 (0.98), with the most common feature being Anemia **(Supplemental Materials)**.

Using a univariate analysis, we found race to be a significant determinant of CAR-T treatment access (Methods,  $p < 0.001$ ,  $\chi^2$  test) **(Table-1)**. ISS Stage was significantly different across Race and was a significant determinant of CAR-T treatment access (Methods,  $p < 0.001$ ,  $\chi^2$  test). We found patient's Socioeconomic status and location to be a significant determinant of CAR-T access (Methods,  $p = 0.007$ , ANOVA). Additionally, we modeled the relationship between UC-Location, Race, and CAR-T access and found patients' UC-Location to be a strong determinant (Methods,  $p < 0.001$ ,  $\chi^2$  test). Furthermore, we ran a regression analysis modeling the relationship between receiving CAR-T treatment that adjusted for ISS Stage, UC-Location, race, ethnicity, number of CRAB features, ADI, and primary insurance coverage. The results were highly significant in showing that these variables played a role in CAR-T treatment access for patients **(Table-2)**.

Our regression model showed that as patients aged, they were less likely (OR = 0.97, [95% CI, 0.96-0.98]) to receive CART treatments when compared to younger patients. Patients who

identified as Hispanic or Latino (OR = 1.26, [95% CI, 0.87-1.81]) were more likely to receive CAR-T treatment than patients who identified as Not Hispanic or Latino. Most strikingly, patients who identified as Black or African American (OR= 0.33, [95% CI, 0.17-0.62) were highly unlikely to receive CAR-T during MM treatments when compared to White patients. Patients with Private primary insurance coverage were more likely to receive CAR-T over patients with Medicaid (OR = 0.71, [95% CI, 0.50-1.01]), Medicare (OR = 0.96, [95% CI, 0.73-1.25]), and Veteran Affairs (OR = 0.61, [95% CI, 0.25-1.53]) insurance.

Results of our model indicated that disease severity was a strong predictor of CAR-T treatment access for patients. Patients who presented with more CRAB features (OR = 1.43, [95% CI, 1.27-1.62]) were more likely to receive a CAR-T treatment. Patients who were diagnosed with ISS Stage II (OR = 1.15, [95% CI, 0.89-1.48]) were more likely to receive a CAR-T treatment than patients with ISS Stage I MM (**Figure-1**).

Our results highlighted that UC-Location was a significant determinant of CAR-T treatment administration. Most CAR-T administration for MM patients was done at UC-2 (43.1%). Our model showed that patients treated at UC-3 (OR = 0.42, [95% CI, 0.30-0.59]) were less likely to receive CAR-T therapy when compared to UC-1. We further investigated individual relationships between UC-Location, ISS Stage, and race, by performing additional regression analysis to understand these differences and found significance for race at UC-Locations ( $p < 0.001$ ), ISS Stage at UC-locations ( $p < 0.001$ ), and number of CRAB features at UC-Locations ( $p < 0.001$ ). Interestingly, our results show that CAR-T treatment access is disparate among different population groups based on Race, disease severity, and UC-Location (**Figure-2a-2c**). The distribution of patients within each UC-Location further explained why Black or African American

patients, and patients with ISS Stage I were less likely to receive CAR-T (**Supplemental Materials**). The distribution showed that locations with fewer CAR-T treatments had the highest amount of Black and Hispanic or Latino patients and patients with ISS stage I. This suggests that disparities in access to CAR-T may be driven by access to locations more likely to offer therapy, rather than individual clinical decision-making.

We further investigated CAR-T therapy access by utilizing a Large Language Model (GPT-4o) deployed within a HIPAA-compliant environment to identify CAR-T eligibility for 270 UCSF patients whose clinical notes contained Assessment & Plan sections. Among these patients, we identified individuals who were considered eligible for CAR-T therapy based on eligibility guidelines but never had a documented discussion about it. We identified CAR-T eligibility for 270 UCSF patients and found patients who identified as other Pacific Islander race had the highest rate of eligibility without discussions at 50%, followed by Black or African American (4.2%), Asian (3.2%), and White (0.6%). These patient populations, despite being eligible for CAR-T therapy, are missing out on potentially life-saving treatments. This finding highlights a critical gap where eligible patients are not being offered or engaged in discussions about CAR-T therapy (**Figure-2**). For patients who had CAR-T discussions with their provider, we further investigated the topic for patients' eligibility decisions. Amongst patients with CAR-T discussions we found the LLM classified 33.3% of them to be Eligible now or potentially eligible in the future and 23.3% who were found to be ineligible (**Figure-3**).

#### 4.5 DISCUSSION

In this retrospective cohort study, we evaluated disease severity, UC-Location, primary insurance coverage, and ADI as independent determinants of CAR-T therapy access for patients diagnosed

with MM within the UC-Health system. Notably, we found that therapy access differed by UC-Locations that offered CAR-T administration. Patients at UC-1, and UC-2 which had a higher percentage of White, and Not Hispanic or Latino patients had higher rates of CAR-T therapy administration than patients treated at UC-3. UC-3 has been known to treat patients for both primary and specialty care. Whereas UC-1 and UC-2 are known for specialty care which may be attributed to higher numbers of CAR-T administration. Differences in access by treatment location could also be attributed to the patient's disease severity.

Our model also revealed disparities in CAR-T access based on patients' race, ethnicity, and age. First, we found that as patients aged, they were less likely to receive CAR-T. This is aligned with clinical practice, as older patients with RRMM may be preferentially administered bispecific antibodies rather than CAR-T given concerns about risks of high-grade cytokine release syndrome (CRS), immune-effector cell-associated neurotoxicity (ICANS), and delayed neurotoxicity such as Parkinsonism<sup>35,36</sup>. Additionally, we found that patients who identified as Black or African American were unlikely to receive CAR-T therapy compared to White patients. This is significant in showing disparities in CAR-T access, given that Black patients are two times more likely to be diagnosed with MM and have worse survival rates than White patients<sup>14</sup>. We suspect that this may be due to the lack of diversity of patients at UC-1, and UC-2 who have the highest rates of CAR-T administration. This discrepancy may also be related to lower rates of enrollment in CAR-T clinical trials.

Beyond UC-Location and race, we discovered that disease severity played a major role in CAR-T therapy access for patients. Patients who were diagnosed with Stage I, or Stage II diagnosis, were more likely to receive CAR-T therapy than patients with Stage III or ISS Stage None diagnosis.

We also found that patients treated at UC-1, and UC-2 had more patients diagnosed with Stage I or Stage II in our dataset. This finding contributes to why UC-1, and UC-2 had more patients with CAR-T therapies than patients at UC-3.

With our understanding of UC-3, primarily treating minorities and patients with lower socioeconomic status, we found it interesting that ADI was not a significant predictor in determining CAR-T access. Considering that socioeconomic status is typically associated with clinical trial access and access to newer treatments in clinical disease, we expected ADI to be significant<sup>38,39</sup>. However, we found it refreshing that Hispanic or Latino patients were more likely to receive CAR-T therapies than White patients. This is a step in the right direction for therapy access when current knowledge shows that Hispanic or Latino patients are less likely to be enrolled in clinical trials and have a higher proportion of patients who live in zip codes with low socioeconomic status and education levels<sup>40</sup>.

Our study revealed that some patients considered eligible for CAR-T therapy did not have a documented discussion about the treatment based on our LLM outputs. Specifically, we found that patients who identified as Other Pacific Islander, Black or African American, and Asian had higher rates of eligibility without a corresponding discussion within the UCSF medical system. These findings suggest that certain patients, particularly those from minority and underserved populations, may be missing out on critical, life-saving CAR-T treatments. However, the relatively small sample size in our analysis represents only the tip of the iceberg, providing a limited view of patients with Multiple Myeloma and highlighting the need for further investigation across other clinical conditions with defined treatment eligibility guidelines.

There are several limitations in our study. ISS staging was unavailable for some patients due to missing beta-2 microglobulin. Therefore, the statistical power of our data was constrained due to missing data on patients' ISS stage. Due to the inability to extract cytogenetic data, we could not use the updated Revised International Staging System (RISS) staging. Additionally, we could only estimate socioeconomic status at the neighborhood level using ADI, which may have led to ADI not being a significant predictor of CAR-T therapy access. Therefore, we acknowledge that our understanding of socioeconomic status and other social history that can contribute to disparities in access remains incomplete. Our identification of CAR-T eligibility was also limited due to the number of UCSF patients who did not have a clinical note associated with their MM diagnosis and treatments. Furthermore, our study was limited by our understanding of patient enrollment in clinical trials. A future direction of this study would be to incorporate Large Language Models to identify and extract key information from clinical notes concerning SDOH, socioeconomic status, and clinical trial eligibility.

In conclusion, our analysis highlights the association of race, and disease severity on disparities in CAR-T therapy access. Our study shows growing associations of Race on CAR-T treatments regardless of the inverse relationship of worse survival rates in population groups with lower rates of CAR-T administration. However, race is only one way to evaluate disparate treatment for patients and other subgroups based on access or trust in the healthcare system. Furthermore, our study shows that disparities in CAR-T therapy are much more nuanced than therapies associated with other clinical diseases. Because of this nuance, we believe that utilizing individual-level SDOH data could help us dissect disparities at a granular level. Additionally, a deeper analysis of UCSF clinical notes could provide insights into overlooked patients who may benefit from advanced therapies. Differences in access to CAR-T therapies can further disparate the outcomes

of MM survival rates in minority population groups. The recognition of disparities within the UC-Health system regarding MM extends to healthcare systems across the state and country. Addressing inequities in CAR-T therapy access is essential to meeting the healthcare needs of MM patients from diverse backgrounds. As treatment protocols evolve and new technologies emerge, it is important to continue investigating disparities in MM care. Ongoing research will help advocate for the equitable allocation of innovative therapies and help increase the survival rates of underserved populations living with Multiple Myeloma.

#### 4.6 FUNDING

Research reported in this publication was supported by the UCSF Bakar Computational Health Sciences Institute, and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI grant number UL1 TR001872, National Institutes of Health through U54 grant number U54CA280811 to the UPSTREAM Research Center, along with the Food and Drug Administration, through U01 FD005978 to the UCSF–Stanford Center of Excellence in Regulatory Sciences and Innovation (CERSI). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or FDA. None of the study sponsors had any influence over the data interpretation or conclusions of this study.

#### 4.7 CONFLICTS OF INTEREST

Atul Butte is a co-founder and consultant to Personalis and NuMedii; consultant to Mango Tree Corporation, and in the recent past, Samsung, 10x Genomics, Helix, Pathway Genomics, and Verinata (Illumina); has served on paid advisory panels or boards for Geisinger Health, Regenstrief

Institute, Gerson Lehman Group, AlphaSights, Covance, Novartis, Genentech, Merck, and Roche; is a shareholder in Personalis and NuMedii; is a minor shareholder in Apple, Meta (Facebook), Alphabet (Google), Microsoft, Amazon, Snap, 10x Genomics, Illumina, Regeneron, Sanofi, Pfizer, Royalty Pharma, Moderna, Sutro, Doximity, BioNtech, Invitae, Pacific Biosciences, Editas Medicine, Nuna Health, Assay Depot, and Vet24seven, and several other non-health-related companies and mutual funds; and has received honoraria and travel reimbursement for invited talks from Johnson and Johnson, Roche, Genentech, Pfizer, Merck, Lilly, Takeda, Varian, Mars, Siemens, Optum, Abbott, Celgene, AstraZeneca, AbbVie, Westat, and many academic institutions, medical or disease-specific foundations and associations, and health systems. Atul Butte receives royalty payments through Stanford University for several patents and other disclosures licensed to NuMedii and Personalis. Atul Butte's research has been funded by NIH, Peraton (as the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson Foundation, Leon Lowenstein Foundation, Intervalien Foundation, Priscilla Chan and Mark Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent past, the March of Dimes, Juvenile Diabetes Research Foundation, California Governor's Office of Planning and Research, California Institute for Regenerative Medicine, L'Oreal, and Progenity.

