Title Page:

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

Title: Real-World Type 2 Diabetes Second-Line Treatment Allocation Among Patients

Authors:
Jaysón Davidson[1,2] [ORCID 0000-0001-9066-9872]
Rohit Vashisht, Ph.D.[1,2] [ORCID 0000-0002-2938-7259]
Kendra Radtke, PharmD, Ph.D.[1]
Ayan Patel[2]
Suneil K. Koliwad, M.D., Ph.D.[3,4] [ORCID 0000-0002-7367-1054]
Atul J. Butte, M.D., Ph.D.[1,2] [ORCID 0000-0002-7433-2740]

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

Corresponding Author:
Atul J. Butte M.D., Ph.D.
Email: atul.butte@ucsf.edu
Address: 490 Illinois St, Box 2933, Floor 02, Room 2207, San Francisco, CA 94158

Keywords: Type 2 diabetes, treatment decision-making, social determinants of health, healthcare disparities, area deprivation index

Word count (excluding title page, abstract, references, figures and tables): 3,484

Tables: 2, Figures: 3

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 ±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 patients. GPT-4 identified adverse social determinants of health (SDOH) across 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]). In UCSF clinical notes, 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.

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[1,2]. Present clinical treatment guidelines target achievement of glycated hemoglobin A1c (HbA1c) levels below 7% for most adults[3,4,5]. Metformin (MF) remains the first-line of therapy, however, newer second-line treatments are often needed when metformin alone fails to achieve glycemic control[6-9]. 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[9,10]. Specific SDOH[10] such as neighborhood disadvantage and socioeconomic status, along with other adverse factors, have been previously studied and shown to significantly impact diabetes management[11-14], medication adherence[15], and behaviors such as physical activity and dietary patterns[16-23]. Real-World characterization studies show wide-spread heterogeneity in the choice of second-line treatments[24-26]. 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)[27] 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.

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[28]. 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 world29. 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 **(eMethods Supp. 2),** and T2D related drug prescriptions for four second-line treatment classes (GLP-1RA, SGLT2i, DPP4i, and SU) **(eMethods Supp. 1),** 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[30].

Area Deprivation Index (ADI)

We utilized the ADI as a metric to estimate patient-level SDOH[27]. 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[31-35] 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[27,36,37].

Study Variables

To describe patient baseline characteristics, we used ADI, HbA1c, Age (in years), patient reported Sex (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[38,39]. 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), sex (Male, Female), primary insurance coverage, race, and ethnicity by using a multinomial regression model [40,41]. 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. Further, we excluded patients whose clinical notes did not contain Social History or SDOH information, resulting in a final cohort of 105 patients with 270 clinical notes **(Supplement-3)**. 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[42]. Outputs from these sample datasets were compared against the ground truth to check for accuracy in identifying SDOH factors **(Supplement-3)**. 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)[42]. 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.

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 (**Table 2**) adjusting for ADI. 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[6]. 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 a 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)**.

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[3,6,8]. 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,

280 and GLP-1RA. We also found that patients with Medicaid as their primary insurance, often
281 associated with low-cost healthcare coverage, were less likely to receive these advanced
282 second-line therapies. This disparity underlying treatment allocation is important, given that
283 SGLT2i and GLP-1RAs have clearly shown benefits in limiting development of serious
284 comorbid conditions including CKD, cardiovascular events, heart failure, and obesity[3,6]. Our
285 findings thus highlight a gap in the fair allocation of effective second-line treatments among
286 patients with T2D.
287
288 Beyond underlying socioeconomic disparities in second-line treatment allocation, we also found
289 a reliance favoring SU when the urgency for glucose lowering was highest. Indeed, in each
290 comparative analysis of specific agents as second-line treatments for T2D, we found that the
291 worse a patient's glycemic control, as measured by HbA1c, the more likely that patient was to
292 receive an SU. This reliance is in line with the traditional prescribing patterns and older ADA
293 guidelines that emphasize SU treatment as the first adjunct to MF for T2D. It also highlights the
294 notion that older more established medication classes might be favored when A1c lowering is
295 the top priority, regardless of SES. This fits with classic literature and ongoing supportive data
296 indicating that optimal HbA1c lowering remains the most important factor in limiting the
297 development of T2D-associated complications[6]. This finding is important because it highlights
298 that patients with a high ADI are more likely to have a more severe A1c at second-line
299 treatment initiation and are more likely to receive SU.
300
301 Our analysis also revealed disparities in second-line treatment allocation based on patient sex
302 and age. First, we found that males were more likely to receive a DPP4i, or a GLP-1RA over
303 SU compared to females, potentially consistent with the fact that adult males have higher T2D
304 rates than females[43]. Advancing age was associated with a higher likelihood of receiving an SU
305 over other second-line treatment choices. This finding may reflect a preference for SU among
306 aged patients to minimize risk of hypoglycemia, cognitive impairment, and autonomic symptoms
307 during management of diabetes[6].
308
309 Our study showed that patients prescribed SU as a second-line treatment had a higher
310 prevalence of documented adverse SDOH factors compared to those prescribed GLP-1RA,

