Longitudinal Analysis of Diabetes Progression in Medicare Patients Using Claims Data

Ben Abijah[1]

[1]*Department of Biostatistics and Epidemiology, University of Massachusetts Amherst, MA*

Abstract

Background: Diabetes is a chronic disease that often progresses from a mild stage to a more complicated one. Understanding the various driving factors of this progression is an important step towards policy implementation to mitigate its effects. In this project, we employ a data-driven approach to examine the longitudinal paths of diabetes progression among Medicare members.

Methods: We analyze synthetic claims data obtained from the Centers for Medicare and Medicaid Services (CMS) from 2022 through 2025 for a cohort of 3,172 prediabetes patients diagnosed of prediabetes in 2022 and followed up for their progression to diabetes mellitus with(out) complications. Transitions between specific states is modeled using continuous-time multi-state models incorporating demographic factors and other comorbidities. Furthermore, we conduct subgroup analyses to delineate differences in progression by sex and specific age categories.

Results: Most prediabetic patients develop diabetes with complications without being diagnosed of diabetes without complications. Obesity, hypertension, and heart disease/failure are some of the factors that significantly impact the risk of diabetes progression. Additionally, males and older adults have higher chances of progression than females and younger adults. Higher risks are observed for obese females and obese older adults. On the other hand, higher risks are associated with males living with hypertension, heart disease/failure, alcohol disorders, and depressive disorders.

Conclusion: The insights learned will be useful in improving healthcare delivery.

Keywords: *Medicare, diabetes, disease progression, survival analysis, multistate models*.

Introduction

Diabetes is a chronic disease that often progresses from a mild stage (i.e., prediabetes) to diabetes mellitus with/without complications. If unattended to, it could eventually lead to other chronic diseases like stroke or kidney failure, or even death. In 2021 alone, diabetes was noted as the underlying cause of death on 399,401 U.S. death certificates [1]. By same year, the American Diabetes Association estimated that about 97.6 million Americans aged 18 or above had prediabetes.

Consequently, understanding the various driving factors behind diabetes progression is an important step towards policy implementation to mitigate its effects. Specifically, given any population, if we can efficiently estimate the risks involved at every stage of the disease – as well as identify the various factors that drive these risks – then health professionals can use medical interventions targeted at preventing its progression or, possibly, delaying the rate of progression. On a larger scale, policymakers can formulate public health policies aimed at protecting the population from this health menace, which cost the country about $412.9 billion in 2022 alone – including $306.6 billion and $106.3 billion in direct and indirect medical costs, respectively [5].

In this project, we employ a data-driven approach to examine the longitudinal paths of diabetes progression among Medicare beneficiaries. The foremost goal is to explore some efficient ways of handling electronic health records – from data ingestion all the way through analyses. Using synthetic healthcare claims data obtained from the Centers for Medicare & Medicaid Services (CMS), we track beneficiaries' progression from being diagnosed with *prediabetes* to *diabetes mellitus without complications*, or *diabetes mellitus with complications*.

To do this, we use the assistance of AI tools for the ingestion of our data and then proceed with statistical modeling to help us understand the dynamics involved in diabetes disease progression. Briefly, we estimate beneficiaries' progression-free survival from either prediabetes or diabetes mellitus without complications, and their time-dependent transition probabilities given these two early stages. We also assess the contribution of demographic features and comorbidities across the progression. Finally, we perform subgroup analyses to delineate differences in progression by sex and age categories.

Methods

Study Design

This is a longitudinal study involving Medicare beneficiaries followed from their time of prediabetes to when higher stages of diabetes were diagnosed. It uses synthetic Medicare claims data obtained from the CMS [2], and included only diagnoses information obtained from the inpatient, outpatient, and professional records. Figure 1 below illustrates the main analytic processes employed for this project.