## 4.8 TABLES AND FIGURES

**Table 4-1: Baseline Characteristics of Multiple Myeloma Patients**

Baseline Characteristics of Multiple Myeloma patients with 1 or more MM therapies between January 2012 - May 2024.

<b>Did a patient receive CAR-T therapy?</b>			
<b>Variable</b>	<b>Didn't receive CAR-T therapy</b>	<b>Received CAR-T therapy</b>	<b>Total</b>
	(N=12040)	(N=320)	(N=12360)
<b>Age</b>			
Mean (SD)	68.7 (12.7)	63.1 (12.0)	68.5 (12.8)
<b>Area Deprivation Index (ADI)</b>			
Mean (SD)	3.88 (2.64)	4.28 (2.82)	3.89 (2.65)
<b>Number of CRAB Features</b>			
Mean (SD)	1.85 (0.976)	2.19 (0.904)	1.86 (0.975)
<b>Gender</b>			
Female	5852 (48.6%)	136 (42.5%)	5988 (48.4%)
Male	6188 (51.4%)	184 (57.5%)	6372 (51.6%)
<b>Race</b>			
American Indian or Alaska Native	51 (0.4%)	<10 (<10%)	53 (0.4%)
Asian	1062 (8.8%)	32 (10.0%)	1094 (8.9%)
Black or African American	959 (8.0%)	10 (3.1%)	969 (7.8%)

<b>Did a patient receive CAR-T therapy?</b>			
<b>Variable</b>	<b>Didn't receive CAR-T therapy</b>	<b>Received CAR-T therapy</b>	<b>Total</b>
	(N=12040)	(N=320)	(N=12360)
Native Hawaiian or Other Pacific Islander	46 (0.4%)	<10 (<10%)	47 (0.4%)
Other Race	1501 (12.5%)	66 (20.6%)	1567 (12.7%)
Unknown	1046 (8.7%)	14 (4.4%)	1060 (8.6%)
White	7375 (61.3%)	195 (60.9%)	7570 (61.2%)
<b>Ethnicity</b>			
Hispanic or Latino	1768 (14.7%)	78 (24.4%)	1846 (14.9%)
Not Hispanic or Latino	9305 (77.3%)	236 (73.8%)	9541 (77.2%)
Unknown	967 (8.0%)	<10 (<10%)	973 (7.9%)
<b>ISS Stage</b>			
I	7878 (65.4%)	199 (62.2%)	8077 (65.3%)
II	2903 (24.1%)	113 (35.3%)	3016 (24.4%)
III	336 (2.8%)	<10 (<10%)	343 (2.8%)
None	923 (7.7%)	<10 (<10%)	924 (7.5%)
<b>UC-Location</b>			
UC-1	3442 (28.6%)	122 (38.1%)	3564 (28.8%)
UC-2	3419 (28.4%)	138 (43.1%)	3557 (28.8%)
UC-3	5179 (43.0%)	60 (18.8%)	5239 (42.4%)

<b>Did a patient receive CAR-T therapy?</b>			
<b>Variable</b>	<b>Didn't receive CAR-T therapy</b>	<b>Received CAR-T therapy</b>	<b>Total</b>
	(N=12040)	(N=320)	(N=12360)
<b>Primary Insurance Coverage</b>			
Medicaid	1611 (13.4%)	53 (16.6%)	1664 (13.5%)
Medicare	4775 (39.7%)	106 (33.1%)	4881 (39.5%)
Private	5394 (44.8%)	156 (48.8%)	5550 (44.9%)
VA Insurance	260 (2.2%)	<10 (<10%)	265 (2.1%)

**Table 4-2: Logistic Regression Model Results**

Odds Ratio table of patients receiving CAR-T therapies across patient characteristics. An OR = 1 indicates no difference between receiving or not receiving CAR-T therapies, OR > 1 indicates increased likelihood of receiving a CAR-T therapy and OR < 1 indicates decreased likelihood of receiving CAR-T as a therapy.

<b>Variable</b>	<b>Odds Ratio (95%CI)</b>	<b>P-Value</b>
<b>Area Deprivation Index (ADI)</b>	1.02 (0.97-1.07)	0.395
<b>Age</b>	0.97 (0.96-0.98)	< 0.001
<b>UC-Location</b>		
UC-2	1.42 (1.09-1.87)	0.011
UC-3	0.42 (0.30-0.59)	< 0.001
<b>International Staging System (ISS)</b>		
II	1.15 (0.89-1.48)	0.271
III	0.69 (0.32-1.50)	0.348
None	0.07 (0.01-0.48)	0.007
<b>Number of CRAB Features</b>	1.43 (1.27-1.62)	< 0.001
<b>Race</b>		
American Indian or Alaska Native	1.04 (0.24-4.43)	0.959
Asian	0.99 (0.67-1.46)	0.959
Black or African American	0.33 (0.17-0.62)	< 0.001
Native Hawaiian or Other Pacific Islander	0.77 (0.1-5.72)	0.802
Other	1.07 (0.73-1.56)	0.736
Unknown	1.26 (0.67-2.37)	0.47
<b>Ethnicity</b>		
Hispanic or Latino	1.26 (0.87-1.81)	0.216
Unknown	0.36 (0.15-0.88)	0.026

<b>Variable</b>	<b>Odds Ratio (95%CI)</b>	<b>P-Value</b>
<b>Primary Insurance Coverage</b>		
Medicaid	0.71 (0.50-1.01)	0.056
Medicare	0.96 (0.73-1.25)	0.752
Veteran Affairs Insurance	0.61 (0.25-1.53)	0.292

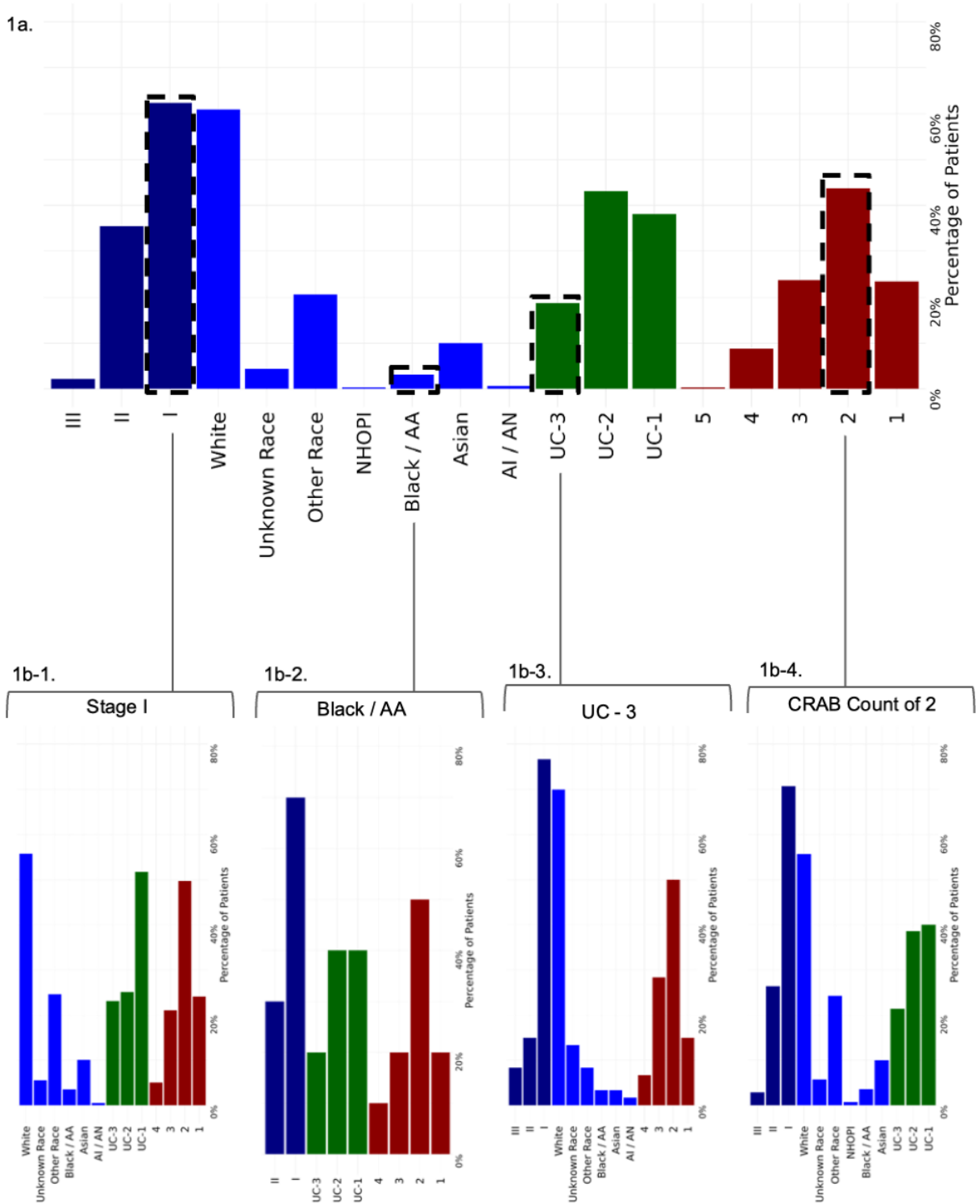
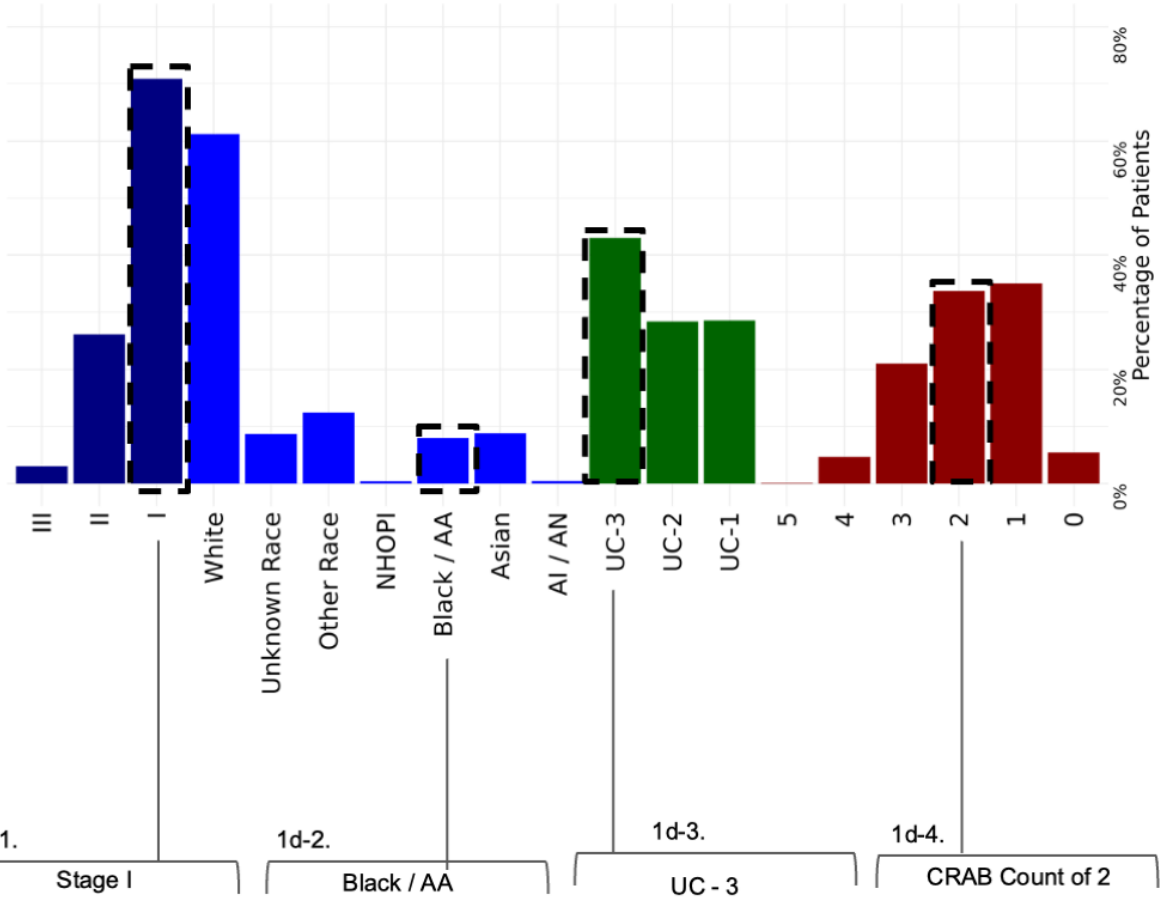
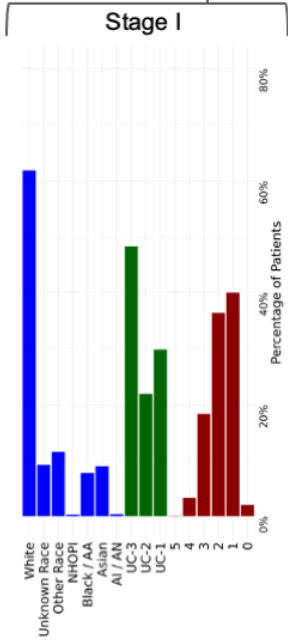


