Analyzing Cardiovascular and Kidney Disease Risks Among Diverse Populations with Type 2 Diabetes Using Compartmental Modeling and the NIH All of Us Database

Jackson Deal, Matthew Jensen, Adrián Martínez Félix, Tamantha Pizarro

Dr. Susan Holechek, Dr. Jordy Cevallos Chavez

Arizona State University

Quantitative Research in the Life and Social Sciences Program

Background

Diabetes Definition

- Diabetes Mellitus is a chronic metabolic disease
- The body doesn't produce enough or cannot use insulin effectively
- This leads to high blood sugar levels

High blood sugar can lead to problems including :

- Neuropathy
- Retinopathy
- Fatigue
- Kidney Disease
- Cardiovascular Disease
- Ketoacidosis (Only Type 1)
- Death

Davis et al. (2019)

Differences in Type 1 vs Type 2 Diabetes

Table 1: Statistical comparison of Type I and Type II Diabetes detailing physical traits, risk factors, and molecular basis.[1-3]

	TYPE I DIABETES	TYPE II DIABETES
ONSET AGE	Childhood/adolescence	After 40 years of age
PHENOTYPE	Often thin or normal weight	Often obese
	Prone to ketoacidosis	No ketoacidosis
INSULIN LEVELS	Absolute insulin deficiency	Relative insulin deficiency and/or resistance
INSULIN ADMINISTRATION	Required for survival	Not required for survival
PANCREAS	Damaged by autoimmune attack	Not damaged
RISK FACTORS	Increased prevalence in relatives	Increased prevalence in relatives
	Autoimmune, genetic	Obesity, physical inactivity, ethnicity, impaired glucose tolerance
PREVALENCE	5-10% of cases	90-95% of cases
TREATMENT	Insulin injections	(1) healthy diet and increased exercise; (2) hypoglycemic tablets; (3) insulin injections

Davis et al. (2019)

Diabetes Stats in the USA

Type 2 Diabetes

More than 37 million

Prevalence

1 in 10 adults

Prediabetes risk

1 in 3 adults have prediabetes

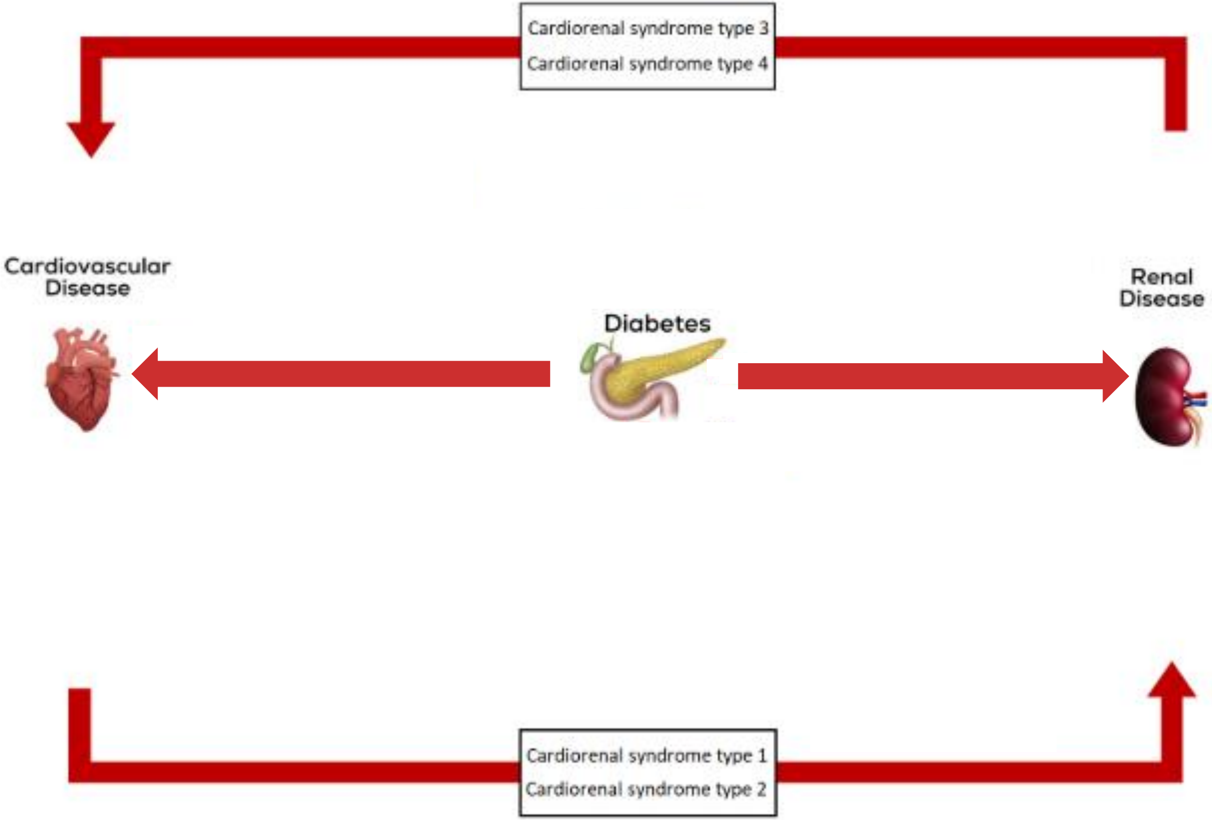
11.6%

Prevalence Underrepresented groups more affected

Centers for Disease Control and Prevention (2025)

Cardiovascular and Kidney Disease

Interaction with cardiovascular complications



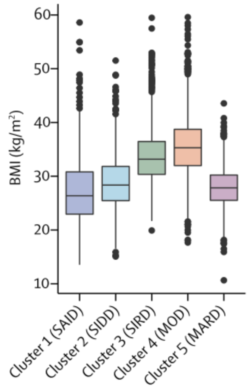
Adapted from Usman, M. et al. (2021)

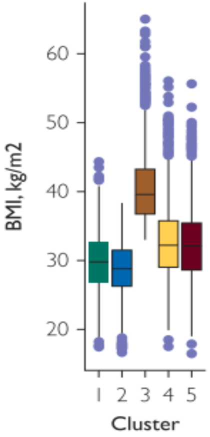
Type 2 Diabetes Subtypes

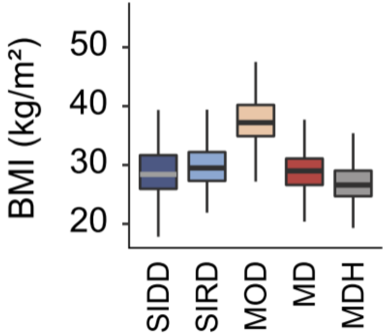
- Severe Autoimmune Diabetes (SAID)
- Severe Insulin-Deficient Diabetes (SIDD)
- Severe Insulin-Resistant Diabetes (SIRD)
- Mild Obesity-Related Diabetes (MOD)
- Mild Age-Related Diabetes (MARD)

Ahlqvist et al. (2018)

Previous Subtype Categorization







Ahlqvist et al. (2018)

- SIDD
- SIRD
- MOD
- MARD
- SAID

Xue et al. (2023)

- SIDD
- SIRD
- MOD
- MARD
- MARD+SIRD

Sliecker et al. (2021)

- SIDD
- SIRD
- MOD
- MD
- MDH

Research Question

How are the Type 2 Diabetes Mellitus subtypes distributed among the US population and how do they contribute to cardiovascular and kidney disease risks?

Aims

Aim 1

Identify type 2 diabetes subtypes and analyze the prevalence of related cardiovascular and kidney complications