311 SGLT2i, and DPP4i. More specifically, these patients faced challenges such as adverse
312 housing, relationships, transportation issues, with many also having caregiving responsibilities
313 for children. This suggests they may have been prescribed a cost-effective treatment rather
314 than a newer second-line option due to non-medical factors. Interestingly, patients prescribed
315 SGLT2i had a higher prevalence of documented adverse employment factors than those
316 prescribed other treatments, indicating that some patients facing employment challenges are
317 still being treated with newer drugs. However, the relatively low number of individual-level
318 SDOH factors captured in our data provides only a partial view of the adverse contributors
319 influencing a patient's second-line treatment prescription.
320
321 There are several limitations in the study. The health disparities associated with SES in the
322 UCHDW, as highlighted in our study, were estimated using the ADI at the neighborhood level,
323 which only provides an approximate measure of a patient's SES. Additionally, the inability to
324 capture a significant amount of individual-level SES data and other SDOH parameters due to
325 the absence of such data elements, limited a comprehensive evaluation of the patient
326 population. We acknowledge that our understanding of SES at the individual level remains
327 incomplete and could potentially be improved in future studies. The sample size of our data was
328 also constrained, primarily due to the stringent cohort selection criteria employed in this study.
329 Furthermore, our study does not account for the approval dates of newer drugs such as
330 SGLT2i, DPP4i, and GLP-1RA which were introduced later during the study period. Lastly, the
331 distribution of patients in this study was slightly skewed towards least economically
332 disadvantaged neighborhoods due to the high number of patients living in higher
333 socioeconomic areas that are treated within the UC-Health System.
334
335 Our findings serve as a preliminary step toward a deeper exploration of how individual SES and
336 SDOH can contribute to health disparities when data is collected in large quantities. We believe
337 that differences in disease severity at second-line treatment allocation may stem from factors
338 such as limited access to healthcare, financial constraints, low health literacy, and geographic
339 location of a patient. Moreover, our findings suggest that disparities in treatment allocation may
340 be influenced by treatment cost and unequal access to healthcare providers who prescribe
341 newer, more effective medications (e.g., GLP-1RA). To further understand the disparities

342 highlighted in our study, a comprehensive examination of high-quality individual-level SDOH
343 factors is needed to effectively address the underlying issues pertaining to treatment selection
344 and disease severity disparities.
345
346 In conclusion, our analysis highlights the association of SDOH with treatment choices,
347 significantly impacting disease state and health outcomes. Our study shows growing associations
348 of SDOH on clinical treatments, with and without adjusting for other covariates. Furthermore, our
349 study shows that SDOH can be utilized to understand differences in care among patients from
350 distinct population groups based on socioeconomic status. The recognition of SDOH disparities
351 within the UC-Health system for T2D extends to healthcare systems across the United States.
352 Understanding the role of SDOH, alongside considerations of disease severity, age, and sex, is
353 crucial in addressing disparities in care among diverse population groups. A comprehensive
354 understanding of treatment disparities will lead to the fair allocation of medical resources to
355 population groups in which disparities exist. Most importantly utilizing SDOH to understand
356 differences in care for multiple disease could drive institutions to improve their delivery of care to
357 disparate population groups. Therefore, we must encourage the generation of high-quality SDOH
358 data at both the neighborhood and individual level to increase the viability of clinical-related
359 questions across disease states.
360
361 **Author Contributions/Acknowledgements**
362 Conceptualization, J.D., R.V., K.R and A.J.B.; writing—original draft preparation, J.D.; writing—
363 review and editing, J.D., R.V., S.K.K, and A.J.B, Area Deprivation Index data, A.P. All authors
364 have read and agreed to the published version of the manuscript.
365
366 **Funding**
367 Research reported in this publication was supported by the UCSF Bakar Computational Health
368 Sciences Institute, and the National Center for Advancing Translational Sciences, National
369 Institutes of Health, through UCSF-CTSI grant number UL1 TR001872, along with the Food and
370 Drug Administration, through U01 FD005978 to the UCSF–Stanford Center of Excellence in
371 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.