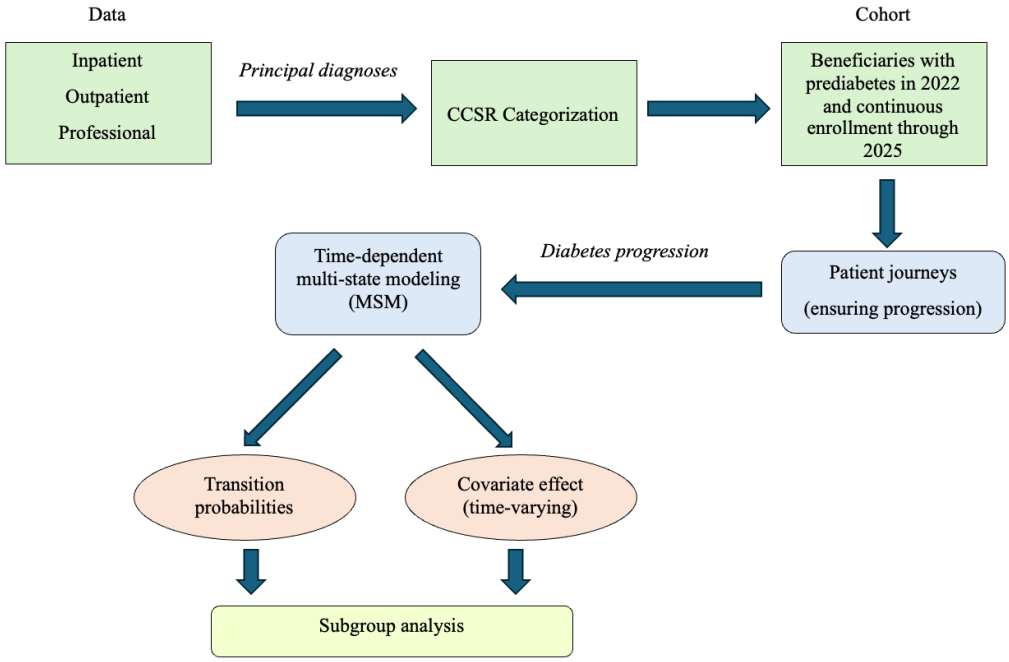

Figure 1. Analytic framework for modelling diabetes progression using patient journey data. Abbreviation: CCSR, Clinical Classifications Software Refined.

Patients' Diagnoses

Briefly, diagnoses were defined as those noted at the initial hospital admission, i.e., principal diagnosis. For all three claim sources (inpatient, outpatient, professional), we used only principal diagnoses because it largely describes the underlying cause of a patient's admission. These diagnoses were then re-categorized using groupings based on the Clinical Classifications Software Refined (CCSR) [3]. Here, we note that beneficiaries had multiple diagnoses of the same condition across years. Consequently, for each patient and diagnosis, we only considered the first time the condition was diagnosed.

Study Cohort and Diabetes Progression

We only consider beneficiaries diagnosed with prediabetes in 2022 who had continuous enrollment through 2025 ($n = 3{,}172$). This includes beneficiaries who "progressed" from a lower stage of diabetes to higher ones, with no regression at any point during the follow-up period. For example, we exclude beneficiaries with observed states, say, $[1,2,1]$ because they regressed from state 2 to state 1 – instead of remaining in state 2, or even advancing to state 3. The assumption here is that once a patient is diagnosed with a more advanced stage of the diseases on the diabetes spectrum, they do not get diagnosed with an earlier stage later, thus ensuring progression. Also, we exclude patients with two or more distinct diabetes stages diagnosed at the same time, because we cannot explicitly determine which diagnosis occurred before the other. This assumes that, at any time point, a patient has one (and only one) of the three diabetes stages and not have multiple stages at once. We note that cohort ingestion was done in *Google Collab*, leveraging the assistance of some of its AI features.

Covariates

Beneficiaries' demographic data was obtained from their membership/enrollment data. This includes *age* at indexing year (i.e., time of prediabetes diagnosis, 2022), *race*, and biological *sex* (male or female). Age was grouped as child (18 years or below), adult below 65 years, and adults aged 65 or above. However, the final analytic cohort did not include any child. There were six different races: White, Hispanic, Black, Asian, North American native, and others. Here, we note that all demographic traits were taken as time-independent, as they do not change over time. Although age changes with time, we only consider beneficiaries' age at baseline (index time).

Seven comorbidities that are plausible risk factors of diabetes were used in this study: *Obesity*, *Hypertension* (Essential hypertension, Hypertension with complications and secondary hypertension), *Heart disease or failure* (Coronary atherosclerosis and other heart disease, Heart failure), *Socioeconomic or lifestyle factors* (Socioeconomic/psychosocial factors, Lifestyle/life management factors), *Alcohol-related disorders*, *Depressive or anxiety disorders* (Depressive disorders, Anxiety and fear-related disorders), and *Sleep wake disorders*. These traits were all coded 1 or 0, indicating presence or absence of the corresponding comorbidity.

Furthermore, comorbidities were considered as time-dependent, in the sense that they changed with time. For example, a beneficiary may not have "hypertension" at the time of prediabetes diagnosis, but may be diagnosed with it, say, 5 months later even when they are still with prediabetes or had advanced to a higher stage. Also, we limit comorbidities to the same indexing time to align all beneficiaries at a comparable starting point (i.e., time of first prediabetes diagnosis in 2022). Finally, comorbidities were updated across the follow-up period as and when they were diagnosed for each patient.