Aim 2

Assess the relationship between type 2 diabetes subtypes and race/ethnicity

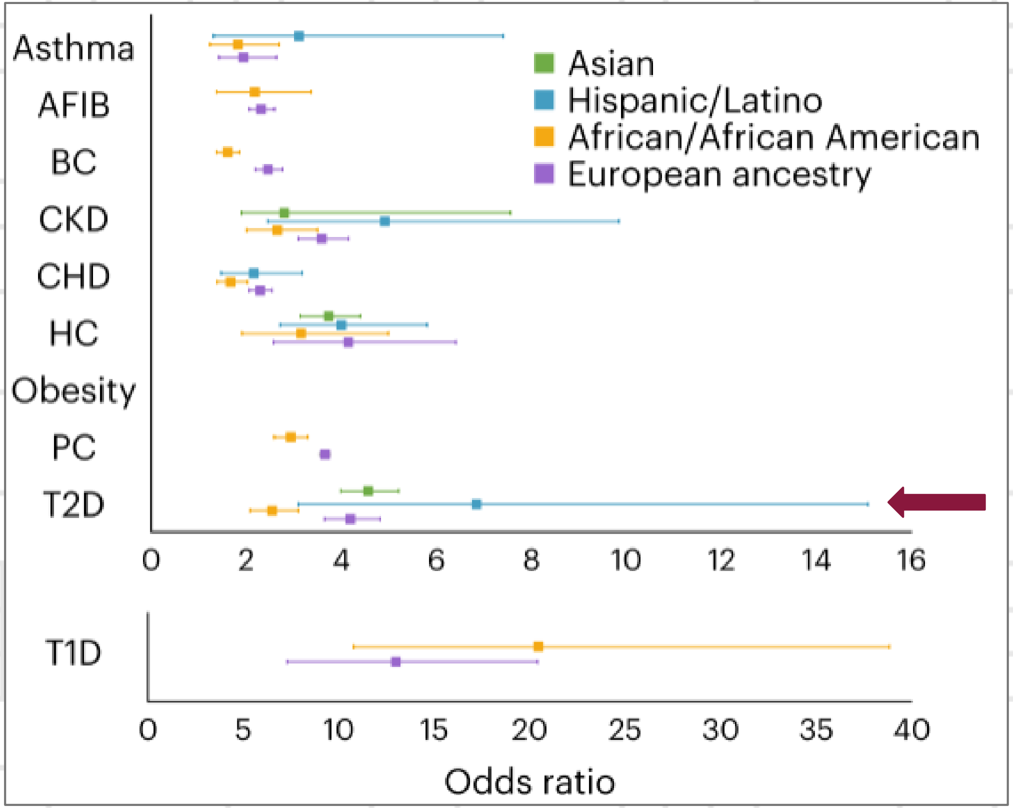
Aim 3

Model the and predict behavior of type 2 diabetes subtypes

Associated Risks Related to Race and Ethnicity

Hispanic/Latino Population at Higher Risk

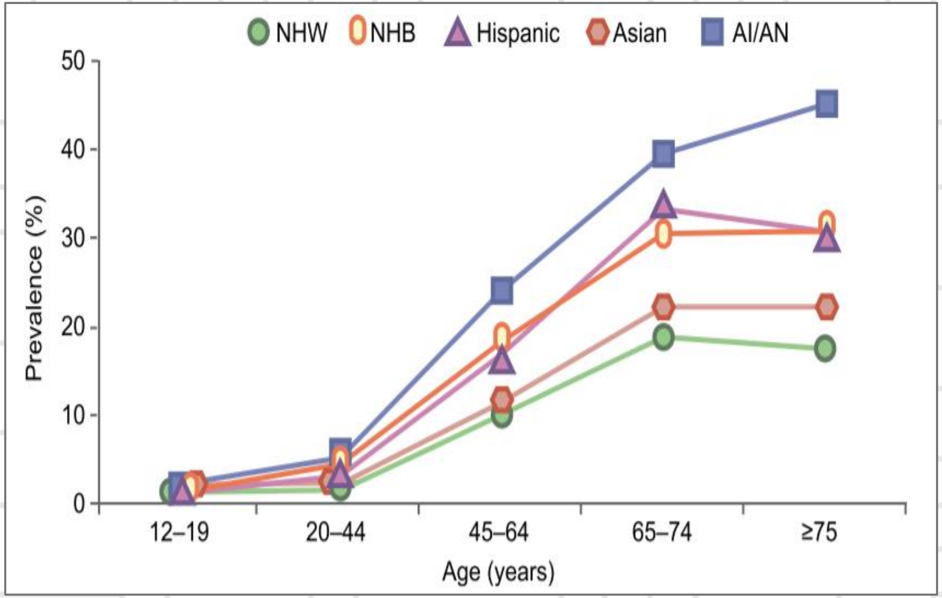
- Based off genetic predisposition (GWAS)
- Hispanics have the highest risk and variation in T2DM, while Africans have the lowest risk



Lennon et al. (2024)

Prevalence Across Race and Ethnicity

- T2DM prevalence **increases with age** in all races and ethnicities
- Non-genetic factors can lead to higher rates of disease in groups with lower genetic risk



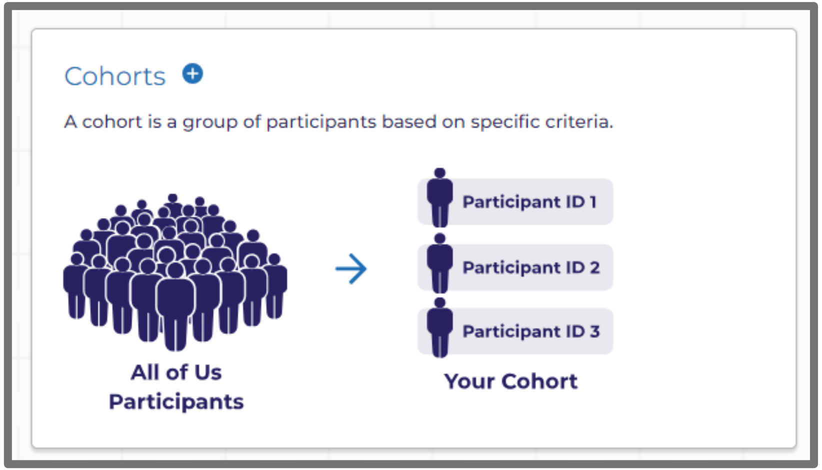
NHW = Non-Hispanic White
NHB = Non-Hispanic Black
AI/AN = American Indian/Alaska Native

Golden et al. (2019)

All of Us Database

All of Us Initiative and Database

1. NIH Biomedical Data Resource (US Participants)
2. 864,000+ Participants
3. Cloud based platform for data analysis
4. Different tiers of access
   - Public, Registered, Controlled
5. All participants have given consent for certified researchers to access their data

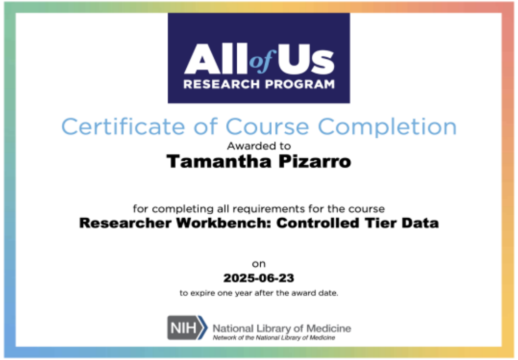


National Institutes of Health, 2025

Certification

1. Avoiding Stigma and Stigmatizing Research
2. Interaction between Biology and Society
3. Group Harms and Cultural Competence
4. Social Responsibility in Research


All of Us RESEARCH PROGRAM
Certificate of Course Completion
Awarded to
Matthew Jensen
for completing all requirements for the course
Researcher Workbench: Controlled Tier Data
on
2025-06-16
to expire one year after the award date.
NIH National Library of Medicine
Network of the National Library of Medicine


All of Us RESEARCH PROGRAM
Certificate of Course Completion
Awarded to
Tamantha Pizarro
for completing all requirements for the course
Researcher Workbench: Controlled Tier Data
on
2025-06-23
to expire one year after the award date.
NIH National Library of Medicine
Network of the National Library of Medicine


All of Us RESEARCH PROGRAM
Certificate of Course Completion
Awarded to
Adrian Martinez Felix
for completing all requirements for the course
Researcher Workbench: Controlled Tier Data
on
2025-06-15
to expire one year after the award date.
NIH National Library of Medicine
Network of the National Library of Medicine

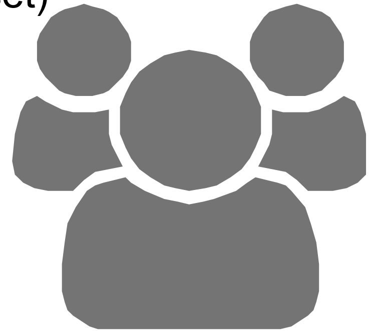

All of Us RESEARCH PROGRAM
Certificate of Course Completion
Awarded to
Jackson Deal
for completing all requirements for the course
Researcher Workbench: Controlled Tier Data
on
2025-06-15
to expire one year after the award date.
NIH National Library of Medicine
Network of the National Library of Medicine

Subtypes and Clustering

Importance of Clustering and Subtypes

1 Subtypes Have Different Characteristics

- Characteristics can impact complications
  - Kidney Function (Creatinine and eGFR)
  - Heart Disease/Stroke (HDL)
  - Elevated blood sugar levels (HBA1c)
  - Other Health Conditions (BMI, Age Onset)



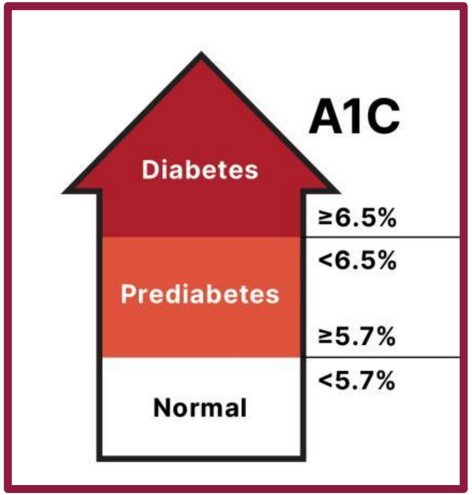
2 Personalized Treatment of T2DM and Complications

- Differing risks can mean different treatments per cluster
- Subtype identification important

Final Cohorts

1 Cohort 1 - Nondiabetic Control

- Include if:
  - A1c below 5.7%
- Exclude if:
  - T2DM or T1DM diagnosis
  - Cardiovascular or kidney complications



2 Cohort 2 - Prediabetes and T2DM

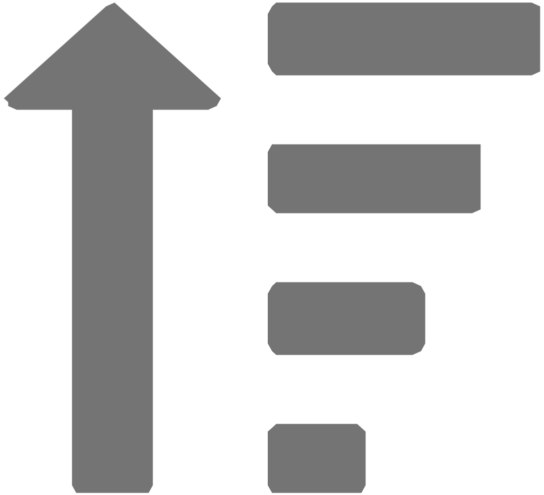
- Include if:
  - Any measurements in clustering variables (A1c, BMI, HDL, Creatinine) AND T2DM/Prediabetes Diagnosis
- Exclude if:
  - Cardiovascular or Kidney complications before diagnosis
  - T1DM Diagnosis
  - Pregnancy related diabetes

American Diabetes Association (2023)

Clustering Tendency - Hopkins Statistic

1 Hopkins Statistic

- Measures how likely the data is to be able to be clustered meaningfully
- Looks at the differences between our data and a uniformly distributed dataset
- Values closer to 1 indicate that our data is more likely to be clustered
- **Our Hopkins Value: ~0.9989**