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.

403 **References:**

404 1. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence
405 estimates for 2019 and projections for 2030 and 2045: Results from the International
406 Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract.
407 2019;157:107843. doi:10.1016/j.diabres.2019.107843
408 2. Centers for Disease Control and Prevention. National Diabetes Statistics Report website.
409 https://www.cdc.gov/diabetes/data/statistics-report/index.html. Accessed November
410 2023.
411 3. American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in
412 Diabetes-2020. Diabetes Care. 2020;43(Suppl 1):S66-S76. doi:10.2337/dc20-S006
413 4. Davies MJ, Aroda VR, Collins BS, et al. Management of hyperglycaemia in type 2
414 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and
415 the European Association for the Study of Diabetes (EASD). Diabetologia.
416 2022;65(12):1925-1966. doi:10.1007/s00125-022-05787-2
417 5. International Diabetes Federation. Recommendations for managing type 2 diabetes in
418 primary care. Accessed April 5, 2023. http://www.idf.org/managing-type2-diabetes
419 6. American Diabetes Association Professional Practice Committee. 9. Pharmacologic
420 Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes-2022.
421 Diabetes Care. 2022;45(Suppl 1):S125-S143. doi:10.2337/dc22-S009
422 7. Hemmingsen B, Lund SS, Gluud C, et al. Intensive glycemic control for patients with type
423 2 diabetes: systematic review with meta-analysis and trial sequential analysis of
424 randomised clinical trials. BMJ. 2011;343:d6898. Published 2011 Nov 24.
425 doi:10.1136/bmj.d6898
426 8. Fang M, Wang D, Coresh J, Selvin E. Trends in Diabetes Treatment and Control in U.S.
427 Adults, 1999-2018. N Engl J Med. 2021;384(23):2219-2228.
428 doi:10.1056/NEJMsa2032271
429 9. Emanuel EJ, Dellgren JL, McCoy MS, Persad G. Fair Allocation of GLP-1 and Dual GLP-
430 1-GIP Receptor Agonists. N Engl J Med. 2024;390(20):1839-1842.
431 doi:10.1056/NEJMp2400978
432 10. Hasbrouck L. Healthy People 2030: An Improved Framework. Health Educ Behav.
433 2021;48(2):113-114. doi:10.1177/1090198121997812