Time to Progression

We created patient journeys synthetically by randomly assigning months January through December to each unique diagnosis. Since this is a longitudinal modeling of diabetes progression, the time of diagnosis of all other conditions – including comorbidities – was calculated relative to the time patients were first diagnosed with prediabetes. Essentially, we track the duration (in months) from prediabetes diagnosis to when the two higher stages of diabetes or comorbidities were diagnosed. Censoring in this study means that the patient did not progress from prediabetes to a specific higher diabetes stage. Given continuous enrollment, censoring is only attributable to end of follow-up period.

Statistical Analyses

First, we use the Kaplan-Meier non-parametric survival estimator to estimate beneficiaries' progression-free survival from either prediabetes or diabetes without complications. Then, motivated by Li et al., 2024 [4] and Siriwardhana et al., 2018 [6], we estimate transition probabilities between three states: (1) prediabetes, (2) diabetes without complications, and (3) diabetes with complications. Using continuous time multi-state models to track diabetes trajectory and risks, we assume three plausible transitions: 1→2, 1→3, or 2→3. However, patients could remain in a current state at given times because not all patients get diagnosed with higher stages.

Our models incorporated demographic factors and comorbidities described earlier. Finally, we perform subgroup analyses to understand differences for specific sex and age categories. Statistical modeling was done in `R` version 4.2.3 (2023-03-15) with relevant packages (e.g., `survival`, `msm`). It is worth mentioning that, although this project uses synthetic data, the methods herein may be transferable to real applications.

Results and Discussion

Preliminary Results

Of the 3,172 beneficiaries in the analytic cohort, 2,453 (77.3%) had prediabetes that never progressed whiles 656 (20.7%) progressed to either diabetes without complications, or diabetes with complications. Patients diagnosed with all three stages of diabetes were 63 (2.0%) [See Appendix, Figure A1 for a distribution of these diagnoses and other comorbidities by year]. The average age at the time of prediabetes diagnoses was 65.6 years (SD=15.6). In terms of groupings, there were 62% adults aged 65 years or above and 38% aged below 65 years. Males made up 48.2% compared to 51.8% females. There were relatively more whites than there were other races. Specifically, the cohort comprised 63.8% Whites, 18% Hispanics, 11.1% Blacks, among others [See Appendix, Figure A2].

Progression-free Survival

We estimated patients' progression-free survival probabilities; the probability that a patient has yet to be diagnosed with a higher stage of diabetes. For example, for a beneficiary with prediabetes, this estimates the probability that at time, say, 6 months, they still had prediabetes and had not yet progressed to any higher stage. We observe in Figure 2 that the survival curve A decays slower than survival curve B. By the fifth month, about 85% of patients were still prediabetic. On the other hand, by the fifth month (after being diagnosed with diabetes without complications), only about 70% of patients still had the condition. This implies that it takes longer on average, for a prediabetic patient to be diagnosed with either diabetes without complications or diabetes with complications, whiles it does not take long for patients with diabetes without complications to be diagnosed with diabetes with complications. In other words, the chances of advancing from state 2 to state 3 is relatively higher than it is from state 1 to either states 2 or 3.

Transition probabilities

Here, we estimate transition probabilities; the probability of being in a higher state (or remaining in a current state) at given times, conditioned on the patient being in that current state. For example, the probability that a patient will be diagnosed with diabetes with complications at time, say, 6

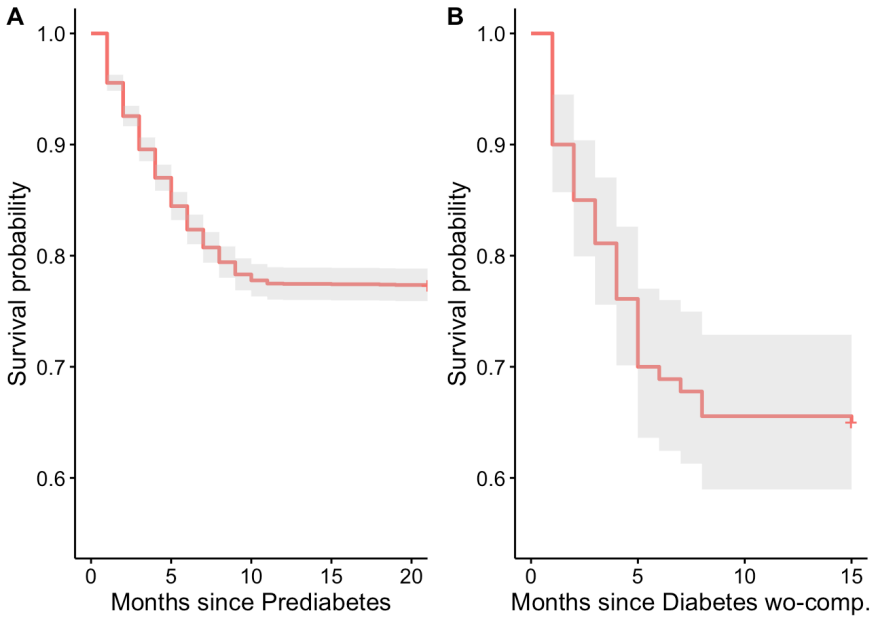

Figure 2. Survival probability curves for diabetes progression. (A) Progression from prediabetes to either diabetes without complications, or diabetes with complications. **(B)** Progression from diabetes without complications to diabetes with complications.