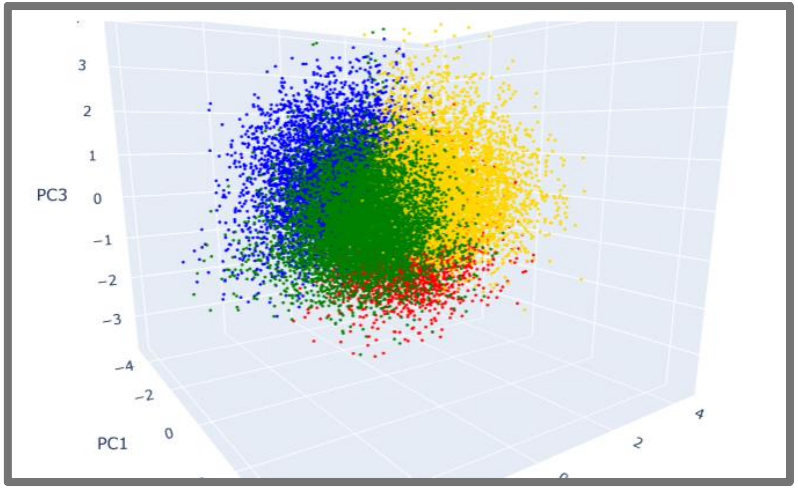


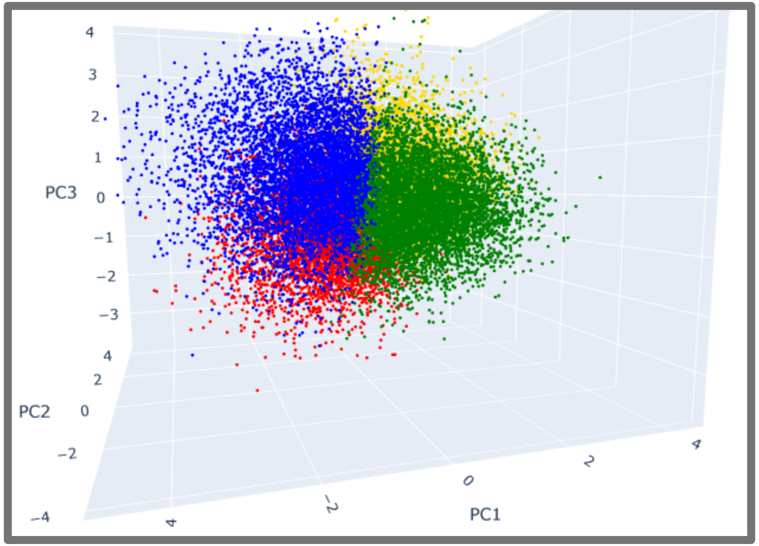
3D Clustering Graph

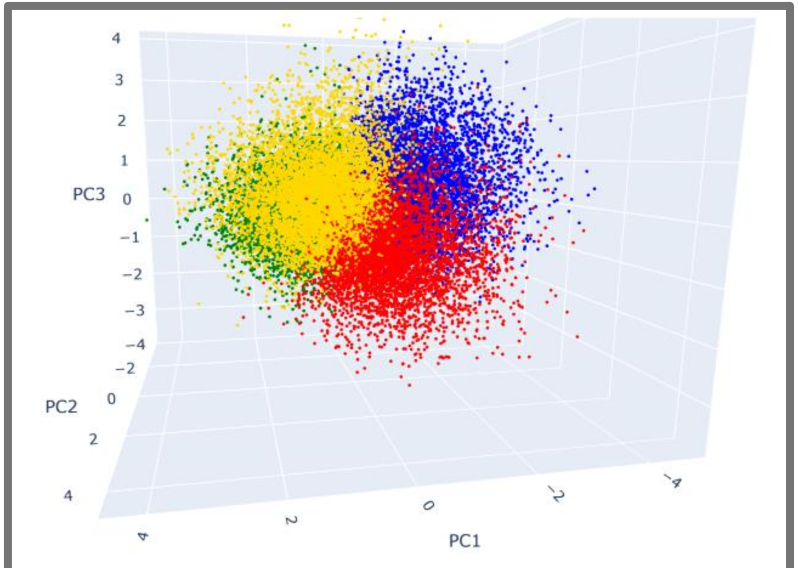
First three Principal Components

- 33.00% + 23.84% + 17.99% = 74.83%
- Stronger separation of clusters

1 - SIDD
2 - SIRD+MARD
3 - MOD
4 - MDH







Identifying Subtypes

Which clusters correspond to which subtypes?

- **Cluster 1** - Severe Insulin Deficient (SIDD)
- **Cluster 2** - Severe Insulin Resistant + Mild Age Related (SIRD+MARD)
- **Cluster 3** - Mild Obesity Related (MOD)
- **Cluster 4** - Mild Diabetes with high HDL cholesterol (MDH)

Median table of characteristics for each cluster. Important to note that each observation is an average of participants' measures

Cluster	Age Onset	BMI	A1c	eGFR	HDL
1 (SIDD)	54	32.348	8.054	93.714	41.750
2 (SIRD+MARD)	66	31.754	6.736	66.709	40.667
3 (MOD)	47	39.276	6.273	99.741	42.523
4 (MDH)	63	29.579	6.200	85.409	58.000

Cluster Stability - Mean Jaccard Coefficients

2 Mean Jaccard Coefficient per cluster

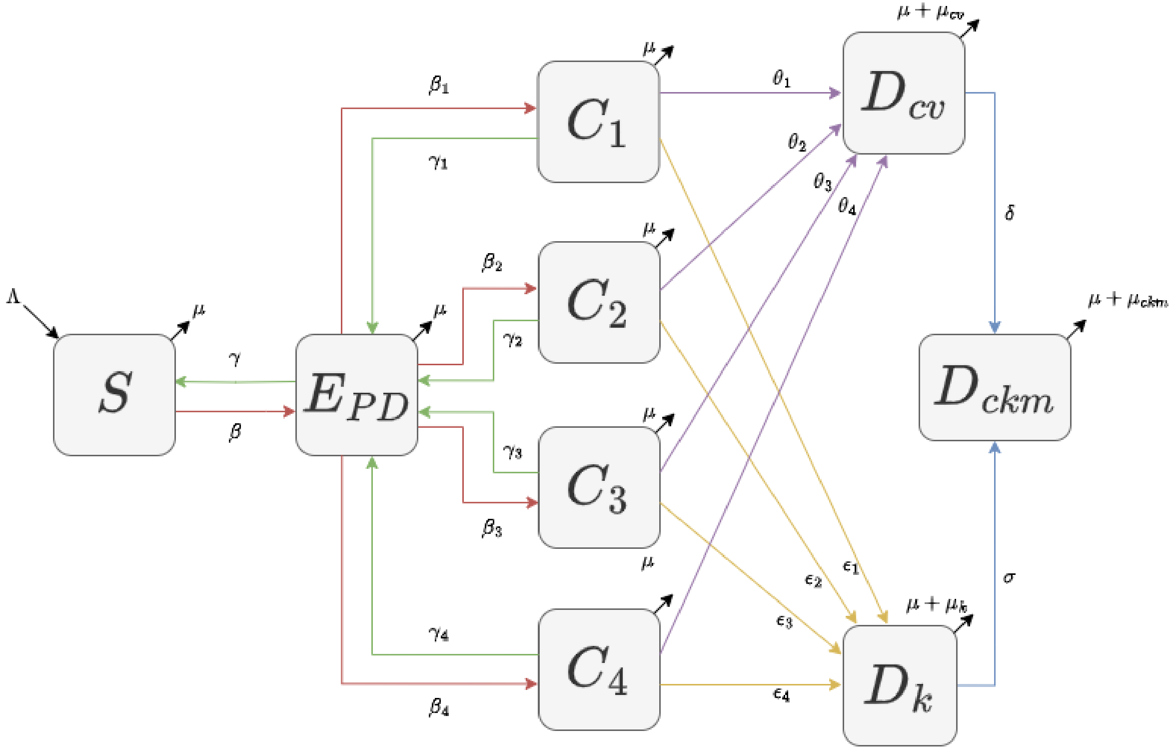
- Measures the similarity of points within the clusters
- Resamples original data to create new datasets which are then clustered
- Take the mean of all iterations
- Used 2000 resamplings
- Values closer to 1 indicate stronger stability

Jaccard similarity means for each cluster considering 2000 bootstraps. Coefficients range between 0 and 1. Typically, numbers above 0.75 are considered stable Anjana et al. (2020)

Cluster	Jaccard Similarity Mean
1 (SIDD)	0.9123
2 (SIRD+MARD)	0.9310
3 (MOD)	0.9807
4 (MDH)	0.8869

Modeling

Current Model



S(t): Susceptible

E_{PD}(t): Prediabetic

C_i(t) for i=1,2,3,4: Diabetic Clusters

D_{cv}(t): Diabetes with a cardiovascular complication

D_k(t): Diabetes with a kidney complication

D_{ckm}(t): Diabetes with a cardiovascular and kidney complication

Model

$$\frac{dS}{dt} = \Lambda - \beta S + \gamma E_{PD} - \mu S, \tag{1a}$$

$$\frac{dE_{PD}}{dt} = \beta S + \gamma_1 C_1 + \gamma_2 C_2 + \gamma_3 C_3 + \gamma_4 C_4 - (\beta_1 + \beta_2 + \beta_3 + \beta_4 + \gamma + \mu) E_{PD}, \tag{1b}$$

$$\frac{dC_1}{dt} = \beta_1 E_{PD} - (\mu + \gamma_1 + \theta_1 + \epsilon_1) C_1, \tag{1c}$$

$$\frac{dC_2}{dt} = \beta_2 E_{PD} - (\mu + \gamma_2 + \theta_2 + \epsilon_2) C_2, \tag{1d}$$

$$\frac{dC_3}{dt} = \beta_3 E_{PD} - (\mu + \gamma_3 + \theta_3 + \epsilon_3) C_3, \tag{1e}$$

$$\frac{dC_4}{dt} = \beta_4 E_{PD} - (\mu + \gamma_4 + \theta_4 + \epsilon_4) C_4, \tag{1f}$$

$$\frac{dD_{CV}}{dt} = \theta_1 C_1 + \theta_2 C_2 + \theta_3 C_3 + \theta_4 C_4 - (\delta + \mu + \mu_{cv}) D_{CV}, \tag{1g}$$

$$\frac{dD_K}{dt} = \epsilon_1 C_1 + \epsilon_2 C_2 + \epsilon_3 C_3 + \epsilon_4 C_4 - (\sigma + \mu + \mu_k) D_K, \tag{1h}$$

$$\frac{dD_{CKM}}{dt} = \delta D_{CV} + \sigma D_K - (\mu + \mu_{ckm}) D_{CKM}. \tag{1i}$$

Mathematical Analysis

Proposition

The closed set

$$D = \{(S, E_{pd}, C_1, C_2, C_3, C_4, C_5, D_{cv}, K, D_{ckm}) \in \mathbb{R}_+^{10} : \mathrm{N} \le \mathrm{N}^*\}$$

is positively-invariant and attracting.