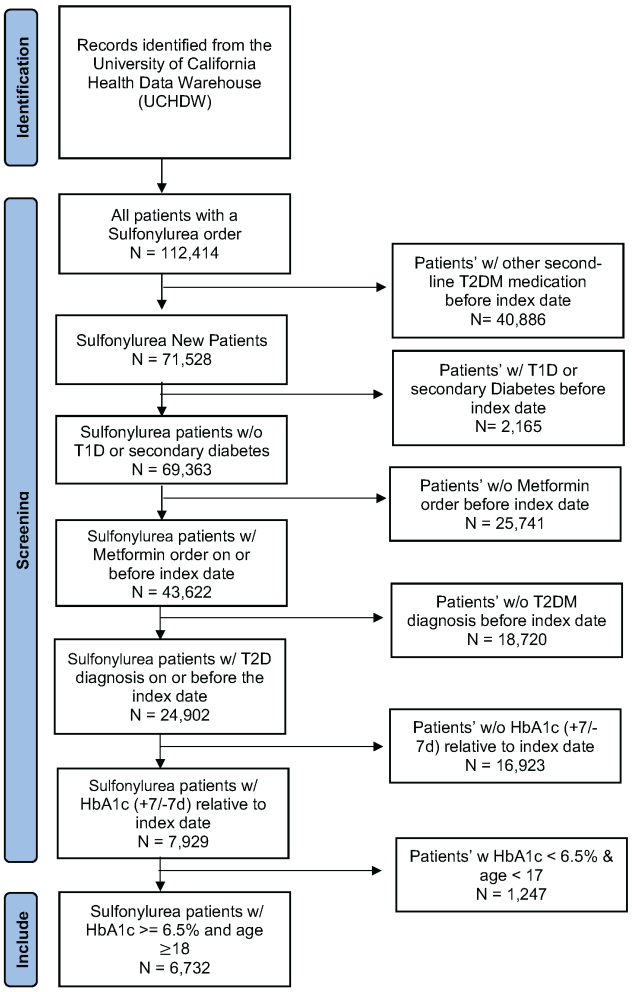
434 11. Schmidt-Busby J, Wiles J, Exeter D, Kenealy T. Understanding 'context' in the self-
435 management of type 2 diabetes with comorbidities: A systematic review and realist
436 evaluation. Diabetes Res Clin Pract. 2018;142:321-334.
437 doi:10.1016/j.diabres.2018.06.003
438 12. Ackermann RT, Holmes AM, Saha C. Designing a natural experiment to evaluate a
439 national health care-community partnership to prevent type 2 diabetes. Prev Chronic Dis.
440 2013;10:E12. doi:10.5888/pcd10.120149
441 13. de Vries McClintock HF, Wiebe DJ, O'Donnell AJ, Morales KH, Small DS, Bogner HR.
442 Neighborhood social environment and patterns of adherence to oral hypoglycemic
443 agents among patients with type 2 diabetes mellitus. Fam Community Health.
444 2015;38(2):169-179. doi:10.1097/FCH.0000000000000069
445 14. Laraia BA, Karter AJ, Warton EM, Schillinger D, Moffet HH, Adler N. Place matters:
446 neighborhood deprivation and cardiometabolic risk factors in the Diabetes Study of
447 Northern California (DISTANCE). Soc Sci Med. 2012;74(7):1082-1090.
448 doi:10.1016/j.socscimed.2011.11.036
449 15. Shaha KC, Sultana S, Saha SK, Shahidullah SM, Jyoti BK. Patient Characteristics
450 Associated With Medication Adherence to Anti-Diabetic Drugs. Mymensingh Med J.
451 2019;28(2):423-428.
452 16. McDoom MM, Cooper LA, Hsu YJ, Singh A, Perin J, Thornton RLJ. Neighborhood
453 Environment Characteristics and Control of Hypertension and Diabetes in a Primary
454 Care Patient Sample. J Gen Intern Med. 2020;35(4):1189-1198. doi:10.1007/s11606-
455 020-05671-y
456 17. Myers CA, Slack T, Broyles ST, Heymsfield SB, Church TS, Martin CK. Diabetes
457 prevalence is associated with different community factors in the diabetes belt versus the
458 rest of the United States. Obesity (Silver Spring). 2017;25(2):452-459.
459 doi:10.1002/oby.21725
460 18. Vijayaraghavan M, Jacobs EA, Seligman H, Fernandez A. The association between
461 housing instability, food insecurity, and diabetes self-efficacy in low-income adults. J
462 Health Care Poor Underserved. 2011;22(4):1279-1291. doi:10.1353/hpu.2011.0131
463 19. Truong HP, Luke AA, Hammond G, Wadhera RK, Reidhead M, Joynt Maddox KE.
464 Utilization of Social Determinants of Health ICD-10 Z-Codes Among Hospitalized