months, given that they had prediabetes. We do this for the first 12 months as well as later months with increments of three. Figure 3 below illustrates transition probability curves conditioned on

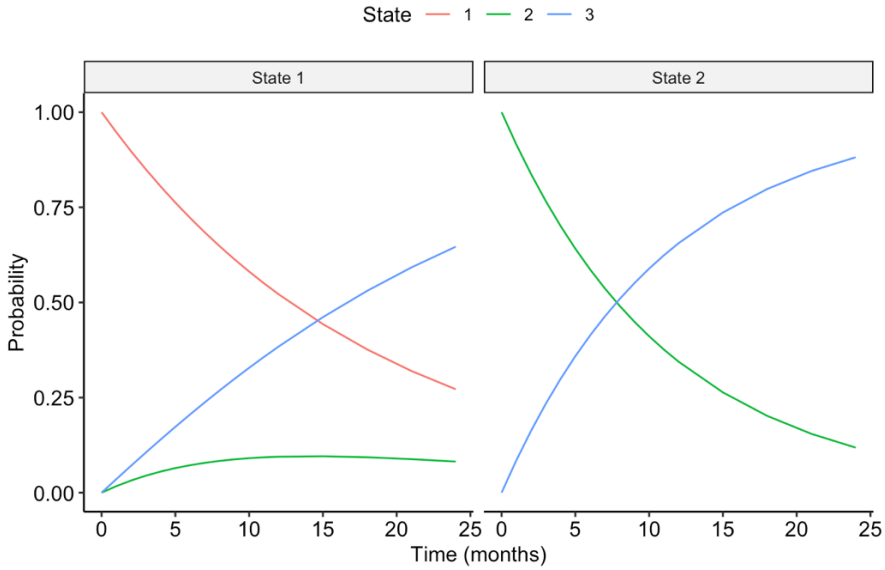

Figure 3. Transition probabilities over time. Left: Probability of transitioning to any state given state 1. **Right**: Probability of transitioning to states 2 or 3 given state 2. Abbreviations: State 1, Prediabetes; State 2, Diabetes mellitus without complications; State 3, Diabetes mellitus with complications.

patients already being in state 1 (left) or state 2 (right). Probabilities were computed while constraining all other covariates as unobserved (i.e., baseline probabilities). Notably, the probability of a patient remaining in a current state decrease over the follow-up period (red curve, left; green curve, right). This is not farfetched, as one would expect that the longer a patient stays with any disease, the more their condition worsens, thus developing into other higher complications. Furthermore, patients are more likely to transition to state 3 given that they were in state 2 than if they were in state 1 (blue curves) – as also observed for progression-free survival. Furthermore, from state 1, the chances of transitioning into state 2 is very slim (green) as compared to state 3 (blue), implying that most patients transition straight to state 3 without passing through the intermediate state 2.

Effects of Risk Factors on Progression

We present some results on diabetes progression risks associated with demographics and comorbidities [Table 1]. The risks of an obese, prediabetic patient progressing on to diabetes with complications are increased as compared to a non-obese, prediabetic patient (HR=1.66). Similarly, prediabetic patients diagnosed with socioeconomic and lifestyle disorders increase their risks of advancing on to the two higher stages of diabetes (HR: 1.37 and 2.24, respectively). Notably, hypertensive patients substantially increase their risks of progressing from any lower stage of

Table 1. Estimated covariate effect for various transitions. Some significant factors affecting progression are highlighted. Confidence intervals are given in braces.

Covariate	1-2*	1-3*	2-3*
Obesity	1.288 (0.949, 1.747)	**1.663 (1.379, 2.006)**	1.412 (0.827, 2.41)
Socioeconomic or lifestyle factors	**1.366 (1.006, 1.854)**	**2.239 (1.835, 2.733)**	1.396 (0.805, 2.422)
Hypertension	**2 (1.449, 2.761)**	**4.808 (3.968, 5.825)**	**1.962 (1.136, 3.388)**
Heart disease or failure	**1.993 (1.435, 2.769)**	**1.497 (1.213, 1.848)**	1.588 (0.918, 2.747)
Sleep wake disorders	0.525 (0.273, 1.008)	0.957 (0.717, 1.279)	0.507 (0.206, 1.247)
Alcohol-related disorders	1.325 (0.502, 3.495)	1.382 (0.853, 2.238)	2.171 (0.423, 11.134)
Depressive or anxiety disorders	1.71 (0.765, 3.824)	**2.185 (1.469, 3.249)**	1.849 (0.58, 5.894)
Sex	1.015 (0.749, 1.375)	0.797 (0.658, 0.965)	0.802 (0.471, 1.364)
Race	1.002 (0.91, 1.102)	1.02 (0.963, 1.082)	1.078 (0.916, 1.269)
Age (years)	**1.014 (1.002, 1.026)**	1.003 (0.996, 1.011)	**1.044 (1.02, 1.068)**