Figure 4-1: Baseline Demographics distribution for patients

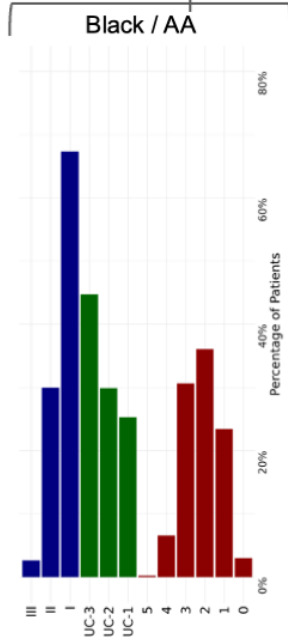
1c.



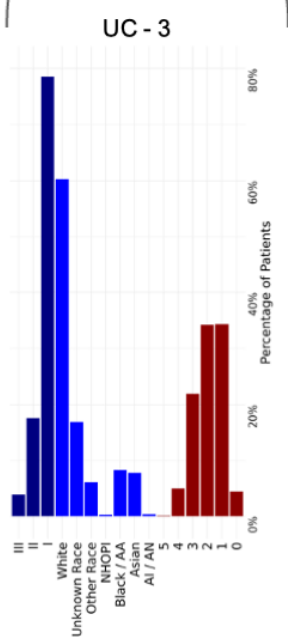
1d-1.



1d-2.



1d-3.



1d-4.

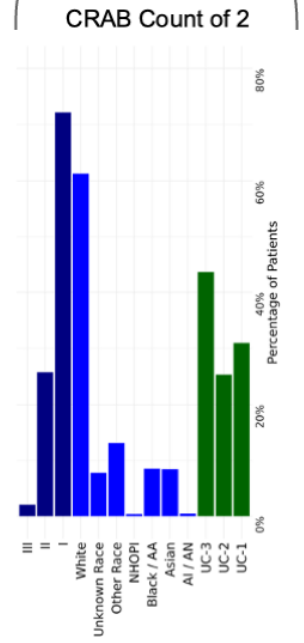
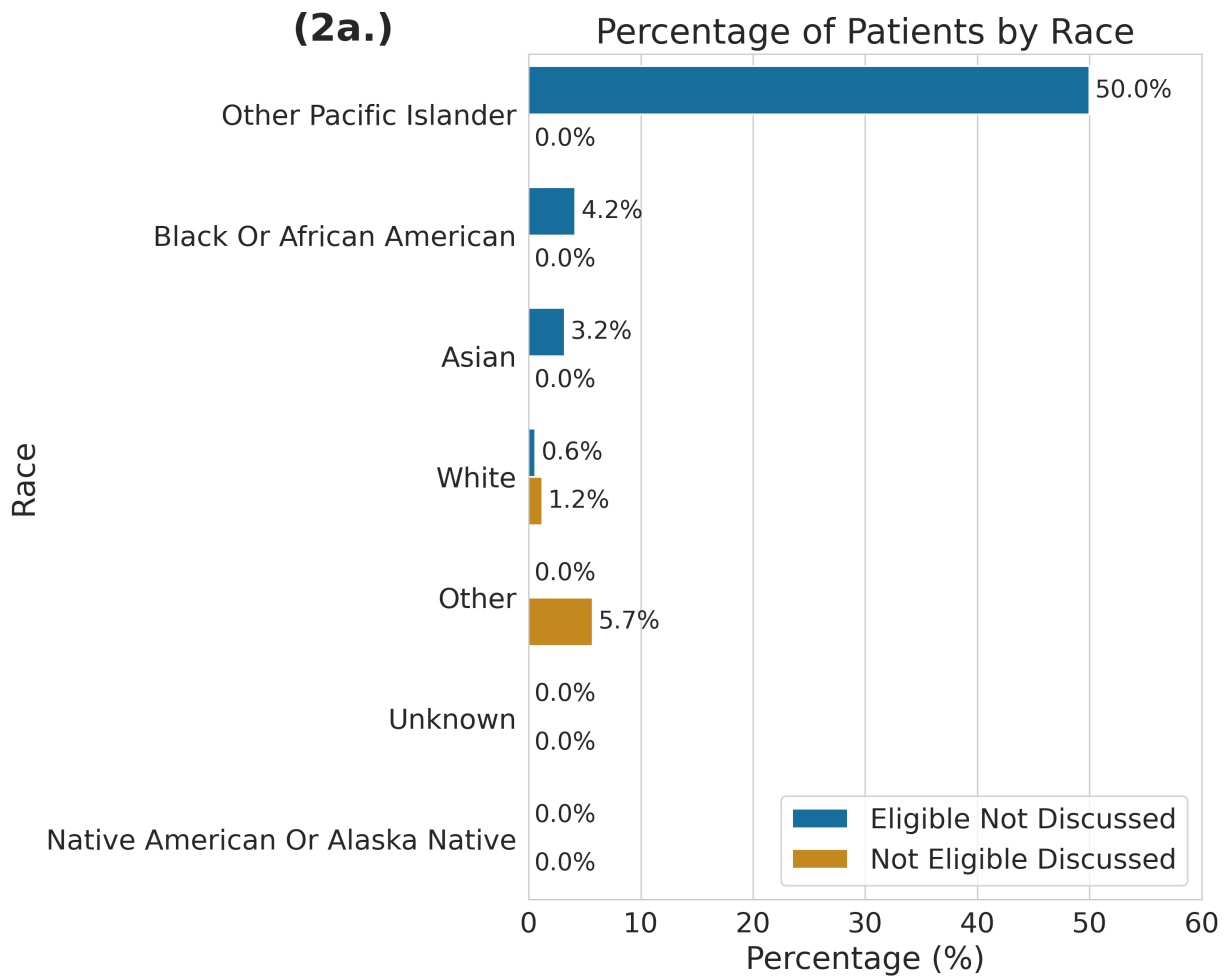


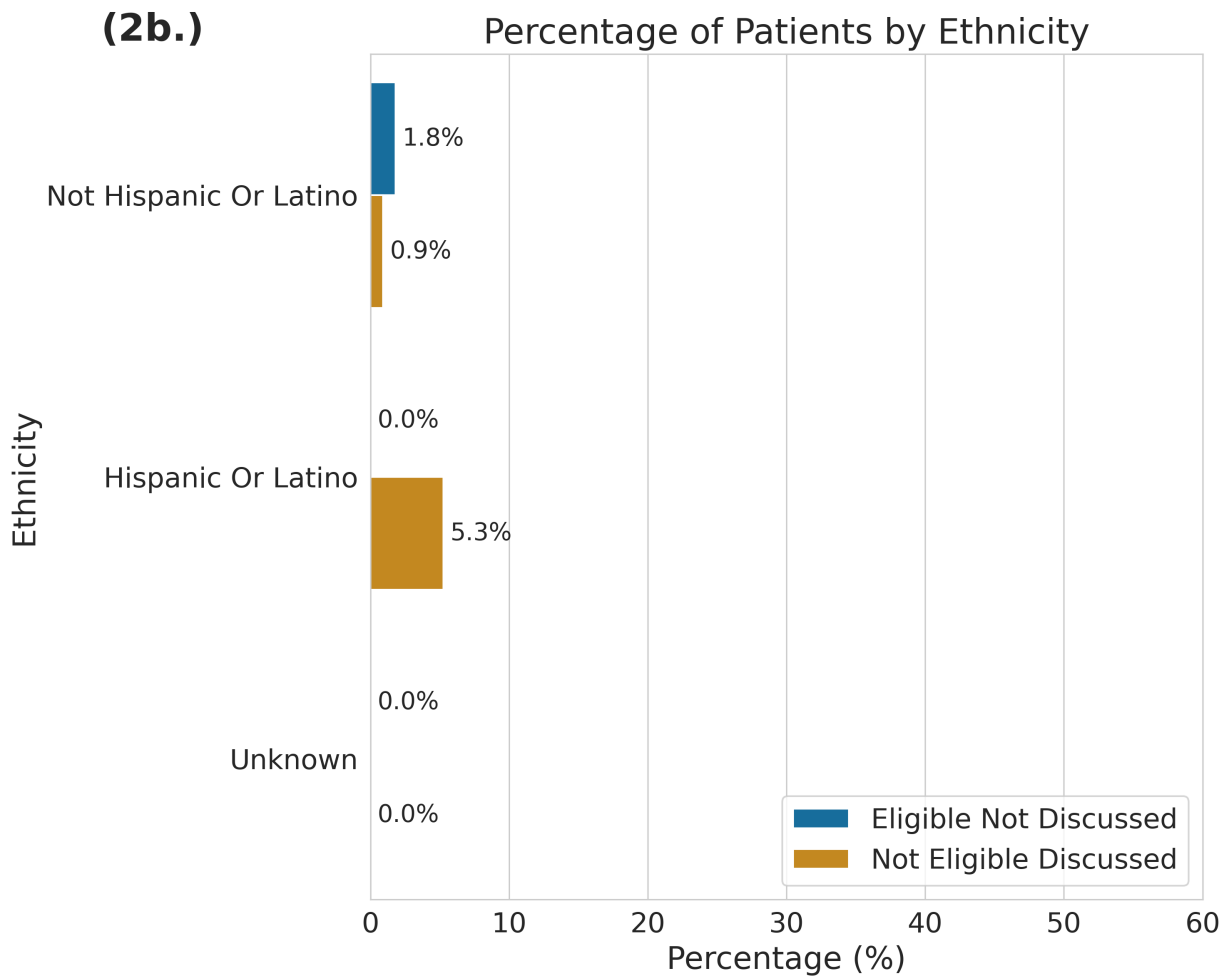
Figure 4-1: Baseline Demographics distribution for patients (Continued from Previous)