Here,

$$N^* = \Lambda/\mu$$

is the asymptotic value for the supersolution N.

Mathematical Analysis

Equilibrium Point

Assuming compartments $C_{1,2,3,4,5}$ are condensed into one compartment C_i, then our model has one interior positive equilibrium point $(S^*, E^*_{PD}, C^*_i, D^*_{CV}, K^*, C^*_{KM})$. By setting

$$\frac{dS}{dt} = \frac{dE_{PD}}{dt} = \frac{dC_i}{dt} = \frac{dD_{CV}}{dt} = \frac{dK}{dt} = \frac{dD_{KM}}{dt} = 0,$$

we get the following equilibrium

$$S^* = \frac{\lambda\left(\beta_i(\epsilon_i + \mu + \theta_i) + (\gamma + \mu)(\gamma_i + \epsilon_i + \theta_i + \mu)\right)}{\beta_i(\theta_i + \mu + \epsilon_i)(\beta + \mu) + \mu(\gamma_i + \theta_i + \mu + \epsilon_i)(\beta + \gamma + \mu)}$$

$$E^*_{PD} = \frac{\beta\lambda(\gamma_i + \theta_i + \mu + \epsilon_i)}{\beta_i(\theta_i + \mu + \epsilon_i)(\beta + \mu) + \mu(\gamma_i + \theta_i + \mu + \epsilon_i)(\beta + \gamma + \mu)}$$

$$C^*_i = \frac{\beta\beta_i\lambda}{\beta_i(\theta_i + \mu + \epsilon_i)(\beta + \mu) + \mu(\gamma_i + \theta_i + \mu + \epsilon_i)(\beta + \gamma + \mu)}$$

$$D^*_{CV} = \frac{\beta\beta_i\theta_i\lambda}{\left(\beta_i(\theta_i + \mu + \epsilon_i)(\beta + \mu) + \mu(\gamma_i + \theta_i + \mu + \epsilon_i)(\beta + \gamma + \mu)\right)(\delta + \mu + \mu_{CV})}$$

$$K^* = \frac{\beta\beta_i\epsilon_i\lambda}{\left(\beta_i(\theta_i + \mu + \epsilon_i)(\beta + \mu) + \mu(\gamma_i + \theta_i + \mu + \epsilon_i)(\beta + \gamma + \mu)\right)(\mu + \mu_K + \sigma)}$$

$$C^*_{KM} = \frac{\beta\beta_i\lambda\left(\delta\theta_i(\mu + \mu_K) + \delta\sigma(\epsilon_i + \theta_i) + \epsilon\sigma(\mu + \mu_{CV})\right)}{\left(\beta_i(\theta_i + \mu + \epsilon_i)(\beta + \mu) + \mu(\gamma_i + \theta_i + \mu + \epsilon_i)(\beta + \gamma + \mu)\right)(\mu + \mu_{CKM})(\delta + \mu + \mu_{CV})(\mu + \mu_K + \sigma)}$$

This equilibrium always exists and is stable (by using Routh-Hurwitz criterion)

Results

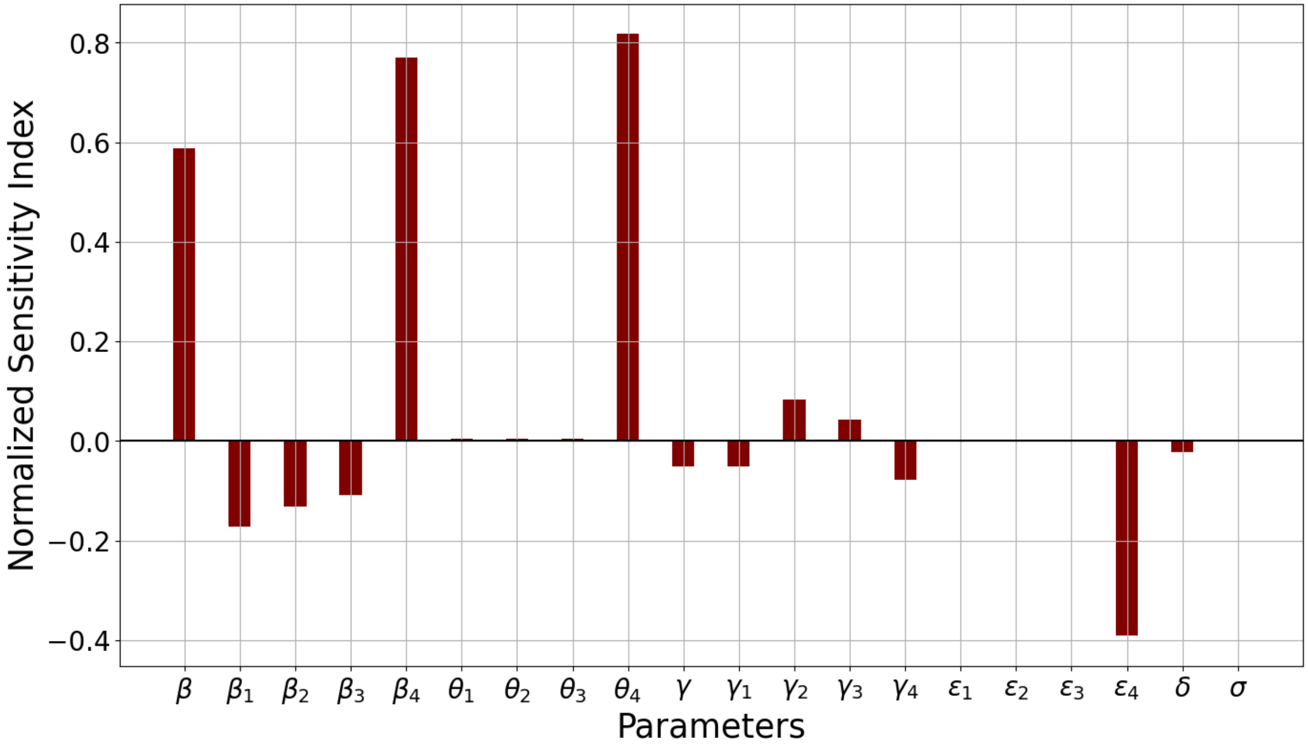
Our Fitted Parameters

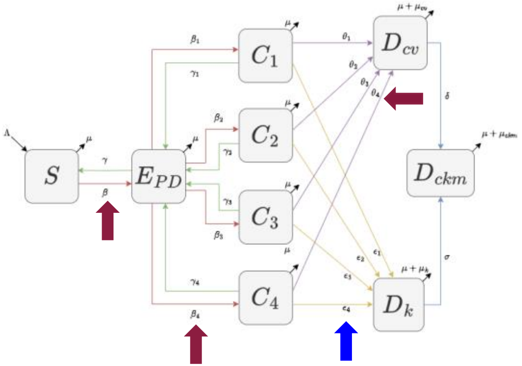
Model diagram of T2D within a population including different subtypes

Parameter	Description	Value	Units	References
Λ	Recruitment rate	987.015	people·month^{-1}	Estimated
β	Transfer from susceptible to prediabetic	0.009802	month^{-1}	Estimated
γ	Recovery from prediabetic to susceptible	0.003767	month^{-1}	Estimated
μ	Natural death rate	0.01271	month^{-1}	CDC
γ_i	Recovery from clusters to prediabetic	[Expanded in following table]	month^{-1}	Estimated
β_i	Transition from prediabetic to clusters	[Expanded in following table]	month^{-1}	Estimated
θ_i	Transition to diabetes with cardiovascular complications	[Expanded in following table]	month^{-1}	Estimated
μ_{CV}	Death rate with cardiovascular complications	0.0015072541	month^{-1}	Xue et al. (2023)
μ_K	Death rate with kidney complications	0.0036783	month^{-1}	Zhao et al. (2025)
μ_{CKM}	Death rate with CKM complications	0.007815	month^{-1}	Zhao et al. (2025)
ϵ_i	Transition to diabetes with kidney disease	[Expanded in following table]	month^{-1}	Estimated
δ	Progression from CV to CKM	0.0003181	month^{-1}	Estimated
σ	Progression from kidney to CKM	0.0003181	month^{-1}	Estimated

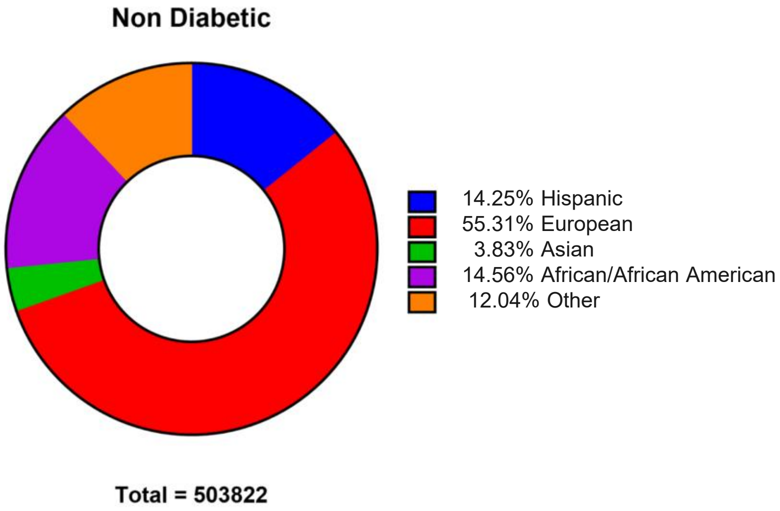
Parameter	Value	Units
β_1	0.01935	month^{-1}
β_2	0.01540	month^{-1}
β_3	0.00785	month^{-1}
β_4	0.01011	month^{-1}
θ_1	0.000013	month^{-1}
θ_2	0.000017	month^{-1}
θ_3	0.000017	month^{-1}
θ_4	0.00619	month^{-1}
ϵ_1	0.000047	month^{-1}
ϵ_2	0.000065	month^{-1}
ϵ_3	0.000088	month^{-1}
ϵ_4	0.01426	month^{-1}
γ_1	0.02103	month^{-1}
γ_2	0.02179	month^{-1}
γ_3	0.00847	month^{-1}
γ_4	0.00372	month^{-1}