* 1 = Prediabetes, 2 = Diabetes mellitus without complication, 3 = Diabetes mellitus with complication.

diabetes to a higher stage, with an even relatively high risk for progression from prediabetes to diabetes with complications (HR: 2.00, 4.81, and 1.96, respectively).

Prediabetic patients with heart disease/failure also have increased risks of progressing to the two higher stages (HR: 1.99 and 1.50, respectively). Depression and anxiety disorders significantly increase the risks of progression from prediabetes to diabetes with complications (HR=2.19). Age of patient also contributes to increased risks for 1→2 and 2→3 progressions. Interestingly, sex and race did not significantly affect diabetes progression, likewise alcohol-related and depressive disorders.

Subgroup Analyses

Now, we present results on age- and sex- stratified subgroup analyses to assess apparent differences in risk of comorbidities and demographics [See Appendix]. In general, males tend to have higher transition probabilities compared to females, implying that they are more likely to progress from lower stages of diabetes to higher ones (blue curves, Figure A3). Similarly, older adults (60 years or above) have higher transition probabilities over the counterparts aged below 60 years (blue curves, Figure A4).

Table A1 presents hazard ratios for the male and female subgroups. We observe that, in both groups, obese patients have increased risks of progression – although relatively higher in females (e.g., HR=1.53, males vs. HR=1.85, females). Unlike females, hypertensive males and those with heart disease/failure increase their risks of progressing from diabetes without complications to diabetes with complications (HR: 2.76 and 3.27, respectively). We also observe that men with alcohol-related disorders are at higher risks of diabetes progression, likewise those with depressive disorders. In terms of age, females face high risks of progression as they age.

The results in Table A2 show notable differences for the age groups. Specifically, older obese adults and those with depressive disorders have slightly higher risks of developing diabetes with no complications from prediabetes than do younger adults (HR: 1.60 and 1.85, adults <65 years; 1.75 and 2.45, adults 60+ years). On the other hand, younger adults with heart disease/failure are at more risk of progression from no complications to complications over older adults (HR=7.06).

Conclusion

This project was a learning exercise on the handling of electronic health data. We sought to analyze longitudinal data on Medicare claims to evaluate risks progression from prediabetes all the way through diabetes with complications. The insights learned will be useful in improving healthcare delivery.

Among others, we find that, on average, there is delayed progression from prediabetes to higher stages of diabetes whereas it does not take long for patients living with diabetes without complications to be diagnosed with its complications. Also, most prediabetic patients tend to develop diabetes with complications without passing through the intermediate diabetes without complications. Obesity, hypertension, and heart disease/failure are some of the factors that significantly impact the risk of progression of diabetes in patients.

The subgroup analyses revealed that males and older adults generally have higher chances of progression than as do females and younger adults. Additionally, higher risk was observed for obese females and obese older adults. On the other hand, higher risks are associated with males living with hypertension, heart disease/failure, alcohol disorders, and depressive disorders.

This project is limited by the fact that the data used were synthetic. Hence, there is the need to incorporate real Medicare claims data to help validate the generalizability of our results. Future work could also incorporate other covariates like use of antidiabetics, as these medications play a key role in delaying the onset/progression of diabetes.

References

1. American Diabetes Association. *Statistics about diabetes*. [Accessed 3 May 2025]. Available from: https://diabetes.org/about-diabetes/statistics/about-diabetes

2. Centers for Medicare & Medicaid Services (CMS). *Synthetic Medicare Enrollment, Fee-for-Service Claims, and Prescription Drug Event*. [Accessed 15 February 2025]. Available from: https://data.cms.gov/collection/synthetic-medicare-enrollment-fee-for-service-claims-and-prescription-drug-event

3. Healthcare Cost and Utilization Project (HCUP). Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses. [Assessed 15 February 2025]. https://hcup-us.ahrq.gov/toolssoftware/ccsr/dxccsr.jsp#download

4. Li RC, Ding S, Ndura K, Patel V, Jurkovitz C. *Building a multistate model from electronic health records data for modeling long-term diabetes complications*. J Clin Transl Sci. 2024 Sep 23;8(1):e133. doi: 10.1017/cts.2024.583. PMID: 39345707; PMCID: PMC11428072.