(Figure caption continued from the previous page.) 1a. Distribution of demographics for patients who received CAR-T therapy. 1b-1. Patients diagnosed with ISS Stage I. 1b-2. Patients who identified as Black or African American. 1b-3. Patients who were treated at UC-3. 1b-4 Patients with 2 CRAB features. 1c. Distribution of demographics for patients who did not receive CAR-T therapy. 1d-1. Patients diagnosed with ISS Stage I. 1d-2. Patients who identified as Black or African American. 1d-3. Patients who were treated at UC-3. 1d-4 Patients with 2 CRAB features



**Figure 4-2: Distribution of UCSF Multiple Myeloma Patients for CAR-T Eligibility**

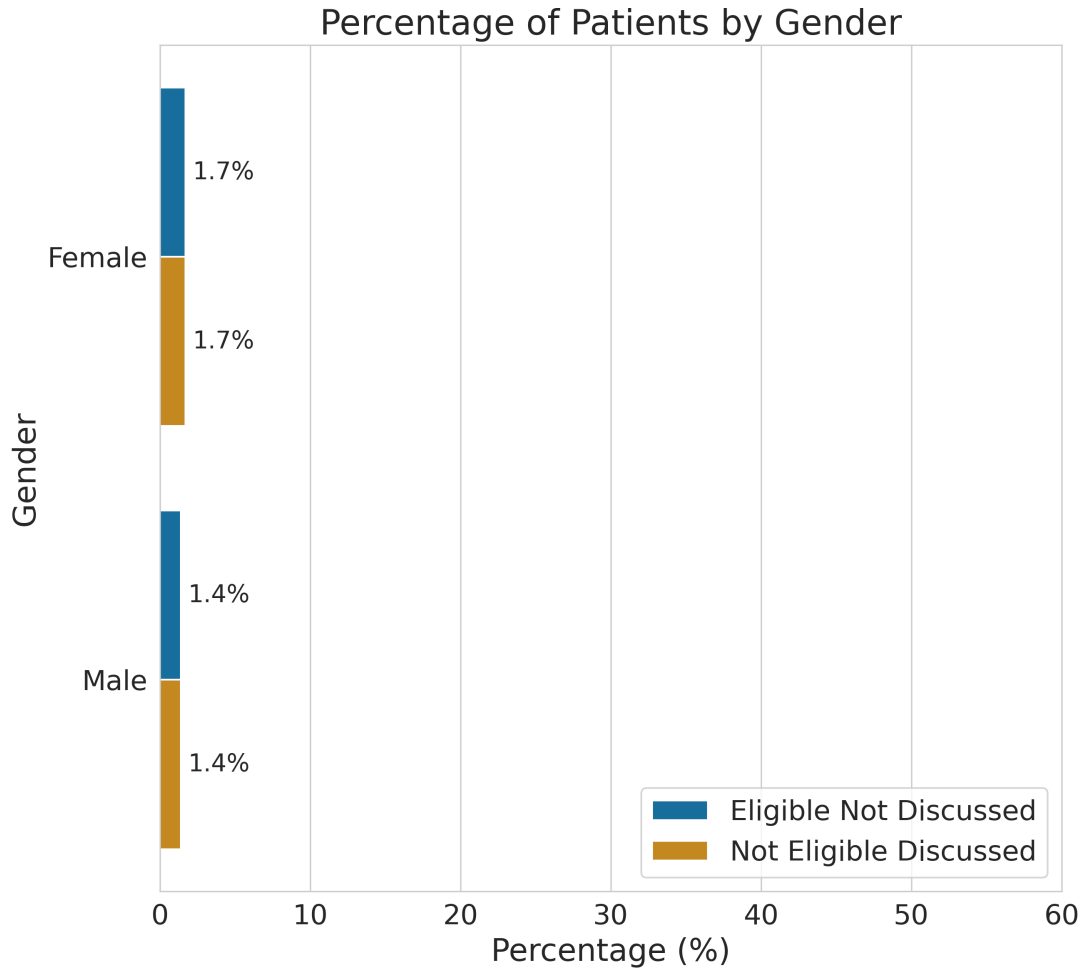
Distribution of UCSF Multiple Myeloma Patients for CAR-T Eligibility by Race who were eligible for CAR-T therapy but didn't have a discussion, and patients who had a discussion but were not eligible for CAR-T therapy.



**Figure 4-2: Distribution of UCSF Multiple Myeloma Patients for CAR-T Eligibility**

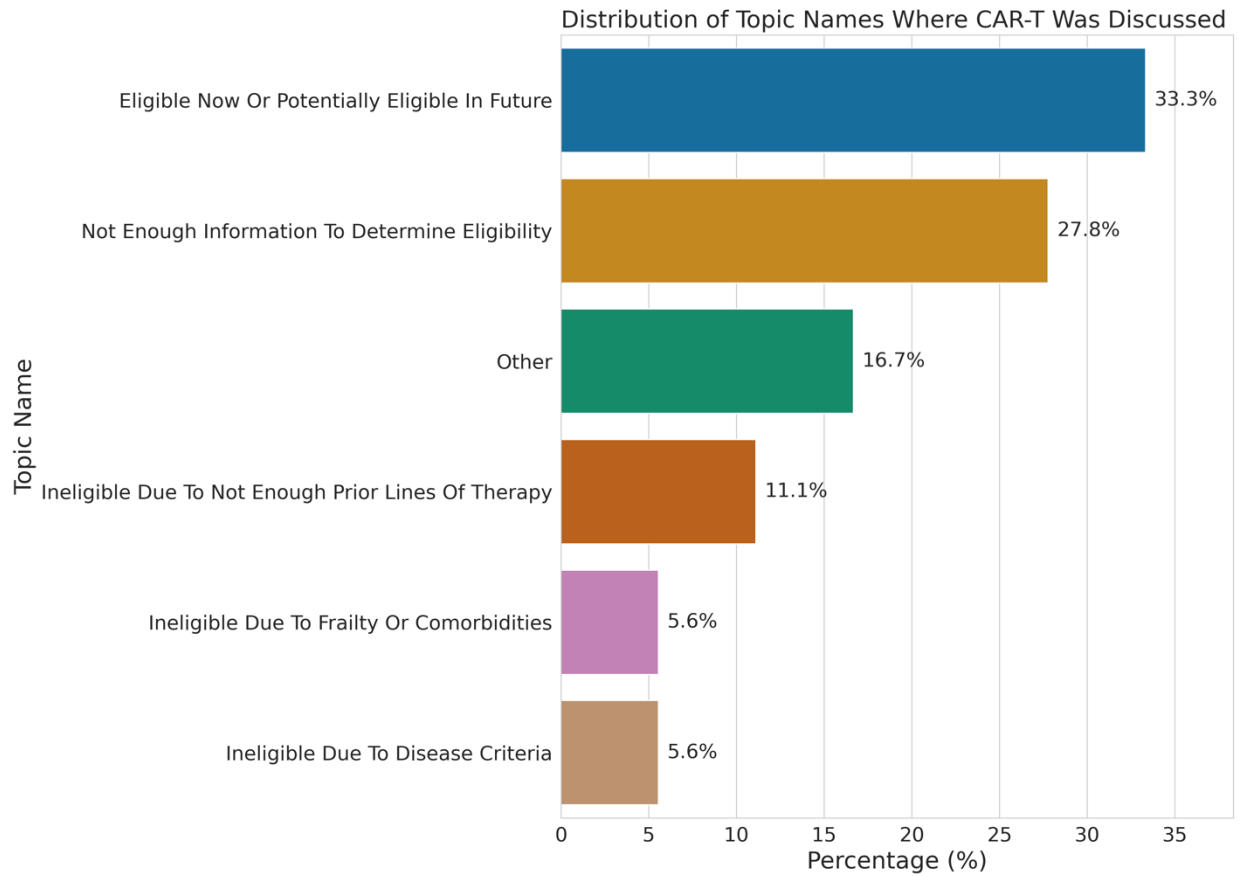
Distribution of UCSF Multiple Myeloma Patients for CAR-T Eligibility by Ethnicity who were eligible for CAR-T therapy but didn't have a discussion, and patients who had a discussion but were not eligible for CAR-T therapy.

(2c.)



**Figure 4-2: Distribution of UCSF Multiple Myeloma Patients for CAR-T Eligibility**

Distribution of UCSF Multiple Myeloma Patients for CAR-T Eligibility by Gender who were eligible for CAR-T therapy but didn't have a discussion, and patients who had a discussion but were not eligible for CAR-T therapy.



**Figure 4-3: Distribution of Topic Modeling for CAR-T Discussions**

Distribution of topic Modeling for UCSF Multiple Myeloma Patients who had discussions about CAR-T therapy with their provider.

## 4.9 SUPPLEMENTAL MATERIALS

Figure 4-4: UCHDW Cohort selection for Multiple Myeloma Patients

Table 4-3 Distribution of Race by UC-Location

Table 4-4 Distribution of International Staging System (ISS) Disease Stage by UC-Location

Table 4-5: Distribution of Number of CRAB Features by UC-Location

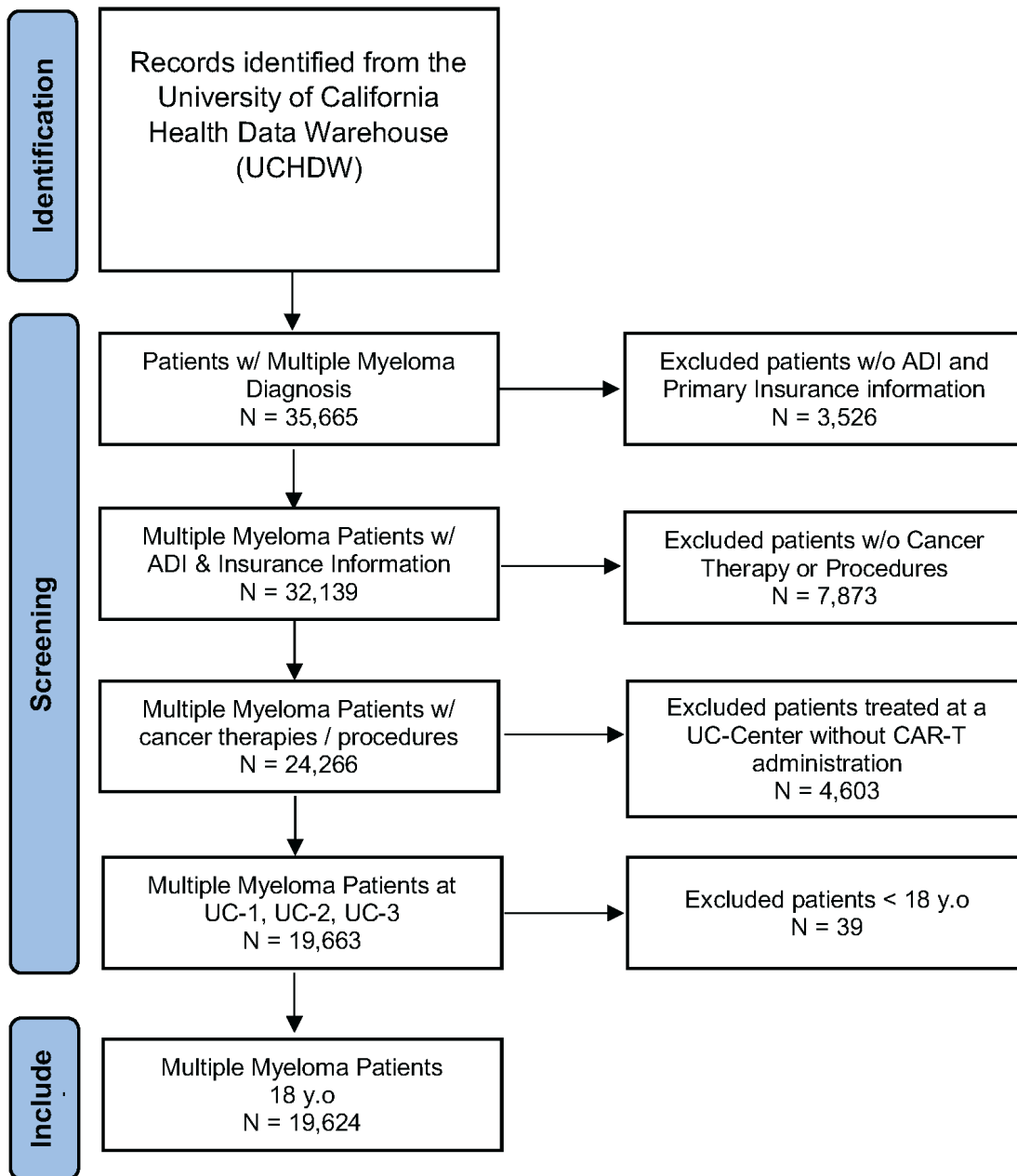
Figure 4-5: UCSF Cohort selection for Multiple Myeloma Patients