465 Patients in the United States, 2016-2017. Med Care. 2020;58(12):1037-1043.
466 doi:10.1097/MLR.0000000000001418
467 20. Walker RJ, Strom Williams J, Egede LE. Influence of Race, Ethnicity and Social
468 Determinants of Health on Diabetes Outcomes. Am J Med Sci. 2016;351(4):366-373.
469 doi:10.1016/j.amjms.2016.01.008
470 21. Hill-Briggs F, Adler NE, Berkowitz SA, et al. Social Determinants of Health and Diabetes:
471 A Scientific Review. Diabetes Care. Published online November 2, 2020.
472 doi:10.2337/dci20-0053
473 22. Haire-Joshu D, Hill-Briggs F. The Next Generation of Diabetes Translation: A Path to
474 Health Equity. Annu Rev Public Health. 2019;40:391-410. doi:10.1146/annurev-
475 publhealth-040218-044158
476 23. Davidson J, Vashisht R, Butte AJ. From Genes to Geography, from Cells to Community,
477 from Biomolecules to Behaviors: The Importance of Social Determinants of Health.
478 Biomolecules. 2022;12(10):1449. Published 2022 Oct 9. doi:10.3390/biom12101449
479 24. Vashisht R, Patel A, Dahm L, et al. Second-Line Pharmaceutical Treatments for Patients
480 with Type 2 Diabetes [published correction appears in JAMA Netw Open. 2023 Nov
481 1;6(11):e2344113]. JAMA Netw Open. 2023;6(10):e2336613. Published 2023 Oct 2.
482 doi:10.1001/jamanetworkopen.2023.36613
483 25. Hripcsak G, Ryan PB, Duke JD, et al. Characterizing treatment pathways at scale using
484 the OHDSI network. Proc Natl Acad Sci U S A. 2016;113(27):7329-7336.
485 doi:10.1073/pnas.1510502113
486 26. Peterson TA, Fontil V, Koliwad SK, Patel A, Butte AJ. Quantifying Variation in Treatment
487 Utilization for Type 2 Diabetes Across Five Major University of California Health
488 Systems. Diabetes Care. 2021;44(4):908-914. doi:10.2337/dc20-0344
489 27. Singh GK. Area deprivation and widening inequalities in US mortality, 1969-1998. Am J
490 Public Health. 2003;93(7):1137-1143. doi:10.2105/ajph.93.7.1137
491 28. Nelson CA, Butte AJ, Baranzini SE. Integrating biomedical research and electronic
492 health records to create knowledge-based biologically meaningful machine-readable
493 embeddings. Nat Commun. 2019;10(1):3045. Published 2019 Jul 10.
494 doi:10.1038/s41467-019-11069-0

495 29. Biedermann, P., Ong, R., Davydov, A. et al. Standardizing registry data to the OMOP
496 Common Data Model: experience from three pulmonary hypertension databases. BMC
497 Med Res Methodol 21, 238 (2021). https://doi.org/10.1186/s12874-021-01434-3
498 30. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of
499 Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS
500 Med. 2007;4(10):e297. doi:10.1371/journal.pmed.0040297
501 31. Fairfield KM, Black AW, Ziller EC, et al. Area Deprivation Index and Rurality in Relation
502 to Lung Cancer Prevalence and Mortality in a Rural State. JNCI Cancer Spectr.
503 2020;4(4):pkaa011. Published 2020 Mar 7. doi:10.1093/jncics/pkaa011
504 32. Ghirimoldi FM, Schmidt S, Simon RC, et al. Association of Socioeconomic Area
505 Deprivation Index with Hospital Readmissions After Colon and Rectal Surgery. J
506 Gastrointest Surg. 2021;25(3):795-808. doi:10.1007/s11605-020-04754-9
507 33. Vince RA Jr, Jiang R, Bank M, et al. Evaluation of Social Determinants of Health and
508 Prostate Cancer Outcomes Among Black and White Patients: A Systematic Review and
509 Meta-analysis [published correction appears in JAMA Netw Open. 2023 Feb
510 1;6(2):e231256]. JAMA Netw Open. 2023;6(1):e2250416. Published 2023 Jan 3.
511 doi:10.1001/jamanetworkopen.2022.50416
512 34. Hu J, Kind AJH, Nerenz D. Area Deprivation Index Predicts Readmission Risk at an
513 Urban Teaching Hospital. Am J Med Qual. 2018;33(5):493-501.
514 doi:10.1177/1062860617753063
515 35. Luningham JM, Seth G, Saini G, et al. Association of Race and Area Deprivation With
516 Breast Cancer Survival Among Black and White Women in the State of Georgia. JAMA
517 Netw Open. 2022;5(10):e2238183. Published 2022 Oct 3.
518 doi:10.1001/jamanetworkopen.2022.38183
519 36. Knighton AJ, Savitz L, Belnap T, Stephenson B, VanDerslice J. Introduction of an Area
520 Deprivation Index Measuring Patient Socioeconomic Status in an Integrated Health
521 System: Implications for Population Health. EGEMS (Wash DC). 2016;4(3):1238.
522 Published 2016 Aug 11. doi:10.13063/2327-9214.1238
523 37. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible -
524 The Neighborhood Atlas. N Engl J Med. 2018;378(26):2456-2458.
525 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