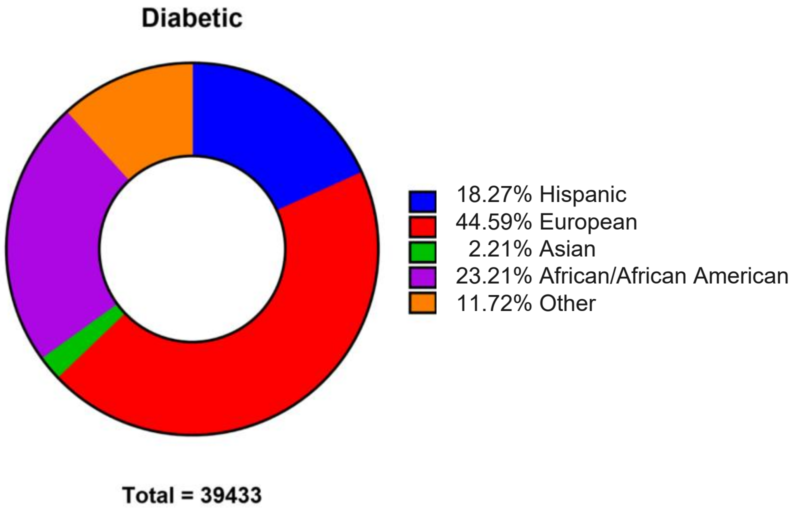
Sensitivity Analysis for D^*_{CV}





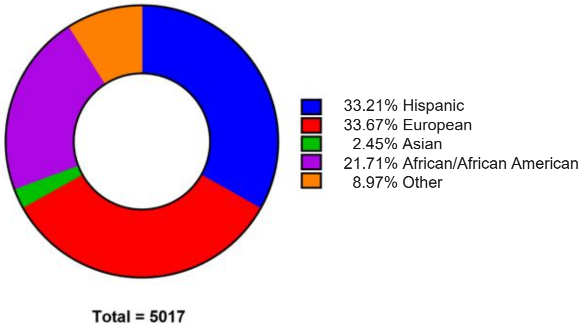
Type 2 Diabetes by Race/Ethnicity (*All of Us*)



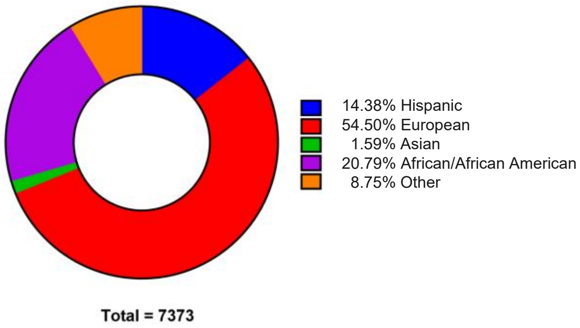


Race/Ethnicity in Each Cluster

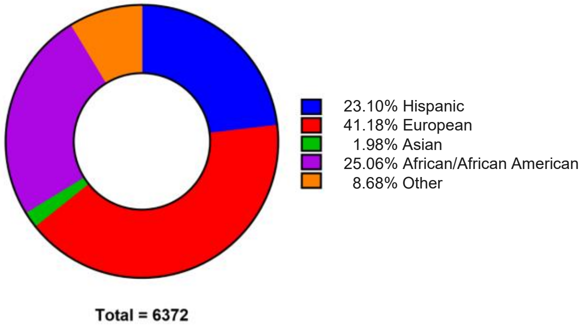
C1: SIDD



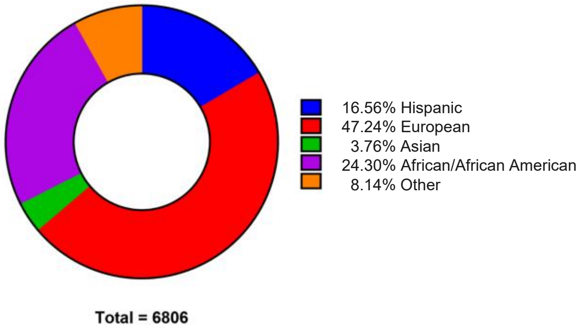
C2: SIRD+MARD



C3: MOD



C4: MDH



Conclusions and Future Studies

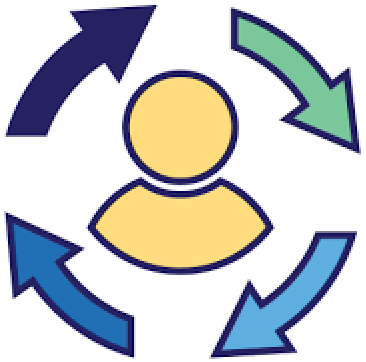
Selected Conclusions

1. Four distinct T2DM subtypes were identified using K-means clustering:
   - Severe Insulin Deficiency Diabetes (SIDD)
   - Severe Insulin Resistant + Mild Age Related (SIRD+MARD)
   - Mild Obesity Related (MOD)
   - Mild Diabetes with high HDL cholesterol (MDH)

2. Subtype classification was associated with different risks supporting the need for personalized management strategies

3. Race and ethnicity influenced subtype distributions suggesting genetic and potential sociodemographic contributions to subtype risk

4. The model could be used as a tool for predicting disease progression allowing tailored interventions based on subtype and risk factors



Future Work

01 Improve the model

- Currently captures general trends
- Want more accurate predictions

02 More complications

- Retinopathy and Neuropathy in clusters with high HBA1c

03 Undiagnosed Individuals

- Current data analysis relies on diabetes diagnosis
- Considering them increases model accuracy

04 Model with Race/Ethnicity and Genetics

- Include race/ethnicity and genetic predisposition in model parameters
- Focus more on Hispanic population

Acknowledgements

We would like to thank Drs. Fabio Milner, Co-Director of the Simon A. Levin Mathematical, Computational and Modeling Sciences Center, and John Nagy for giving us the opportunity to participate in the Quantitative Research in the Life and Social Sciences program (QRLSSP). We also recognize the work of the many administrative staff members –Riley Pass, Kimberly Dietmeyer, Tracey Simpson, and Elaine Rudolph– who supported this effort, and the academic support personnel: Jordy Cevallos Chavez, Carlos Bustamante Orellana, Unnati Gohil, Mason Manning, Tamantha Pizarro, and Abhishek Yadav. This research was conducted as part of the 2025 Simon A. Levin Center's (MCMSC) QRLSSP at Arizona State University (ASU) and has been partially supported by The College of Liberal Arts and Sciences of ASU.

We gratefully acknowledge *All of Us* participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's *All of Us Research Program* for making available the participant data and cohort data examined in this study. This study used data from the *All of Us* Research Program's Controlled Tier Dataset version 8, available to authorized users on the researcher workbench

THANK YOU

Any Questions?

References

1. Afkarian, M., Sachs, M. C., Kestenbaum, B., Hirsch, I. B., Tuttle, K. R., Himmelfarb, J., & de Boer, I. H. (2013). Kidney Disease and Increased Mortality Risk in Type 2 Diabetes. Journal of the American Society of Nephrology, 24(2), 302–308. Kidney Disease and Increased Mortality Risk in Type 2 Diabetes
2. Ahlqvist, E., Storm, P., Käräjämäki, A., Martinell, M., Dorkhan, M., Carlsson, A., Vikman, P., Prasad, R. B., Aly, D. M., Almgren, P., Wessman, Y., Shaat, N., Spégel, P., Mulder, H., Lindholm, E., Melander, O., Hansson, O., Malmqvist, U., Lernmark, Å., … Groop, L. (2018). Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *The Lancet Diabetes and Endocrinology*, *6*(5), 361–369. https://doi.org/10.1016/S2213-8587(18)30051-2
3. American Diabetes Association. (2023). Understanding Diabetes Diagnosis. Diabetes.org; American Diabetes Association. https://diabetes.org/about-diabetes/diagnosis
4. Anjana, R. M., Baskar, V., Nair, A. T. N., Jebarani, S., Siddiqui, M. K., Pradeepa, R., Unnikrishnan, R., Palmer, C., Pearson, E., & Mohan, V. (2020). Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: A data-driven cluster analysis: The INSPIRED study. BMJ Open Diabetes Research and Care, 8(1). https://doi.org/10.1136/bmjdrc-2020-001506
5. Cao, J., Yan, W., Ma, X., Huang, H., & Yan, H. (2021). Insulin-like Growth Factor 2 mRNA-Binding Protein 2-a Potential Link between Type 2 Diabetes Mellitus and Cancer. In *Journal of Clinical Endocrinology and Metabolism* (Vol. 106, Issue 10, pp. 2807–2818). Endocrine Society. https://doi.org/10.1210/clinem/dgab391
6. Centers for Disease Control and Prevention. (2024). National diabetes statistics report. Centers for Disease Control and Prevention. https://www.cdc.gov/diabetes/php/data-research/index.html