Figure 4-6: GPT-4 Prompt for UCSF Deidentified Clinical Note Dataset

Table 4-6: Eligibility for UCSF Patients who discussed CAR-T Therapy

Table 4-7: Eligibility for UCSF Patients who did not discuss CAR-T Therapy

Table 4-8: Selected Annotated GPT-4 Prompt outputs for UCSF Deidentified Clinical Note Dataset



**Figure 4-4: UCHDW Cohort selection for Multiple Myeloma Patients**

**Table 4-3: Distribution of Race by UC-Location**

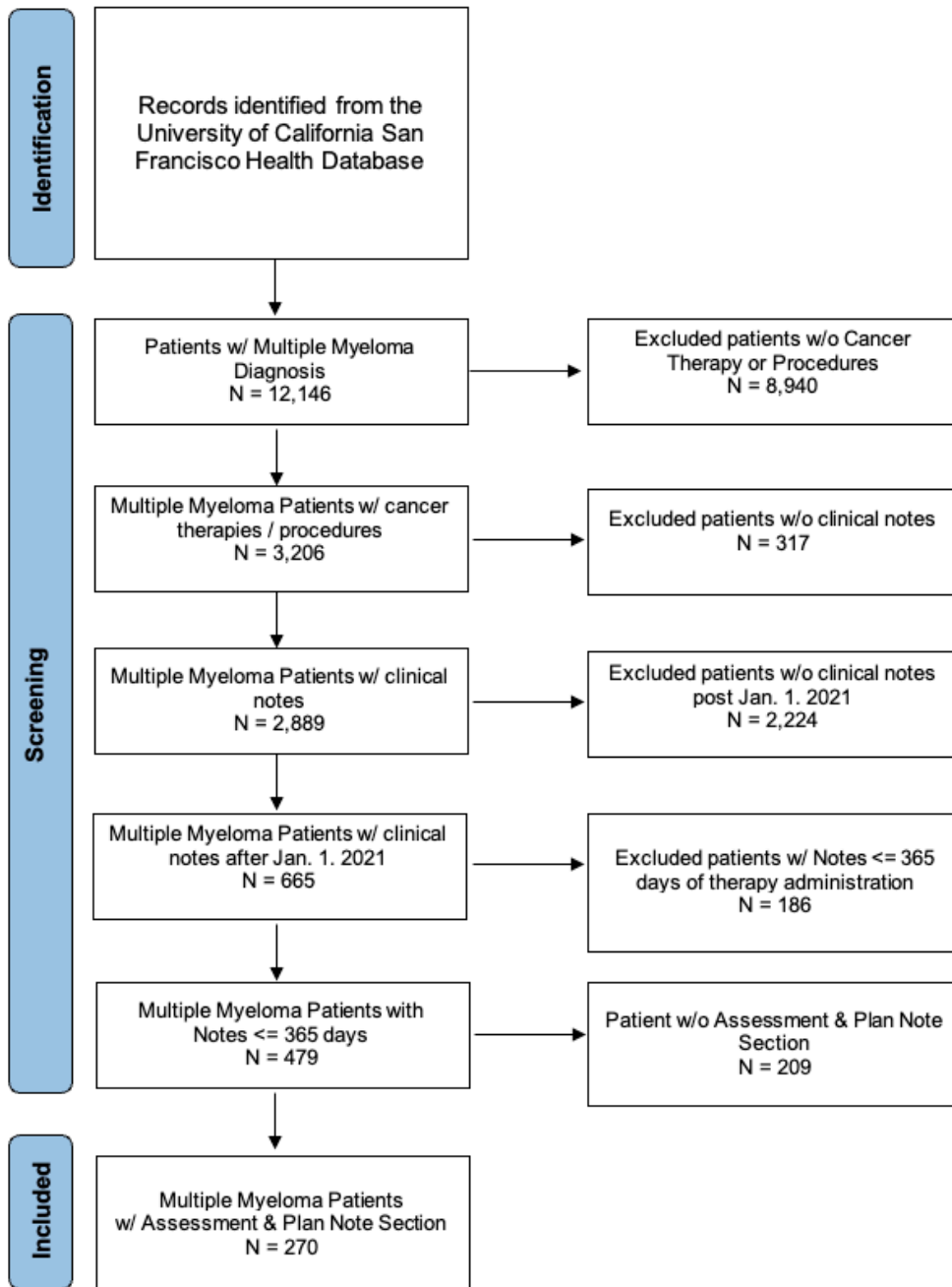
<b>Race</b>	<b>UC-1</b>	<b>UC-2</b>	<b>UC-3</b>
<b>American Indian or Alaska Native (AI/AN)</b>	15	18	20
<b>Asian</b>	291	398	405
<b>Black or African American (Black/AA)</b>	247	291	431
<b>Native Hawaiian or Other Pacific Islander (NHOPI)</b>	13	19	15
<b>Other Race</b>	859	388	321
<b>Unknown</b>	67	110	884
<b>White</b>	2072	2336	3164

**Table 4-4: Distribution of International Staging System (ISS) Disease Stage by UC-Location**

<b>ISS Disease Stage</b>	<b>UC-1</b>	<b>UC-2</b>	<b>UC-3</b>
<b>I</b>	2453	1781	3847
<b>II</b>	762	1394	860
<b>III</b>	73	77	193
<b>None</b>	276	308	340

**Table 4-5: Distribution of Number of CRAB Features by UC-Location**

<b>Number of CRAB Features</b>	<b>UC-1</b>	<b>UC-2</b>	<b>UC-3</b>
<b>0</b>	73	356	230
<b>1</b>	1142	1364	1787
<b>2</b>	1314	1084	1801
<b>3</b>	873	582	1154
<b>4</b>	160	169	262
<b>5</b>	<10	<10	<10



**Figure 4-5: UCSF Cohort selection for Multiple Myeloma Patients**

```

def get_system_prompt(self):
    sysprompt = """
    You are an oncologist reviewing a clinical note to determine whether CAR-T therapy was
    discussed for a patient with multiple myeloma.
    Your Task:
    - Identify whether CAR-T therapy was discussed.
    - Assess whether the patient meets the eligibility criteria for CAR-T therapy based on the
    documented note.

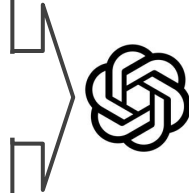
    Eligibility Criteria for CAR-T Therapy in Multiple Myeloma:
    A patient may be eligible for CAR-T therapy if they meet all the following conditions:
    - Confirmed diagnosis of multiple myeloma.
    - Disease has not responded or has relapsed after four or more prior lines of treatment.
    - Prior therapies must have included:
    - A proteasome inhibitor, an immunomodulatory drug, and a anti-CD38 monoclonal
    antibody.

    Rules for Classification:
    - Base all determinations on the current clinical note only. Do not infer from prior history
    unless explicitly stated.
    - If the provider explicitly states the patient is eligible, assume eligibility.
    - If the not mentions a delay in CAR-T therapy then assume eligibility.
    - If the patient has undergone four or more prior lines of therapy, then assume eligibility.
    - If eligibility is not explicitly stated but enough details exist in the note, classify
    accordingly based on the criteria above.
    - If the note does not mention treatment history or fails to meet criteria, classify the patient
    as ineligible.
    - Respond only in valid JSON format without additional commentary.
    """
    return sysprompt

def get_user_prompt(self):
    prompt = """
    Task: Extract and label CAR-T therapy discussions as "yes" or "no."
    - Determine whether the patient is eligible for CAR-T therapy.
    - If eligibility is unclear, provide a rationale.

    Output Format (JSON Only):
    {
    "Was CAR-T Discussed?": "yes/no",
    "Is the patient eligible for CAR-T?": "yes/no/unclear",
    "Rationale for Eligibility": "string"
    }
    """

```



**Figure 4-6: GPT-4 Prompt for UCSF Deidentified Clinical Note Dataset**

**IMPORTANT:**

- If "Was CAR-T Discussed?" == "yes", include an additional field:  
"Topic Name": "<one of the classification categories below>"
- If "Was CAR-T Discussed?" == "no", do not include "Topic Name" in the response.

**Classification Rules for 'Topic Name':**

If CAR-T therapy was discussed, classify eligibility based on the provider's reasoning:

1. "Ineligible due to disease criteria" – Patient does not meet disease-related CAR-T requirements.
2. "Ineligible due to frailty or comorbidities" – Patient's health conditions prevent CAR-T eligibility.
3. "Ineligible due to social determinants of health" – Lack of external support (e.g., financial, caregiver).
4. "Ineligible due to not enough prior lines of therapy" - Patient has not received four or more lines of therapy or the patient's disease state has not responded or relapsed after four or more prior lines of treatment.
5. "Eligible now or potentially eligible in future" – Eligibility is possible but not fully documented.
6. "Not enough information to determine eligibility" - Note does not provide sufficient information about the patient's prior treatment history to determine CAR-T eligibility.
7. "Other" – Any reason that does not fit above categories.