5. Parker ED, Lin J, Mahoney T, Ume N, Yang G, Gabbay RA, ElSayed NA, Bannuru RR. *Economic Costs of Diabetes in the U.S. in 2022*. Diabetes Care. 2024 Jan 1;47(1):26-43. doi: 10.2337/dci23-0085. PMID: 37909353.

6. Siriwardhana C, Lim E, Davis J, Chen JJ. *Progression of diabetes, ischemic heart disease, and chronic kidney disease in a three chronic conditions multistate model*. BMC Public Health. 2018 Jun 18;18(1):752. doi: 10.1186/s12889-018-5688-y. PMID: 29914451; PMCID: PMC6006736.

Appendix

Data Availability

Data used for this project are available at the official data website of the Centers for Medicare & Medicaid Services (CMS): https://data.cms.gov/collection/synthetic-medicare-enrollment-fee-for-service-claims-and-prescription-drug-event

Code Availability

Notebooks used for data processing and statistical analyses can her found on GitHub repository: https://github.com/benabijah/EHR

Figures

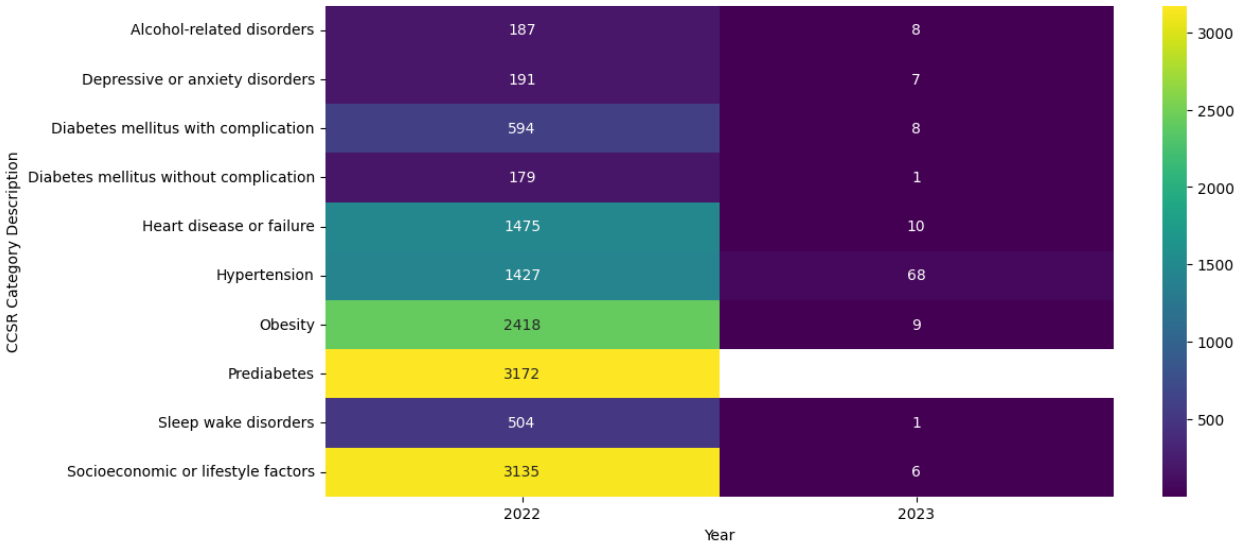

Figure A1. Distribution of CCSR Categories by year. Abbreviation: CCSR, Clinical Classifications Software Refined.

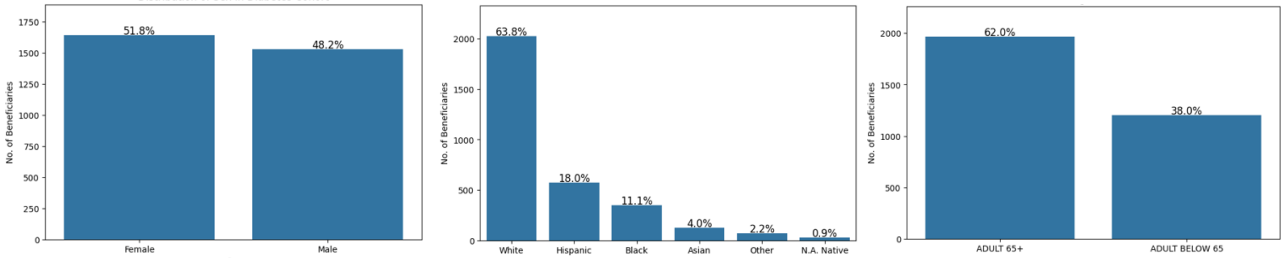

Figure A2. Demographic distribution of diabetes cohort. **Left**: Distribution by sex. **Middle**: Distribution by race. **Right**: Distribution by age categories.