562 **Figure Legends:**



563

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

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

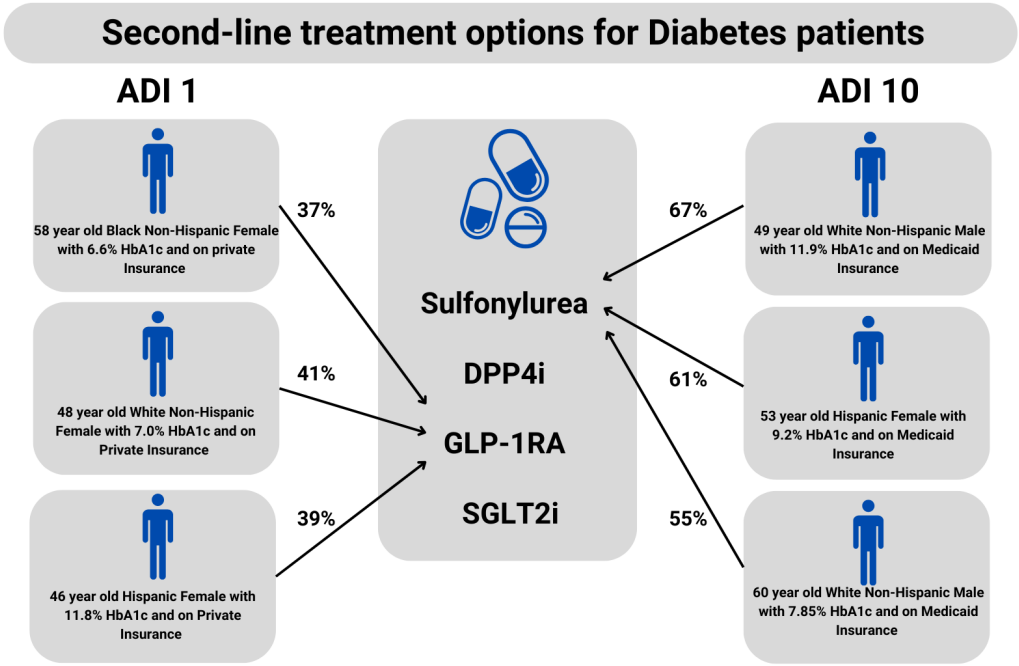


Figure 2: Second-line treatment options for Type 2 diabetes patients

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.