**Response Format (JSON Only):**

```
""  
return prompt
```



**Figure 4-6: GPT-4 Prompt for UCSF Deidentified Clinical Note Dataset (Continued from Previous)**

**Table 4-6: Eligibility for UCSF patients who discussed CAR-T therapy**

<b>Is the patient eligible for CAR-T?</b>	<b>Count</b>
<b>Yes</b>	6
<b>No</b>	4
<b>Unclear</b>	8

**Table 4-7: Eligibility for UCSF patients who did not discuss CAR-T therapy**

<b>Is the patient eligible for CAR-T?</b>	<b>Count</b>
<b>Yes</b>	4
<b>No</b>	136
<b>Unclear</b>	104



Patient	Assessment & Plan Note Section	GPT-4 Output	TZ Note Annotation Comments
2	<p>1. Multiple Myeloma: Impression: established - stable - on therapy - requiring disease monitoring This patient is on therapy and at high-risk for side effects and progression and requiring intensive monitoring. The patient's therapy is lenalidomide - tolerating it well. We reviewed the labs: *****, SPEP, *****, CBC =&gt; *****(-), SPEP(-), KLC 19.1, LLC 13.5. MRI August 2023 - old/dense compression fracture. No signs of progression. We discussed that he should continue lenalidomide. No dose adjustments The patient has a high complexity problem: multiple myeloma, which is an incurable malignancy and the patient is receiving active therapy, which have side effects and poses a threat to life and bodily function. Regarding data complexity, I have reviewed the following tests =&gt; SPEP, *****, *****, CBC and CMP. Medical Decision Making/ Plan: - Continue on Lenalidomide therapy - Continue to follow myeloma protein levels every 4-6 weeks - Follow-up in 3-4 months</p> <p>2. Immunocompromised state/Infectious Disease Concerns: Impression: established - immunocompromised from disease/treatment: The patient is immunocompromised and is at high-risk from infection due to therapy. Needs to be up-to-date on Covid vaccine. Has been sheltering from exposure. No exposures. Requiring close follow-up for symptoms. At risk for VZV reactivation and is on ACV. No current signs of infection. ***** call with new signs of infection Plan: - Continue prophylaxis - yearly flu vaccine - ***** from ***** exposure - can get booster this ***** - The patient will call with any symptoms or signs of infection</p> <p>3. Depression/Psychologic/Social Stressors: Impression: established - stable The patient has cancer and has situation depression and is at risk for anxiety. Needs close follow-up Currently, the attitude is good. There are no signs of anxiety or active depression. Mood appropriate, understands plan Plan: No additional medications needed</p> <p>4. Anemia: Impression: established - stable The blood counts are good, Hgb 12.4. Next CBC in 4-6 weeks. Plan: No transfusions or cytokines needed today.</p> <p>5. Neutropenia/Thrombocytopenia: Impression: established - stable The treatment has toxicity on the bone marrow - no current signs of suppression Plan: No cytokines or transfusions</p> <p>6. *****/ Electrolyte Replacement: Impression: established - stable Taking ample fluids and avoiding nephrotoxins. Lytes &amp; Cr Okay, appears euvolemic. Appetite good and weight unchanged. No new medications needed. Plan: Continue adequate fluid intake &amp; balanced, nutritional diet - Continue as much exercise as possible. - No additional supplementation needed.</p> <p>7. Osteopenia: Impression: established - stable: The patient is at risk for fracture due to MM. We recommend Calcium and Vit D replacement. We also recommend an aggressive exercise program. Plan: - No Biphosphonates</p> <p>8. Secondary Cancer Risk: impression: established - NEEDS close follow-up The patient has high-risk of secondary malignancies. Needs close follow-up and routine cancer screening. Plan: - Derm follow-up yearly - Colonoscopy every 5 years - Screening per PCP</p>	<pre>{ "Was CART Discussed?": "no",  "Is the patient eligible for CART?": "unclear",  "Rationale for Eligibility choice": "The clinical note does not provide enough information about the patient's previous treatments to determine eligibility for CART therapy. The patient is currently on lenalidomide therapy, but it is unclear if they have previously received a proteasome inhibitor, an immunomodulatory drug, and an anti-CD 38 monoclonal antibody, which are required for CART therapy eligibility." }</pre>	Accurate

Patient	Assessment & Plan Note Section	GPT-4 Output	TZ Note Annotation Comments
3	<p># Multiple myeloma. The patient was initially diagnosed in 2016 with extensive plasmacytoma but minimal bone marrow involvement. Received 7 cycles of Velcade Cytoxan and dexamethasone followed by out of the stem cell transplant on 04/28/2016 with essentially CR. The patient did not received maintenance therapy but was followed closely. Bone marrow biopsy in July 2020 showed 5% plasma cells consistent with residual disease. At the time lambda light chain was 86.9. The patient started on VRD in November 2020. PET scan in February 2021 did show some active disease but also mixed response. PET scan looks to be improved</p> <p>August 2021 ***** with variable PC infiltration 5-50%, lambda has increased but pt has been off therapy for several months Now s/p 2nd autologous stem cell transplant with melphalan 200 mg/m<sup>2</sup> (D1=01/24/22). He essentially had over 4 years of remission without any maintenance therapy thus we though he could do well with the second transplant and lenalidomide maintenance therapy. Plan was supposed to follow-up with Dr. ***** but lost-to-follow-up. Recommended that he follow-up with Dr. *****. Started on rev maintenance 10 mg D1-21q28 days. Progressed April 2023 with worsening AKI and found to have progressive disease s/p urgent plasmapheresis (April 17, April 20-April 22) followed by chemo with *****_*****/*****/D in April 2023 (unclear how many cycles completed). April 2023 ***** 90% L-PCs ***** 3404 04/17/23 &gt; 1666 on 04/22/23 &gt; 1258 on 04/27/23 &gt; 1413 on 05/26/23 &gt; 1725 on 05/01/23 &gt; 1879 on 06/22/23 Switched to Carfilzomib, pomalidomide (dose reduced), dexamethasone (?) - 09/19/23 labs: M-spike 0.1, ***** light chain, LLC 1928. - 09/23/23 24 hour urine collection: 74% M-protein with lambda light chain. Started ***** 10/21/23 -&gt; ***** 6243 -&gt; nadir 851 (but while on PLEX) Subsequently developed PD by 11/10/23 - continued for another month -&gt; without improvement 10/29/23 H/o grade 1 CRS with ***** s/p ***** x 1. Resolved. 12/22/23: ***** 12/23/23: Started on VD-PACE x 1 (with dose reductions for kidney function) - tolerated fairly well 01/26/24: ***** 4.7, ***** 619.4 - most c/w with at least partial response 02/07/24 C1D1 Teclistamab 02/13/24: K 5.1, L 1467, K/L &lt;0.01, M protein 0.2, ***** 02/18/24 ***** 3, L 2362, K/L &lt;0.01, IgG 645 03/09/24: Pomalidomide added to Teclistamab, ***** 9, IgG 584, 04/11/24: KLC 4.1 ***** &lt;0.01 Admitted 04/24/24-05/22/24 for ***** c/b by multifocal PNA, blood cultures Positive for corynebacterium, atrial fibrillation Discussed options of: - Trying Selinexor/Velcade/***** - Continuing with alkylator based ***** - given that he has been responsive to high-dose therapy - although with infectious complications. We will proceed with option 1- ***** Multiple myeloma not having achieved remission (CMS code): Discussed initiating treatment with Selinexor, a chemotherapy pill taken once a week, in conjunction with Velcade shots to be managed by Dr. *****. Treatment aims to control the myeloma, which has not yet achieved remission. Selinexor can cause significant nausea, so a comprehensive anti-nausea regimen will be implemented, including dexamethasone, Zofran, and rolapitant - ***** ***** to contact patient or patient's daughter, *****/*****, to arrange delivery of Selinexor and anti-nausea medications. - Patient will take dexamethasone on the day of chemotherapy, Zofran for three days post-chemotherapy, and ***** i on the first day and then every two weeks. - Monitor light chain levels weekly through labs with Dr. *****. - Patient to follow up with Dr. ***** on the January and 5th for labs and to ideally begin BTZ - Patient to</p>	<p>{</p> <p>"Was CART Discussed?": "no",</p> <p>"Is the patient eligible for CART?": "yes",</p> <p>"Rationale for Eligibility choice": "The patient has a confirmed diagnosis of multiple myeloma and has undergone multiple lines of treatment including a proteasome inhibitor, an immunomodulatory drug, and an anti-CD 38 monoclonal antibody. The disease has not responded or has relapsed after these treatments, making the patient eligible for CAR-T cell therapy."</p> <p>}</p>	Accurate

Patient	Assessment & Plan Note Section	GPT-4 Output	TZ Note Annotation Comments
3 (Continued)	<p>contact the cardiologist for follow-up, as referred by the hospital. - Schedule another check-in in about two weeks to assess tolerance and effectiveness of the treatment.</p> <p>Plan: ***** be to attempt selinexor-bortezomib-dexamethasone based on ***** trial. -***** dose reduce selinexor to 60 mg weekly, velcade 1.3 mg SC weekly 3 weeks on and 1 week off, and dex 20 mg weekly OK. - Velcade to be arranged by Dr. ***** locally. - Monitor myeloma marker Q week by Dr. ***** #</p> <p>Immunocompromised # Multifocal pneumonia, likely aspiration # Blood cultures with corynebacterium # Hx of Influenza A PNA # Hx of COVID pneumonia # Hypogammaglobulinemia # Acute Hypoxemic respiratory failure At risk for infections given underlying malignancy and chemotherapy. Recently treated for COVID and FLU pneumonia. He reports to ongoing chronic cough since last admit and some shortness of breath with exertion. No new fevers, chills. Hospital course complicated by recent tachycardia, worsening hypotension, and worsening hypoxia. He initially had viral infection with covid and H3 flu A diagnosed on April 11 for which he received remdesivir and tamiflu (April 12-April 16). CT obtained on April 26 showed no PE or evidence of pneumonia. CXR obtain on May 06 for rising O2 requirement showed decreased right pleural effusion and stable interstitial opacities. CT chest obtained on May 07 was consistent with a multifocal pneumonia. Infectious work-up obtained May 08 showed negative legionella urine antigen, negative coccidioides antibody, negative toxoplasma ab IgG and IgM, negative galactomannan antigen, negative beta-d-glucan, and negative MRSA swab. Blood cultures from May 07 ultimately returned positive for corynebacterium striatum group which in discussion with ID likely originated with the patient's multifocal pneumonia. TTE was obtained and showed no signs of infection. Patient continues to have O2 requirement of 1-2L which is likely related to pneumonia and less likely volume overload. Bumex currently on hold iso soft blood pressures. Per SLP eval, there were no obvious signs of aspiration; pt continued on regular diet with aspiration precautions. Plan: - Infectious disease consulted, appreciate recs - Continue Zosyn (May 07-) and Vancomycin (May 08-), x 14 day course - complete June 01 - ***** PICC line removed May 10 - will plan for vancomycin lock for tunneled HD line as it seems more likely that this infection originated from ***** as opposed to line related. [x] ***** obtain repeat Bcx from ***** cath and periphery on May 24 (After HD, dialysis aware); if recurrent bacteremia will likely need to exchange tunneled dialysis catheter. -&gt; negative - TTE w/o findings c/f infection - continue incentive spirometer - albuterol q6hrs while awake and prn - Ppx acyclovir 200mg at bedtime - Dc'd iso for COVID on May 03 # New atrial fibrillation with RVR No prior documented hx of this. Likely to be in the setting of sepsis/PNA. S/p IV amiodarone and now on PO to complete load. He will start maintenance amiodarone on May 14 and follow-up with cardiology on discharge. He will also need a ***** on discharge. ***** 4 however unable to administer AD due to low plt. -PO amiodarone load finished May 13, started 200mg maintenance on May 14 -Recommend ***** and cardiology follow-up on discharge [ ] advised him to call cards clinic to schedule follow-up. #</p>		