References

7. Golden, S.H., Yajnik, C., Phatak, S. et al. Racial/ethnic differences in the burden of type 2 diabetes over the life course: a focus on the USA and India. Diabetologia 62, 1751–1760 (2019). https://doi.org/10.1007/s00125-019-4968-0
8. Henson, J., Tziannou, A., Rowlands, A. v., Edwardson, C. L., Hall, A. P., Davies, M. J., & Yates, T. (2024). Twenty-four-hour physical behaviour profiles across type 2 diabetes mellitus subtypes. *Diabetes, Obesity and Metabolism*, *26*(4), 1355–1365. https://doi.org/10.1111/dom.15437
9. Hennig, C. (2007). Cluster-wise assessment of cluster stability. Computational Statistics and Data Analysis, 52(1), 258–271. https://doi.org/10.1016/j.csda.2006.11.025
10. Hwang, Y. C., Ahn, H. Y., Jun, J. E., Jeong, I. K., Ahn, K. J., & Chung, H. Y. (2023). Subtypes of type 2 diabetes and their association with outcomes in Korean adults - A cluster analysis of community-based prospective cohort. Metabolism: Clinical and Experimental, 141. https://doi.org/10.1016/j.metabol.2023.155514
11. Karter, A. J., Nundy, S., Parker, M. M., Moffet, H. H., & Huang, E. S. (2014). Incidence of remission in adults with type 2 diabetes: The diabetes & aging study. *Diabetes Care*, *37*(12), 3188–3195. https://doi.org/10.2337/dc14-0874
12. Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning*. STHDA. https://xsliulab.github.io/Workshop/2021/week10/r-cluster-book.pdf

References

12. Kouidere, A., Balatif, O., Ferjouchia, H., Boutayeb, A., & Rachik, M. (2019). Optimal control strategy for a discrete time to the dynamics of a population of diabetics with highlighting the impact of living environment. *Discrete Dynamics in Nature and Society*, *2019*. https://doi.org/10.1155/2019/6342169
13. Lazo-Porras, M., Bernabe-Ortiz, A., Ruiz-Alejos, A., Smeeth, L., Gilman, R. H., Checkley, W., Málaga, G., & Miranda, J. J. (2020). Regression from prediabetes to normal glucose levels is more frequent than progression towards diabetes: The CRONICAS Cohort Study. Diabetes Research and Clinical Practice, 163. https://doi.org/10.1016/j.diabres.2019.107829
14. Lennon, N. J., Kottyan, L. C., Kachulis, C., Abul-Husn, N. S., Arias, J., Belbin, G., Below, J. E., Berndt, S. I., Chung, W. K., Cimino, J. J., Clayton, E. W., Connolly, J. J., Crosslin, D. R., Dikilitas, O., Velez Edwards, D. R., Feng, Q. P., Fisher, M., Freimuth, R. R., Ge, T., … Loos, R. (2024). Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nature Medicine*, *30*(2), 480–487. https://doi.org/10.1038/s41591-024-02796-z
15. Murphy, S. L., Kochanek, K. D., Xu, J., & Arias, E. (2024). Mortality in the United States, 2023. Mortality in the United States, 2023, 521. https://doi.org/10.15620/cdc/170564
16. Pigeyre, M., Hess, S., Gomez, M. F., Asplund, O., Groop, L., Pare, G., & Gerstein, H. (2022). Validation of the classification for type 2 diabetes into five subgroups: a report from the ORIGIN trial. *Diabetologia*, *65*, 206-215.

References

16. Office of Minority Health. (2024b, September 25). Diabetes and Hispanic Americans. Minorityhealth.hhs.gov. https://minorityhealth.hhs.gov/diabetes-and-hispanic-americans
17. Rapoport, M., Chetrit, A., Cantrell, D., Novikov, I., Roth, J., & Dankner, R. (2021). Years of potential life lost in pre-diabetes and diabetes mellitus: Data from a 40-year follow-up of the Israel study on Glucose intolerance, Obesity and Hypertension. *BMJ Open Diabetes Research and Care*, *9*(1). https://doi.org/10.1136/bmjdrc-2020-001981
18. Usman, M. S., Khan, M. S., & Butler, J. (2021, June). *The Interplay Between Diabetes, Cardiovascular Disease, and Kidney Disease*. PubMed; American Diabetes Association. https://www.ncbi.nlm.nih.gov/books/NBK571718/
19. Slieker, R. C., Donnelly, L. A., Fitipaldi, H., Bouland, G. A., Giordano, G. N., Åkerlund, M., Gerl, M. J., Ahlqvist, E., Ali, A., Dragan, I., Festa, A., Hansen, M. K., Aly, D. M., Kim, M., Kuznetsov, D., Mehl, F., Klose, C., Simons, K., Pavo, I., … Pearson, E. R. (2021). *Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study Mild diabetes MDH Mild diabetes with high HDL-cholesterol MOD Mild obesity-related diabetes RHAPSODY Risk Assessment and ProgreSsiOn of DIabetes SIDD Severe insulin-deficient diabetes SIRD Severe insulin-resistant diabetes*. https://doi.org/10.1007/s00125-021-05490-8/Published
20. Tanabe, H., Sato, M., Miyake, A., Shimajiri, Y., Ojima, T., Narita, A., Saito, H., Tanaka, K., Masuzaki, H., Kazama, J. J., Katagiri, H., Tamiya, G., Kawakami, E., & Shimabukuro, M. (2024). Machine learning-based reproducible prediction of type 2 diabetes subtypes. *Diabetologia*. https://doi.org/10.1007/s00125-024-06248-8

References

20. Tripathi, P., Vyawahare, A., Kadam, N., Tiwari, D., Biswas, M. das, Kathrikolly, T., Sharma, B., Vijayakumar, V., & Kuppusamy, M. (2024). Comparison of clustering and phenotyping approaches for subclassification of type 2 diabetes and its association with remission in Indian population. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-71126-7
21. Tyagi, S., Kaur Bhatia, S., Das, A., & Mittal, T. (2024). Dynamic analysis of a diabetes mathematical model including heredity parameter and multiple time delays. *Nonlinear Dynamics*. https://doi.org/10.1007/s11071-024-10776-8
22. (19) Shariq Usman, M., Shahzeb Khan, M., & Butler, J. (2023.). *CHRONIC KIDNEY DISEASE AND TYPE 2 DIABETES The Interplay Between Diabetes, Cardiovascular Disease, and Kidney Disease Burden of Diabetes and Associated Cardiorenal Disorders*. https://www.ncbi.nlm.nih.gov/books/NBK571718/
23. Vidal, T. M., Williams, C. A., Ramoutar, U. D., & Haffizulla, F. (2022). Type 2 Diabetes Mellitus in Latinx Populations in the United States: A Culturally Relevant Literature Review. *Cureus*. https://doi.org/10.7759/cureus.23173
24. Xue, Q., Li, X., Wang, X., Ma, H., Heianza, Y., & Qi, L. (2023). Subtypes of Type 2 Diabetes and Incident Cardiovascular Disease Risk: UK Biobank and All of Us Cohorts. *Mayo Clinic Proceedings*, *98*(8), 1192–1204. https://doi.org/10.1016/j.mayocp.2023.01.024
25. Zhao, S., Liu, H., & Liu, A. (2025). Associations of cardiovascular-kidney-metabolic conditions with all-cause and cardiovascular mortality among US adults: a population-based cohort study. BMC Cardiovascular Disorders, 25(1), 443. https://doi.org/10.1186/s12872-025-04910-6

Supplementary Materials

Data Cleaning

What we did

1. Calculated Age at Onset
2. Calculated eGFR from Creatinine
3. Excluded outliers
   - Domain knowledge
   - Box plots and quartiles
4. Calculated averages of measurements for each person

What we used

1. Person ID, DoB, condition start
2. Person ID, measurement time and value, DoB, sex at birth
3. Measurement value, measurement units
4. Person ID and measurement value

Conditions Data

- Person ID
- Condition name
- Condition start time

Demographic Data

- Person ID
- Date of Birth
- Sex at birth
- Race/Ethnicity

Measurement Data

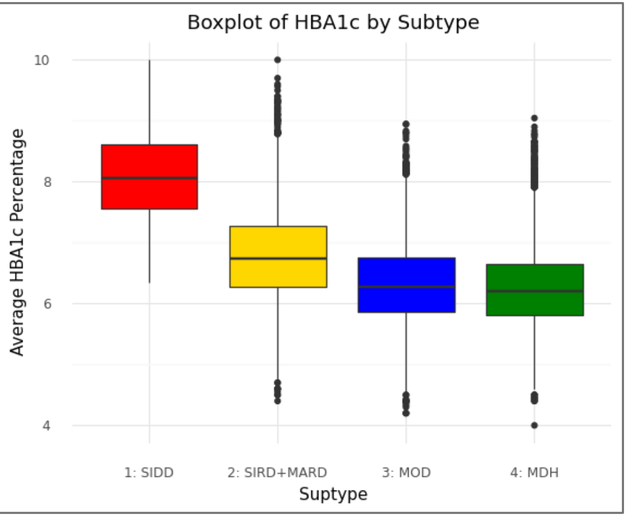
- Person ID
- Measurement name
- Measurement value
- Measurement time
- Measurement Unit

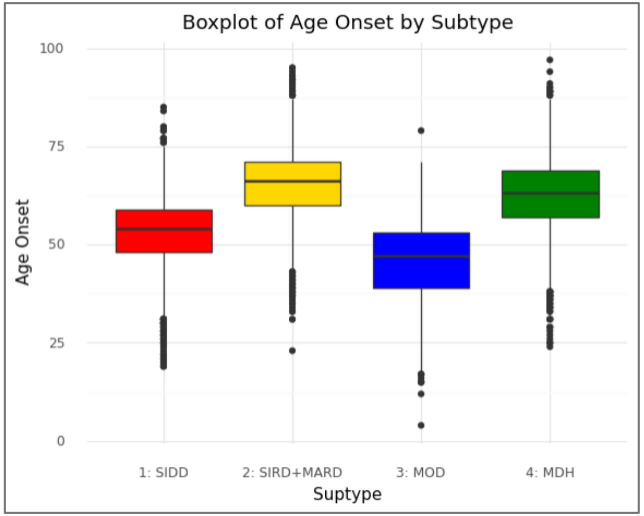
Mathematical Analysis

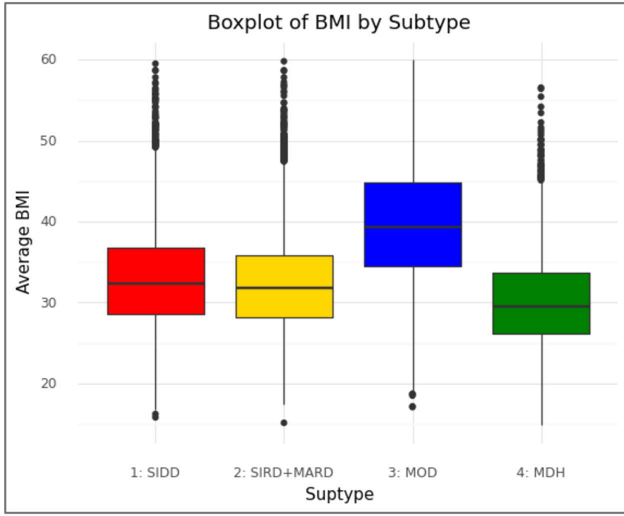
Equilibrium Point

$$J(S^*, E^*_{PD}, C^*_i, D^*_{CV}, K^*, C^*_{KM}) = \begin{bmatrix} -\beta-\mu & \gamma & 0 & 0 & 0 & 0 \\ \beta & -\beta_i-\gamma-\mu & \gamma_i & 0 & 0 & 0 \\ 0 & \beta_i & -\gamma_i-\epsilon_i-\theta_i-\mu & 0 & 0 & 0 \\ 0 & 0 & \theta_i & -\delta-\mu-\mu_{CV} & 0 & 0 \\ 0 & 0 & \epsilon_i & 0 & -\mu-\mu_K-\sigma & 0 \\ 0 & 0 & 0 & \delta & \sigma & -\mu-\mu_{CKM} \end{bmatrix}. \tag{3}$$