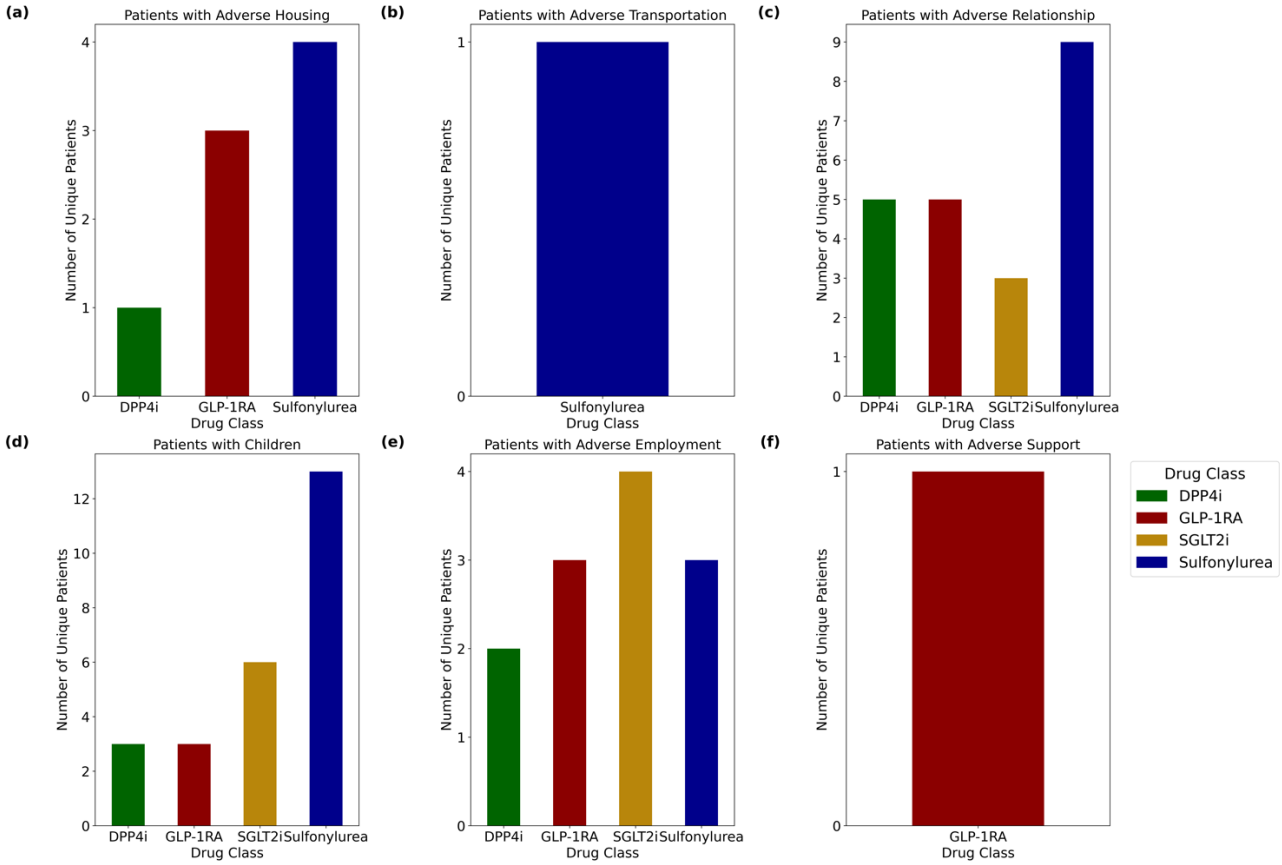


Figure 3: GPT-4 extraction and identification of Individual-Level SDOH Factors

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.

Table 1: Baseline Characteristics of Type 2 Diabetes Patients with Second-Line Treatments

Variable	No. (%) of patients					
	DPP4i (N=2918)	**GLP-1RA** (N=2736)	**SGLT2i** (N=2704)	**Sulfonylurea** (N=6732)	**Total** (N=15090)	**P-Value**
Age						
Mean (SD)	62.7 (13.4)	55.5 (13.7)	60.3 (13.4)	59.0 (12.9)	59.3 (13.4)	< 0.001
HbA1c						
Mean (SD)	8.59 (1.85)	8.83 (2.06)	8.62 (1.88)	9.19 (2.02)	8.91 (1.99)	< 0.001
Sex						
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%)	
Race						
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%)	
Native Hawaiian or Other Pacific Islander	29 (1.0%)	20 (0.7%)	29 (1.1%)	81 (1.2%)	159 (1.1%)	
Other Race	560 (19.2%)	502 (18.3%)	480 (17.8%)	1535 (22.8%)	3077 (20.4%)	
Unknown	247 (8.5%)	332 (12.1%)	300 (11.1%)	430 (6.4%)	1309 (8.7%)	
White	1341 (46.0%)	1307 (47.8%)	1192 (44.1%)	3197 (47.5%)	7037 (46.6%)	
Ethnicity						
Hispanic or Latino	708 (24.3%)	794 (29.0%)	742 (27.4%)	2389 (35.5%)	4633 (30.7%)	< 0.001
Not Hispanic or Latino	1977 (67.8%)	1697 (62.0%)	1740 (64.3%)	3946 (58.6%)	9360 (62.0%)	
Unknown	233 (8.0%)	245 (9.0%)	222 (8.2%)	397 (5.9%)	1097 (7.3%)	
Area Deprivation Index (ADI)						
1	368 (12.6%)	308 (11.3%)	337 (12.5%)	519 (7.7%)	1532 (10.2%)	< 0.001
2	293 (10.0%)	262 (9.6%)	266 (9.8%)	551 (8.2%)	1372 (9.1%)	
3	323 (11.1%)	302 (11.0%)	298 (11.0%)	529 (7.9%)	1452 (9.6%)	
4	353 (12.1%)	360 (13.2%)	315 (11.6%)	771 (11.5%)	1799 (11.9%)	
5	317 (10.9%)	328 (12.0%)	336 (12.4%)	773 (11.5%)	1754 (11.6%)	
6	293 (10.0%)	336 (12.3%)	326 (12.1%)	947 (14.1%)	1902 (12.6%)	
7	295 (10.1%)	283 (10.3%)	250 (9.2%)	772 (11.5%)	1600 (10.6%)	