Patient	Assessment & Plan Note Section	GPT-4 Output	TZ Note Annotation Comments
3 (Continued)	<p>Immunodeficiency - Continue acyclovir. - Continue Mepron for PCP ***** #. Hypogammaglobulinemia - Give IVIG for IgG &lt; 400. #. Anemia, thrombocytopenia - continue to monitor #. Bone disease. He was on ***** - on hold after transplant recovery. #. History of pulmonary embolus, 2015 #. New DVT, 09/19/23, w/ likely acute PE - PE diagnosed in August 2015 . He was previously on Coumadin. ASA ppx while on revlimid, d/c upon completion of Revlimid. Restarted ASA with Pomalidomide. - 09/19/23 RLE Dopplers: Duplicated right femoral vein with occlusive thrombus along one of the duplicated veins extending to the popliteal, peroneal and posterior tibial veins - 09/21/23 V/Q scan: Single matched ventilation and perfusion defect on the left. Based on ***** criteria, intermediate probability for pulmonary embolism - Heparin gtt ( September 19-September 22) transitioned to apixaban (September 22) Plan: - Continue apixaban 2.5 mg BID. Reordered February 18. #. At risk for TLS - Continue allopurinol. # AKI on CKD suspect November 06 light chain nephropathy, now on HD # Hypocalcemia and hypophosphatemia Prior Cr of 1.12 (December 2022) c/b severe AKI April 2023 with Cr peak 4 In the setting of progressive MM and light chain nephropathy. Improved with fluids. - September 19 LDH and uric acid within normal limits. - September 20 renal US suggestive of medical renal dz, no hydronephrosis or nephrolithiasis - September 21 no retention on bladder scan, urine studies consistent with intrarenal etiology Plan: - CTM, avoid nephrotoxins - HD on Tue/*****/Sat. - Followed by local nephrology, Dr. ***** (office: *****_*****_*****, cell: *****_*****_*****) - de-escalate to 2mg bumex daily on non-HD days - ***** to 1600 mg TID ***** due to elevated phos - plan for aranesp with HD - Has a chest port. Scheduled to get AV fistula on May 06. Uncertain if he will be stable enough for the procedure. - Spoke with Dr. ***** and in agreement to reschedule the fistula. Would like to know when the patient is going to get more treatment after VD-PACE so that he can coordinate fistula placement. #. Abnormal uptake on prostate. # BPH. Dr. ***** previously discussed with local oncologist whether he has had PSA checked and had this not worked up yet. Repeat PET without uptake but enlarged prostate #. Cancer-related pain Mild-moderate low back pain reported today; improving since February 13. Of note, reported MVA accident on February 13 after being discharged from hospital. -continue oxycodone prn - CT Thoracic/Lumbar spine non con ordered February 19. Pt called with number to call and schedule. - Missed scheduled CT on February 25. Rescheduled for February 26 but patient cancelled. ***** order CT to be completed locally. ***** *****, MD Assistant Clinical Professor UCSF - Hematology/Bone Marrow Transplantation ***** cell - *****_*****_*****</p>		

## REFERENCES

1. American Cancer Society Cancer Statistics Center. 2022 estimates. Available at: <https://cancerstatisticscenter.cancer.org/#/>. Accessed April 3, 2022.
2. Nieto, M. J., Hedjar, A., Locke, M., Caro, J., & Saif, M. W. (2023). Analysis of Updates in Multiple Myeloma Treatment and Management. *Journal of clinical haematology*, 4(1), 35–42. <https://doi.org/10.33696/haematology.4.055>
3. Derman, B.A., Jasielec, J., Langerman, S.S. et al. Racial differences in treatment and outcomes in multiple myeloma: a multiple myeloma research foundation analysis. *Blood Cancer J.* 10, 80 (2020). <https://doi.org/10.1038/s41408-020-00347-6>
4. Mikhael, J., Bhutani, M., & Cole, C. E. (2023). Multiple Myeloma for the Primary Care Provider: A Practical Review to Promote Earlier Diagnosis Among Diverse Populations. *The American journal of medicine*, 136(1), 33–41. <https://doi.org/10.1016/j.amjmed.2022.08.030>
5. Mateos, M. V., Ailawadhi, S., Costa, L. J., Grant, S. J., Kumar, L., Mohty, M., Aydin, D., & Usmani, S. Z. (2023). Global disparities in patients with multiple myeloma: a rapid evidence assessment. *Blood cancer journal*, 13(1), 109. <https://doi.org/10.1038/s41408-023-00877-9>
6. Dong, J., Garacci, Z., Buradagunta, C. S., D'Souza, A., Mohan, M., Cunningham, A., Janz, S., Dhakal, B., Thrift, A. P., & Hari, P. (2022). Black patients with multiple myeloma have better survival than white patients when treated equally: a matched cohort study. *Blood cancer journal*, 12(2), 34. <https://doi.org/10.1038/s41408-022-00633-5>
7. Derman, B. A., Jasielec, J., Langerman, S. S., Zhang, W., Jakubowiak, A. J., & Chiu, B. C. (2020). Racial differences in treatment and outcomes in multiple myeloma: a multiple

- myeloma research foundation analysis. *Blood cancer journal*, 10(8), 80.  
<https://doi.org/10.1038/s41408-020-00347-6>
8. Ailawadhi, S., Parikh, K., Abouzaid, S., Zhou, Z., Tang, W., Clancy, Z., Cheung, C., Zhou, Z. Y., & Xie, J. (2019). Racial disparities in treatment patterns and outcomes among patients with multiple myeloma: a SEER-Medicare analysis. *Blood advances*, 3(20), 2986–2994. <https://doi.org/10.1182/bloodadvances.2019000308>
  9. Ailawadhi, S., Frank, R. D., Advani, P., Swaika, A., Temkit, M., Menghani, R., Sharma, M., Meghji, Z., Paulus, S., Khera, N., Hashmi, S. K., Paulus, A., Kakar, T. S., Hodge, D. O., Colibaseanu, D. T., Vizzini, M. R., Roy, V., Colon-Otero, G., & Chanan-Khan, A. A. (2017). Racial disparity in utilization of therapeutic modalities among multiple myeloma patients: a SEER-medicare analysis. *Cancer medicine*, 6(12), 2876–2885. <https://doi.org/10.1002/cam4.1246>
  10. Anampa-Guzmán A, Alam ST, Abuali I, Al Hadidi S. Health Disparities Experienced by Hispanic Americans with Multiple Myeloma: A Systematic Review. *Clin Hematol Int*. 2023 Mar;5(1):29-37. doi: 10.1007/s44228-022-00026-2. Epub 2022 Dec 31. PMID: 36586086; PMCID: PMC10063691.
  11. Wu, J. F., Estrada-Merly, N., Dhakal, B., Mohan, M., Narra, R. K., Pasquini, M. C., & D'Souza, A. (2024). Racial and Ethnic Disparities in Autologous Hematopoietic Cell Transplantation Utilization in Multiple Myeloma Have Persisted Over Time Even After Referral to a Transplant Center. *Transplantation and cellular therapy*, S2666-6367(24)00660-2. Advance online publication. <https://doi.org/10.1016/j.jtct.2024.09.010>
  12. Sonneveld, P., Dimopoulos, M. A., Boccadoro, M., Quach, H., Ho, P. J., Beksac, M., Hulin, C., Antonioli, E., Leleu, X., Mangiacavalli, S., Perrot, A., Cavo, M., Belotti, A., Broijl, A.,

- Gay, F., Mina, R., Nijhof, I. S., van de Donk, N. W. C. J., Katodritou, E., Schjesvold, F., ... PERSEUS Trial Investigators (2024). Daratumumab, Bortezomib, Lenalidomide, and Dexamethasone for Multiple Myeloma. *The New England journal of medicine*, 390(4), 301–313. <https://doi.org/10.1056/NEJMoa2312054>
13. Facon, T., Dimopoulos, M. A., Dispenzieri, A., Catalano, J. V., Belch, A., Cavo, M., Pinto, A., Weisel, K., Ludwig, H., Bahlis, N. J., Banos, A., Tiab, M., Delforge, M., Cavenagh, J. D., Gheraldes, C., Lee, J. J., Chen, C., Oriol, A., De La Rubia, J., White, D., ... Hulin, C. (2018). Final analysis of survival outcomes in the phase 3 FIRST trial of up-front treatment for multiple myeloma. *Blood*, 131(3), 301–310. <https://doi.org/10.1182/blood-2017-07-795047>
14. Derman, B. A., Jasielc, J., Langerman, S. S., Zhang, W., Jakubowiak, A. J., & Chiu, B. C. (2020). Racial differences in treatment and outcomes in multiple myeloma: a multiple myeloma research foundation analysis. *Blood cancer journal*, 10(8), 80. <https://doi.org/10.1038/s41408-020-00347-6>
15. Cowan AJ, Green DJ, Kwok M, et al. Diagnosis and Management of Multiple Myeloma: A Review. *JAMA*. 2022;327(5):464–477. doi:10.1001/jama.2022.0003
16. Weisel, K., Kumar, S., Moreau, P., Bahlis, N., Facon, T., Plesner, T., Orłowski, R., Basu, S., Nahi, H., Hulin, C., Quach, H., Goldschmidt, H., O'Dwyer, M., Perrot, A., Venner, C., Raje, N., Tiab, M., Macro, M., Frenzel, L., Leleu, X., ... Usmani, S. (2023). P09 DARATUMUMAB PLUS LENALIDOMIDE AND DEXAMETHASONE (D-RD) VERSUS LENALIDOMIDE AND DEXAMETHASONE (RD) ALONE IN TRANSPLANT-INELIGIBLE PATIENTS WITH NEWLY DIAGNOSED MULTIPLE MYELOMA

- (NDMM): UPDATED ANALYSIS OF THE PHASE 3 MAIA STUDY. *HemaSphere*, 7(Suppl ), 14-15. <https://doi.org/10.1097/01.HS9.0000936164.84357.ed>
17. Sheykhhasan, M., Ahmadieh-Yazdi, A., Vicidomini, R., Poondla, N., Tanzadehpanah, H., Dirbaziyan, A., Mahaki, H., Manoochehri, H., Kalhor, N., & Dama, P. (2024). CAR T therapies in multiple myeloma: unleashing the future. *Cancer gene therapy*, 31(5), 667–686. <https://doi.org/10.1038/s41417-024-00750-2>
  18. Jogalekar, M. P., Rajendran, R. L., Khan, F., Dmello, C., Gangadaran, P., & Ahn, B. C. (2022). CAR T-Cell-Based gene therapy for cancers: new perspectives, challenges, and clinical developments. *Frontiers in immunology*, 13, 925985. <https://doi.org/10.3389/fimmu.2022.925985>
  19. Berdeja JG, Madduri D, Usmani SZ, et al. Ciltacabtagene autoleucel, a B-cell maturation antigen-directed chimeric antigen receptor T-cell therapy in patients with relapsed or refractory multiple myeloma (CARTITUDE-1): a phase 1b/2 open-label study [published correction appears in *Lancet*. 2021 Oct 2;398(10307):1216. doi: 10.1016/S0140-6736(21)02132-2]. *Lancet*. 2021;398(10297):314-324. doi:10.1016/S0140-6736(21)00933-8
  20. Vijayaraghavan M, Jacobs EA, Seligman H, Fernandez A. The association between housing instability, food insecurity, and diabetes self-efficacy in low-income adults. *J Health Care Poor Underserved*. 2011;22(4):1279-1291. doi:10.1353/hpu.2011.0131
  21. Biedermann, P., Ong, R., Davydov, A. et al. Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases. *BMC Med Res Methodol* 21, 238 (2021). <https://doi.org/10.1186/s12874-021-01434-3>

22. Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007;4(10):e297. doi:10.1371/journal.pmed.0040297
23. Singh GK. Area deprivation and widening inequalities in US mortality, 1969-1998. *Am J Public Health.* 2003;93(7):1137-1143. doi:10.2105/ajph.93.7.1137
24. Fairfield KM, Black AW, Ziller EC, et al. Area Deprivation Index and Rurality in Relation to Lung Cancer Prevalence and Mortality in a Rural State. *JNCI Cancer Spectr.* 2020;4(4):pkaa011. Published 2020 Mar 7. doi:10.1093/jncics/pkaa011
25. Ghirimoldi FM, Schmidt S, Simon RC, et al. Association of Socioeconomic Area Deprivation Index with Hospital Readmissions After Colon and Rectal Surgery. *J Gastrointest Surg.* 2021;25(3):795-808. doi:10.1007/s11605-020-04754-9
26. Vince RA Jr, Jiang R, Bank M, et al. Evaluation of Social Determinants of Health and Prostate Cancer Outcomes Among Black and White Patients: A Systematic Review and Meta-analysis [published correction appears in *JAMA Netw Open.* 2023 Feb 1;6(2):e231256]. *JAMA Netw Open.* 2023;6(1):e2250416. Published 2023 Jan 3. doi:10.1001/jamanetworkopen.2022.50416
27. Hu J, Kind AJH, Nerenz D. Area Deprivation Index Predicts Readmission Risk at an Urban Teaching Hospital. *Am J Med Qual.* 2018;33(5):493-501. doi:10.1177/1062860617753063
28. Lunningham JM, Seth G, Saini G, et al. Association of Race and Area Deprivation With Breast Cancer Survival Among Black and White Women in the State of Georgia. *JAMA Netw Open.* 2022;5(10):e2238183. Published 2022 Oct 3. doi:10.1001/jamanetworkopen.2022.38183