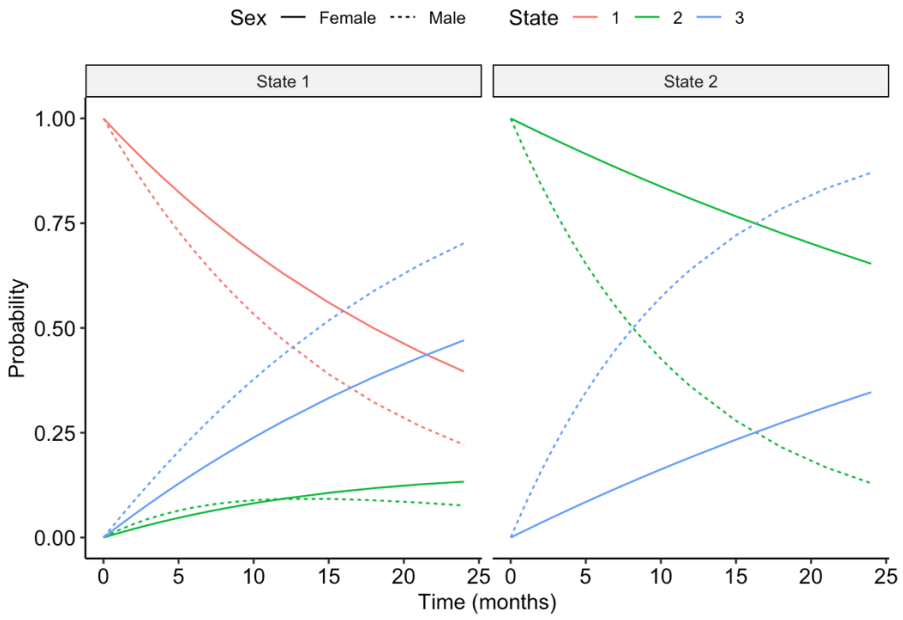

Figure A3. Sex-stratified transition probabilities across time. *Left*: Probability of transitioning to any state given state 1. *Right*: Probability of transitioning to states 2 or 3 given state 2. Abbreviations: State 1, Prediabetes; State 2, Diabetes mellitus without complications; State 3, Diabetes mellitus with complications.

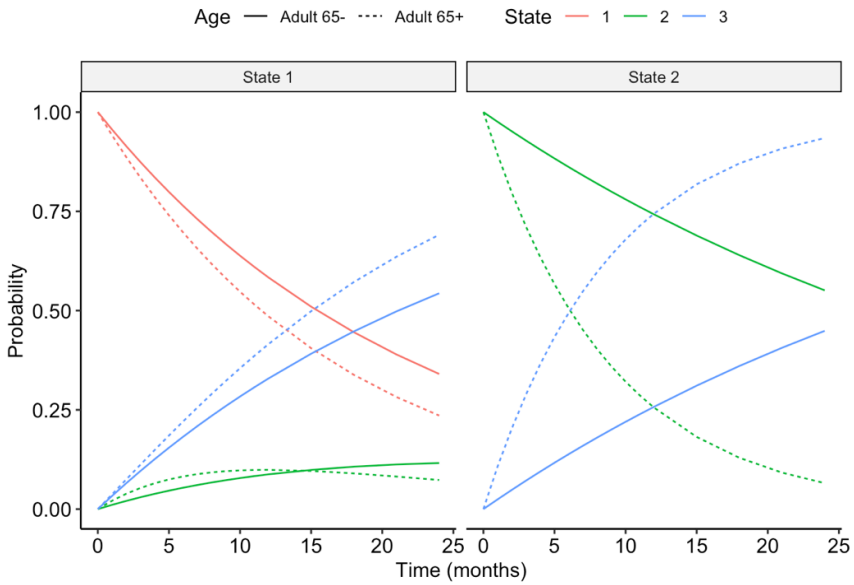

Figure A4. Age-stratified transition probabilities across time. *Left*: Probability of transitioning to any state given state 1. *Right*: Probability of transitioning to states 2 or 3 given state 2. Abbreviations: State 1, Prediabetes; State 2, Diabetes mellitus without complications; State 3, Diabetes mellitus with complications.

Tables

Table A1. Sex-stratified estimated covariate effect for various transitions. Some significant factors affecting progression are highlighted. Confidence intervals are given in braces.