8	236 (8.1%)	245 (9.0%)	213 (7.9%)	676 (10.0%)	1370 (9.1%)	
9	231 (7.9%)	176 (6.4%)	208 (7.7%)	659 (9.8%)	1274 (8.4%)	
10	209 (7.2%)	136 (5.0%)	155 (5.7%)	535 (7.9%)	1035 (6.9%)	
Primary Insurance						
Medicaid	700 (24.0%)	607 (22.2%)	720 (26.6%)	2515 (37.4%)	4542 (30.1%)	< 0.001
Medicare	880 (30.2%)	495 (18.1%)	634 (23.4%)	1762 (26.2%)	3771 (25.0%)	
Private	1273 (43.6%)	1580 (57.7%)	1301 (48.1%)	2349 (34.9%)	6503 (43.1%)	
Veterans Affairs	65 (2.2%)	54 (2.0%)	49 (1.8%)	106 (1.6%)	274 (1.8%)	

596

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

600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624

625 **Table 2: Results of the Multinomial Regression Model**

Variable	DPP4i		GLP-1RA		SGLT2i	
	OR (95% CI)	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**
Area Deprivation Index (ADI)	0.96 (0.95-0.98)	< 0.001	0.94 (0.92-0.96)	< 0.001	0.95 (0.93-0.97)	< 0.001
HbA1c	0.90 (0.88-0.92)	< 0.001	0.90 (0.87-0.92)	< 0.001	0.88 (0.85-0.90)	< 0.001
Age	1.02 (1.01-1.02)	< 0.001	0.98 (0.97-0.98)	< 0.001	1.01 (1.00-1.01)	0.001
Sex						
Female	1.10 (1.00-1.20)	0.043	1.44 (1.31-1.58)	< 0.001	0.79 (0.72-0.86)	< 0.001
Race						
American Indian or Alaska Native	1.33 (0.74-2.41)	0.345	1.52 (0.87-2.64)	0.138	0.85 (0.42-1.74)	0.662
Asian	1.19 (1.05-1.35)	0.008	0.62 (0.53-0.72)	< 0.001	1.25 (1.09-1.43)	0.001
Black or African American	0.81 (0.67-0.98)	0.031	1.03 (0.86-1.22)	0.759	1.06 (0.88-1.27)	0.533
Native Hawaiian or Other Pacific Islander	0.86 (0.56-1.33)	0.507	0.53 (0.32-0.87)	0.012	1.01 (0.65-1.56)	0.969
Other Race	1.21 (1.06-1.38)	0.004	1.04 (0.91-1.19)	0.552	1.04 (0.90-1.19)	0.604
Unknown	1.35 (1.10-1.65)	0.004	1.54 (1.27-1.86)	< 0.001	1.81 (1.49-2.20)	< 0.001
Ethnicity						
Hispanic or Latino	0.73 (0.64-0.83)	< 0.001	0.78 (0.69-0.88)	< 0.001	0.90 (0.79-1.02)	0.104
Unknown	1.02 (0.83-1.25)	0.859	0.92 (0.75-1.14)	0.459	0.90 (0.73-1.12)	0.346
Primary Insurance						
Medicare	1.19 (1.05-1.36)	0.007	1.30 (1.12-1.51)	< 0.001	0.93 (0.81-1.07)	0.297
Private Health Insurance	1.67 (1.49-1.87)	< 0.001	2.39 (2.13-2.69)	< 0.001	1.56 (1.40-1.75)	< 0.001
Veterans Affairs	1.62 (1.17-2.24)	0.004	2.26 (1.60-3.21)	< 0.001	1.29 (0.91-1.84)	0.158

626

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

632