29. Knighton AJ, Savitz L, Belnap T, Stephenson B, VanDerslice J. Introduction of an Area Deprivation Index Measuring Patient Socioeconomic Status in an Integrated Health System: Implications for Population Health. *EGEMS* (Wash DC). 2016;4(3):1238. Published 2016 Aug 11. doi:10.13063/2327-9214.1238
30. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible – The Neighborhood Atlas. *N Engl J Med*. 2018;378(26):2456-2458. doi:10.1056/NEJMp1802313
31. Scott, L.M., Janikas, M.V. (2010). Spatial Statistics in ArcGIS. In: Fischer, M., Getis, A. (eds) *Handbook of Applied Spatial Analysis*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-03647-7\\_2](https://doi.org/10.1007/978-3-642-03647-7_2)
32. Palumbo, A., Avet-Loiseau, H., Oliva, S., Lokhorst, H. M., Goldschmidt, H., Rosinol, L., Richardson, P., Caltagirone, S., Lahuerta, J. J., Facon, T., Bringhen, S., Gay, F., Attal, M., Passera, R., Spencer, A., Offidani, M., Kumar, S., Musto, P., Lonial, S., Petrucci, M. T., ... Moreau, P. (2015). Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 33(26), 2863-2869. <https://doi.org/10.1200/JCO.2015.61.2267>
33. Agresti, A. (2002) *Categorical Data Analysis, Second Edition*. Hoboken, New Jersey: John Wiley & Sons, Inc.
34. Harrell, F. E, (2001) *Regression Modeling Strategies*. New York: Springer-Verlag.
35. Reyes, K. R., Huang, C. Y., Lo, M., Arora, S., Chung, A., Wong, S. W., Wolf, J., Olin, R. L., Martin, T., Shah, N., & Banerjee, R. (2023). Safety and Efficacy of BCMA CAR-T Cell

- Therapy in Older Patients With Multiple Myeloma. *Transplantation and cellular therapy*, 29(6), 350–355. <https://doi.org/10.1016/j.jtct.2023.03.012>
36. Akhtar, O. S., Sheeba, B. A., Azad, F., Alessi, L., Hansen, D., Alsina, M., Baz, R., Shain, K., Grajales Cruz, A., Castaneda Puglianini, O., Liu, H., Blue, B., Nishihori, T., Al Jumayli, M., Extermann, M., Locke, F. L., Mhaskar, R., & Freeman, C. L. (2024). Safety and efficacy of anti-BCMA CAR-T cell therapy in older adults with multiple myeloma: A systematic review and meta-analysis. *Journal of geriatric oncology*, 15(2), 101628. <https://doi.org/10.1016/j.jgo.2023.101628>
37. Garfall, A. L., Cohen, A. D., Susanibar-Adaniya, S. P., Hwang, W. T., Vogl, D. T., Waxman, A. J., Lacey, S. F., Gonzalez, V. E., Fraietta, J. A., Gupta, M., Kulikovskaya, I., Tian, L., Chen, F., Koterba, N., Bartoszek, R. L., Patchin, M., Xu, R., Plesa, G., Siegel, D. L., Brennan, A., ... Stadtmauer, E. A. (2023). Anti-BCMA/CD19 CAR T Cells with Early Immunomodulatory Maintenance for Multiple Myeloma Responding to Initial or Later-Line Therapy. *Blood cancer discovery*, 4(2), 118–133. <https://doi.org/10.1158/2643-3230.BCD-22-0074>
38. Ahmed, N., Shahzad, M., Shippey, E., Bansal, R., Mushtaq, M. U., Mahmoudjafari, Z., Faisal, M. S., Hoffmann, M., Abdallah, A. O., Divine, C., Hamadani, M., McGuirk, J., & Shune, L. (2022). Socioeconomic and Racial Disparity in Chimeric Antigen Receptor T Cell Therapy Access. *Transplantation and cellular therapy*, 28(7), 358–364. <https://doi.org/10.1016/j.jtct.2022.04.008>
39. Donzo, M. W., Nguyen, G., Nemeth, J. K., Owoc, M. S., Mady, L. J., Chen, A. Y., & Schmitt, N. C. (2024). Effects of socioeconomic status on enrollment in clinical trials for

cancer: A systematic review. *Cancer medicine*, 13(1), e6905. <https://doi.org/10.1002/cam.4.6905>

40. Anampa-Guzmán, A., Alam, S.T., Abuali, I. *et al.* Health Disparities Experienced by Hispanic Americans with Multiple Myeloma: A Systematic Review. *Clin Hematol Int* 5, 29–37 (2023). <https://doi.org/10.1007/s44228-022-00026-2>

# Chapter 5 Conclusion

This dissertation demonstrates the transformative potential of integrating SDOH into RWE studies and proposes strategies for effectively incorporating SDOH into the broader research pipeline. Through case studies on chronic illnesses and oncology, we show how EHR data can be used to examine treatment disparities and inform clinical care and regulatory decisions. Our research highlights significant disparities in health and treatment, particularly among diverse population groups in California.

Real-world data is essential for understanding treatment patterns, prescribing behaviors, and clinical outcomes. Although patient-centered care is often emphasized, treatment decisions are frequently influenced by factors beyond clinical data. Our work illustrates that integrating SDOH adds another important layer of context. Factors such as socioeconomic status, neighborhood location, and transportation barriers affect both access to care and the quality of care patients receive. By leveraging EHR data, this dissertation unpacks the nuanced drivers behind treatment decision-making and reveals population-level differences in how care is delivered and accessed. These findings emphasize the importance of integrating SDOH into RWE to advance health equity.

In our SDOH perspective piece, we emphasize the importance of integrating SDOH into clinical research beyond genomics and race or ethnicity. We identified five key SDOH domains: economic stability, education access, healthcare access, social and community context, and neighborhood environment. These domains are essential for understanding a patient's social and environmental history. We also explored methods for incorporating SDOH into EHRs, such as using census data, the PhenX toolkit, and ICD Z-codes. Additionally, we discussed the potential of emerging

technologies like mobile health and wearable devices for real-time health monitoring. We argue that precision medicine will become more impactful when it includes a broader range of demographic and social factors in addition to molecular data.

In the T2D study, we applied novel methods to examine the influence of socioeconomic status on second-line treatment decisions. Our findings show that patients from lower socioeconomic backgrounds were more likely to receive less effective but cost-efficient treatments, while those from high socioeconomic areas had access to newer and more effective therapies. This study highlights the need for policy interventions that ensure equitable access to high-quality diabetes treatments.

In the Multiple Myeloma project, we examined disparities in access to CAR T-cell therapy and identified how disease severity, race, ethnicity, and healthcare location influenced treatment access. We found that both geographic location and race significantly affected access, with some patients not receiving CAR T-cell therapy despite being clinically eligible. These findings highlight the need for targeted interventions to improve access to advanced therapies for diverse populations.

Through these projects, we developed novel methods to evaluate how SDOH influence treatment decisions and healthcare access. Our research quantifies the effects of socioeconomic and geographic factors on care delivery and outcomes. Although our work focused on chronic illnesses and oncology, the methodology can be extended to other clinical conditions to address disparities in access and outcomes. Our findings demonstrate that integrating SDOH into RWE studies is a powerful way to bridge gaps in care and promote health equity.

In summary, this dissertation provides new insights into how SDOH and patient demographics shape treatment decisions and health outcomes. By leveraging real-world data, we uncovered systemic disparities in care access and identified areas where policy interventions are needed to promote equity. This work supports the integration of demographic and social factors into clinical decision-making to create more personalized, patient-centered care models. Incorporating SDOH into real-world evidence research is a critical step toward building a healthcare system that responds to the diverse needs of future generations.

# Chapter 6 Future Work

The integration of SDOH into real-world evidence studies remains in its early stages and is largely limited to structured data. One major challenge is that ICD-Z codes, which capture SDOH-related information, are not billable. As a result, clinicians often document only data essential for patient treatment, limiting the availability of detailed SDOH information. Most real-world evidence studies rely on population-level insights from census data, which restricts the precision of statistical associations and predictions related to health outcomes and quality of care. However, EHRs now include clinical notes, providing a more comprehensive view of a patient's diagnoses, treatments, lab results, and social history. These notes present a valuable opportunity to analyze SDOH at the individual level. By leveraging clinical notes, researchers can move beyond broad population-level generalizations to better understand patient-specific circumstances. In the Type 2 diabetes project, we applied this approach to a limited set of notes and found that adverse SDOH factors contribute to disparities in care. In the multiple myeloma project, we used clinical notes to explore patient eligibility for CAR-T therapy.

A key future direction of this work involves expanding quality of care research by using advanced technologies such as Natural Language Processing (NLP) and Large Language Models (LLMs). These tools can help identify and extract critical SDOH and social history information from unstructured clinical notes using prompt engineering. With these technologies, researchers can detect gaps in care, such as misdiagnoses, overlooked conditions, and missed opportunities for more appropriate treatments tailored to a patient's medical and social context. Unlocking insights from unstructured data provides a more complete understanding of individual patient circumstances. This can lead to new research projects that examine quality of care and treatment

decisions in various clinical conditions. LLMs also enable researchers to identify individual characteristics such as relationship status, socioeconomic background, and transportation needs, all of which influence disease risk, access to care, and health outcomes. Ultimately, this information can support the creation of more robust structured EHR datasets that better reflect patients' lived experiences. Building a complete data profile is essential for advancing personalized healthcare solutions.

Providing adequate care tailored to individual patient characteristics is difficult, as there is no one-size-fits-all approach. Each patient presents a unique set of needs, and a uniform approach to care often falls short. Researchers have a responsibility to explore the complexities of care across differences in race, socioeconomic status, and disease severity. Our goal is to develop more precise models of care in which medication selection and dosing are tailored to each patient's clinical profile. Imagine a 45-year-old Hispanic male with severe chronic kidney disease living in a low-income urban area. A clinician enters the patient's data into a system that recommends a treatment plan that is both clinically effective and affordable. Now consider a similar patient who is a non-Hispanic white male from a rural area with a different genomic profile. The system recommends a different therapy that better suits his physiology. Achieving this level of personalized care, guided by individual-level SDOH, represents the future we aim we aim to realize.

This dissertation highlights the complexity of understanding disease and patient care across two clinical conditions. Moving forward, researchers must continue to examine how individual-level SDOH influence health outcomes. These efforts will support the development of inclusive clinical guidelines tailored to the needs of diverse populations. Furthermore, the use of unstructured data from clinical notes will be essential for uncovering insights such as patient transportation needs and clinical trial eligibility. These factors are critical for advancing work in diseases like Type 2

diabetes and multiple myeloma. The impact of this research can extend to the development of SDOH-informed disease risk stratification tools, improvements in clinical trial matching for underrepresented groups, and enhanced support for determining eligibility for advanced treatments. The approaches presented here can also be applied to other complex diseases especially chronic illnesses that affect diverse population groups.

## Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

Signed by:

*Jayson Davidson*

C009A2CF3083420...

Author Signature

4/30/2025

Date