	Males			Females		
Covariate	**1-2**	**1-3**	**2-3**	**1-2**	**1-3**	**2-3**
Obesity	1.216 (0.794, 1.862)	**1.527 (1.193, 1.954)**	0.931 (0.46, 1.883)	1.432 (0.93, 2.203)	**1.845 (1.376, 2.473)**	**2.876 (1.185, 6.979)**
Socioeconomic or lifestyle factors	**1.618 (1.049, 2.495)**	**2.203 (1.698, 2.859)**	1.459 (0.695, 3.065)	1.274 (0.828, 1.962)	**2.19 (1.606, 2.988)**	1.869 (0.811, 4.308)
Hypertension	**2.051 (1.299, 3.24)**	**4.225 (3.284, 5.436)**	**2.762 (1.29, 5.914)**	**2.142 (1.356, 3.384)**	**5.84 (4.328, 7.88)**	1.787 (0.783, 4.079)
Heart disease or failure	**2.372 (1.5, 3.75)**	**1.46 (1.116, 1.908)**	**3.268 (1.466, 7.284)**	**1.827 (1.147, 2.911)**	**1.443 (1.01, 2.062)**	1.166 (0.525, 2.593)
Sleep wake disorders	0.587 (0.267, 1.289)	1.065 (0.75, 1.513)	0.719 (0.288, 1.8)	0.41 (0.128, 1.317)	0.758 (0.443, 1.298)	0 (0, Inf)
Alcohol-related disorders	**3.054 (1.157, 8.066)**	1.056 (0.411, 2.715)	**4.799 (1.124, 20.499)**	0 (0, Inf)	1.505 (0.736, 3.081)	0 (0, Inf)
Depressive or anxiety disorders	2.328 (0.834, 6.497)	**2.794 (1.657, 4.712)**	1.758 (0.534, 5.781)	1.157 (0.312, 4.284)	1.733 (0.946, 3.175)	0.946 (0.079, 11.373)
Race	1.077 (0.947, 1.226)	0.983 (0.906, 1.066)	1.084 (0.884, 1.329)	0.941 (0.815, 1.086)	1.061 (0.973, 1.157)	1.259 (0.966, 1.641)
Age (years)	1.005 (0.987, 1.023)	1.005 (0.995, 1.015)	1.025 (0.992, 1.059)	**1.02 (1.005, 1.036)**	1.003 (0.992, 1.014)	**1.069 (1.035, 1.103)**

1 = Prediabetes, 2 = Diabetes mellitus without complication, 3 = Diabetes mellitus with complication.

Table A2. Age-stratified estimated covariate effect for various transitions. Some significant factors affecting progression are highlighted. Confidence intervals are given in braces.

	Adults <65			Adults 65+		
Covariate	**1-2**	**1-3**	**2-3**	**1-2**	**1-3**	**2-3**
Obesity	1.399 (0.752, 2.6)	**1.599 (1.18, 2.168)**	1.929 (0.343, 10.855)	1.207 (0.845, 1.723)	**1.752 (1.378, 2.229)**	1.175 (0.671, 2.058)
Socioeconomic or lifestyle factors	**2.065 (1.097, 3.889)**	**2.541 (1.851, 3.489)**	0.514 (0.089, 2.979)	1.234 (0.866, 1.758)	**2.073 (1.612, 2.666)**	1.639 (0.884, 3.04)
Hypertension	1.693 (0.848, 3.378)	**5.177 (3.818, 7.02)**	**5.36 (1.239, 23.194)**	**2.213 (1.528, 3.206)**	**4.489 (3.525, 5.717)**	**2.347 (1.291, 4.268)**
Heart disease or failure	1.536 (0.505, 4.675)	**1.71 (1.061, 2.757)**	**7.056 (1.482, 33.594)**	**2.087 (1.474, 2.955)**	**1.639 (1.294, 2.076)**	1.733 (0.969, 3.1)
Sleep wake disorders	0.529 (0.205, 1.365)	0.723 (0.48, 1.089)	0.566 (0.095, 3.382)	0.555 (0.222, 1.387)	1.162 (0.771, 1.752)	0.397 (0.119, 1.327)
Alcohol-related disorders	1.314 (0.315, 5.486)	1.364 (0.664, 2.801)	0 (0, Inf)	1.445 (0.358, 5.842)	1.292 (0.616, 2.712)	3.001 (0.513, 17.562)
Depressive or anxiety disorders	2.079 (0.634, 6.823)	**1.848 (1.009, 3.386)**	1.275 (0.096, 17.013)	1.392 (0.474, 4.087)	**2.448 (1.482, 4.043)**	2.226 (0.623, 7.954)
Sex	0.642 (0.346, 1.194)	0.82 (0.603, 1.115)	0.273 (0.035, 2.156)	1.129 (0.794, 1.605)	0.798 (0.624, 1.019)	0.806 (0.447, 1.455)
Race	1.068 (0.872, 1.309)	1.039 (0.941, 1.148)	1.526 (0.982, 2.372)	0.983 (0.88, 1.098)	1.016 (0.945, 1.092)	0.981 (0.806, 1.193)

1 = Prediabetes, 2 = Diabetes mellitus without complication, 3 = Diabetes mellitus with complication.