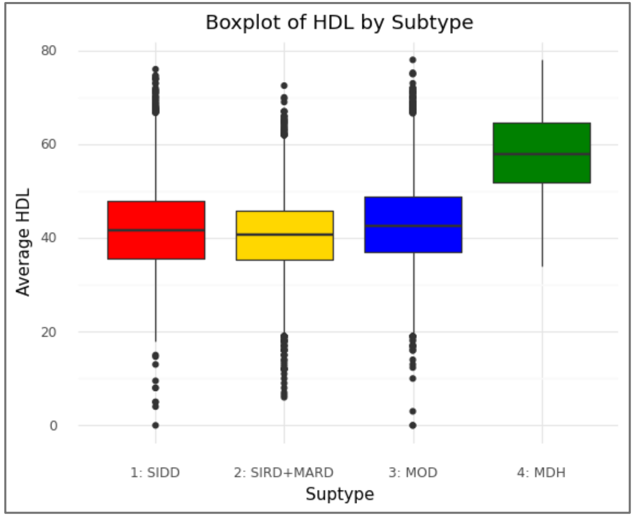
Identifying Subtypes

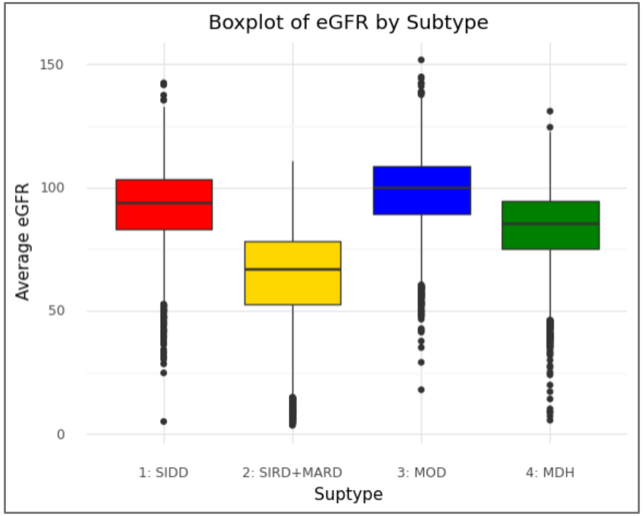






Identifying Subtypes





Mathematical Analysis

Equilibrium Point

$$J^* = \begin{bmatrix} -\beta - \mu & \gamma & 0 \\ \beta & -\beta_i - \gamma - \mu & \gamma_i \\ 0 & \beta_i & -\gamma_i - \epsilon_i - \theta_i - \mu \end{bmatrix}$$

- Stability proven by Routh-Hurwitz criterion.

Mathematical Analysis

Equilibrium Point

$$J^* = \begin{bmatrix} -\beta - \mu & \gamma & 0 \\ \beta & -\beta_i - \gamma - \mu & \gamma_i \\ 0 & \beta_i & -\gamma_i - \epsilon_i - \theta_i - \mu \end{bmatrix}$$

- Stability proven by Routh-Hurwitz criterion.

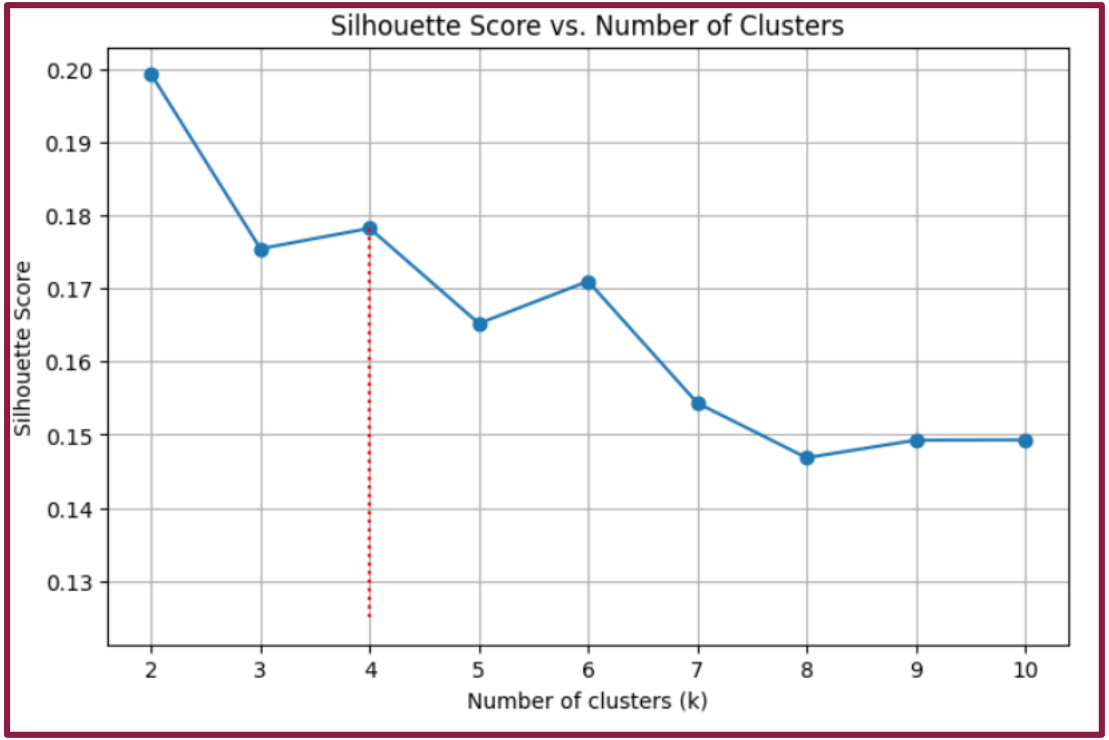
Clustering

1 K-means Clustering

- Participant Averages
- Sklearn kmeans
- N = 25,568
- HbA1c, BMI, Age at Onset, eGFR, HDL

2 Silhouette Method

- Shows optimal number of clusters
- All previous studies we have looked at found 4-5 clusters
- Precedent for using second highest silhouette score



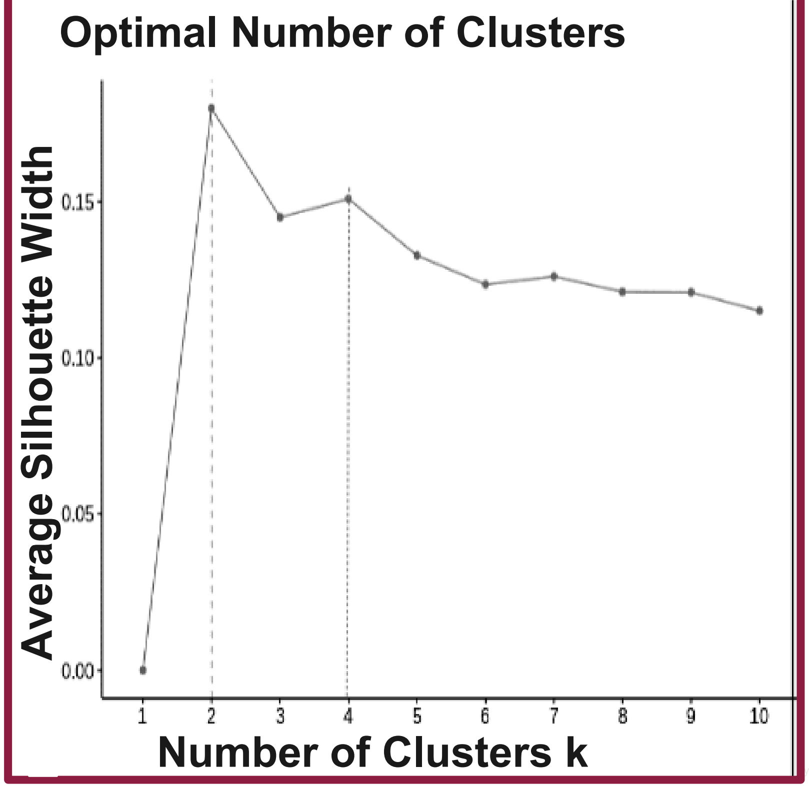
Anjana et al., 2020 Type 2 Diabetes Subtypes in Asian and Indian Populations

1 **General Findings**

- 7 Clustering variables
- 4 replicable clusters
- N=19,084
- MARD, SIDD, IROD, CIRDD
- New ones are combinations of ones from other papers

2 **Silhouette Method**

- Despite 2 being higher, they chose k=4



2 Cluster Medians to Compare

Cluster	Age Onset	BMI	A1c	eGFR	HDL
1*	65	30.4000	6.55	76.308	48.000
2*	49	36.768	6.80	98.422	42.000

*Note: Cluster 1 and 2 here *do not* directly correspond to cluster 1 and 2 in other slides

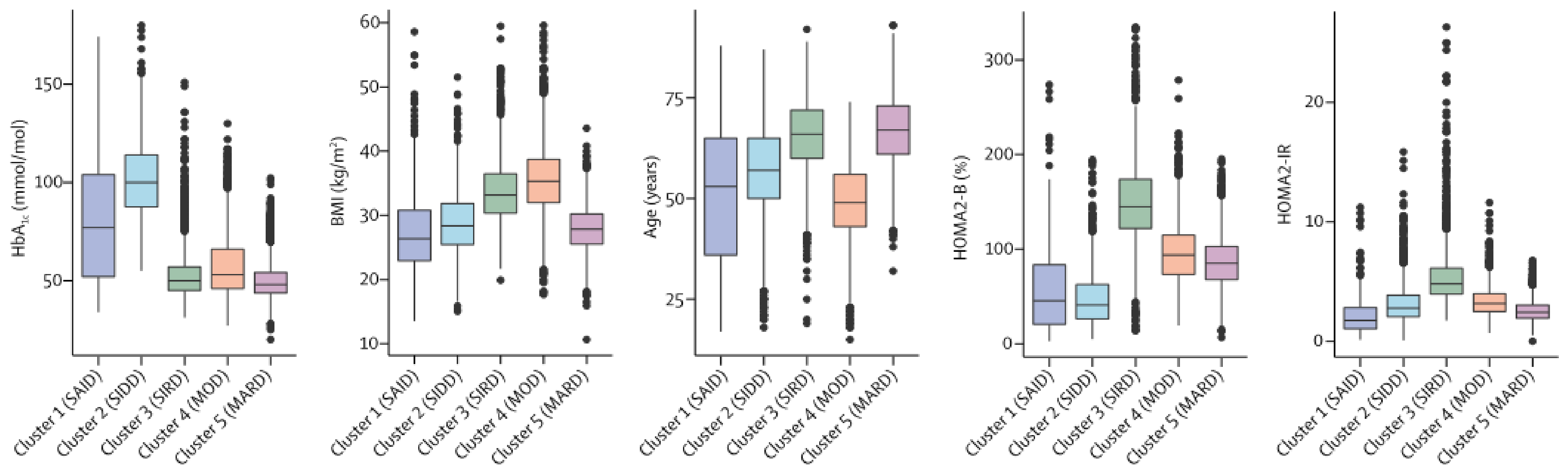
Our Cluster Medians to Compare

Which clusters correspond to which subtypes?

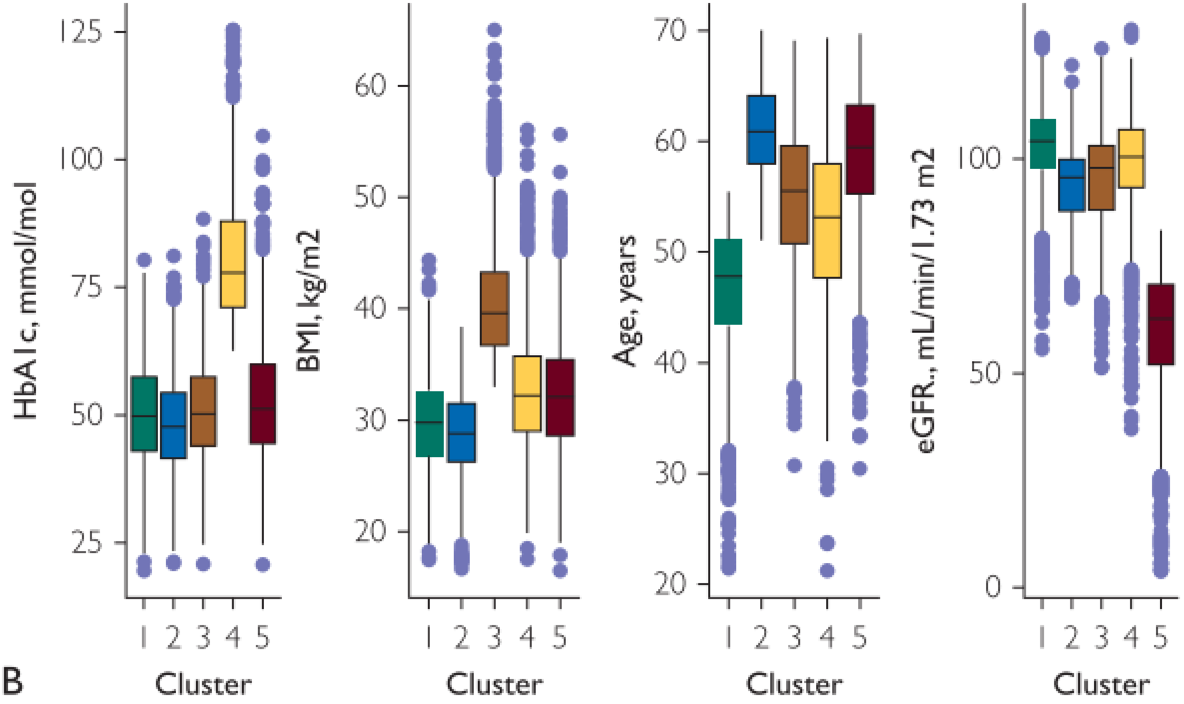
- Cluster 1 - Severe Insulin Deficient (SIDD)
- Cluster 2 - Severe Insulin Resistant + Mild Age Related (SIRD+MARD)
- Cluster 3 - Mild Obesity Related (MOD)
- Cluster 4 - Mild Diabetes with high HDL cholesterol

Cluster	Age Onset	BMI	A1c	eGFR	HDL
1	54	32.348	8.054	93.714	41.750
2	66	31.754	6.736	66.709	40.667
3	47	39.276	6.273	99.741	42.523
4	63	29.579	6.200	85.409	58.000

Ahlqvist Medians to Compare

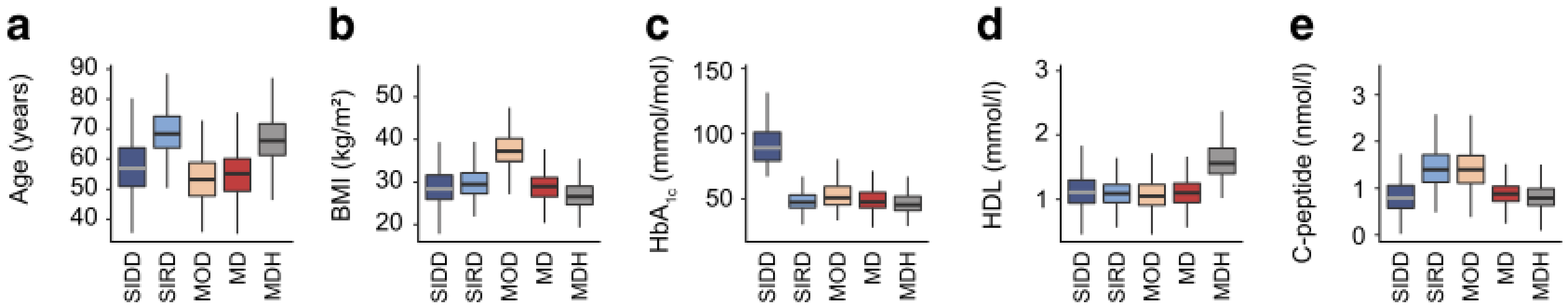


Xue Medians to Compare



B

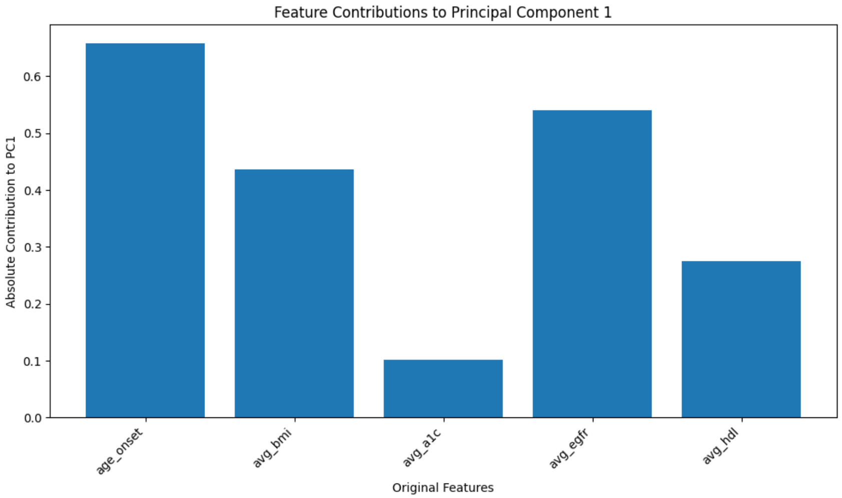
Sliecker Medians to Compare

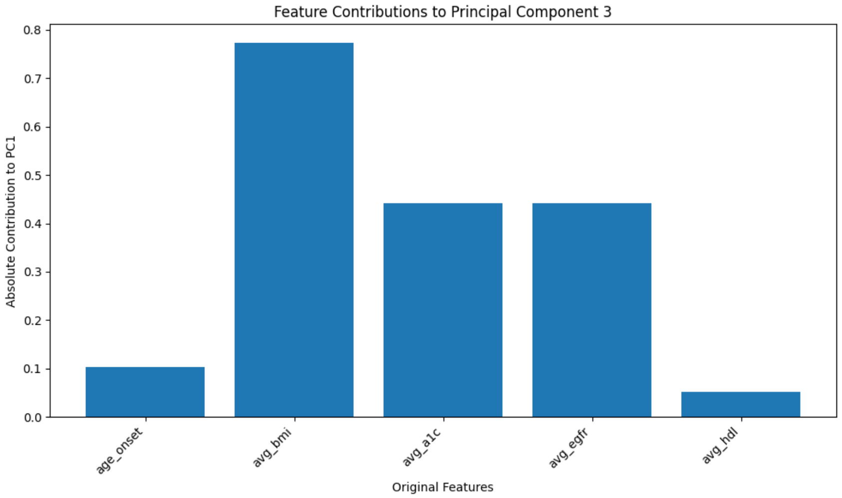


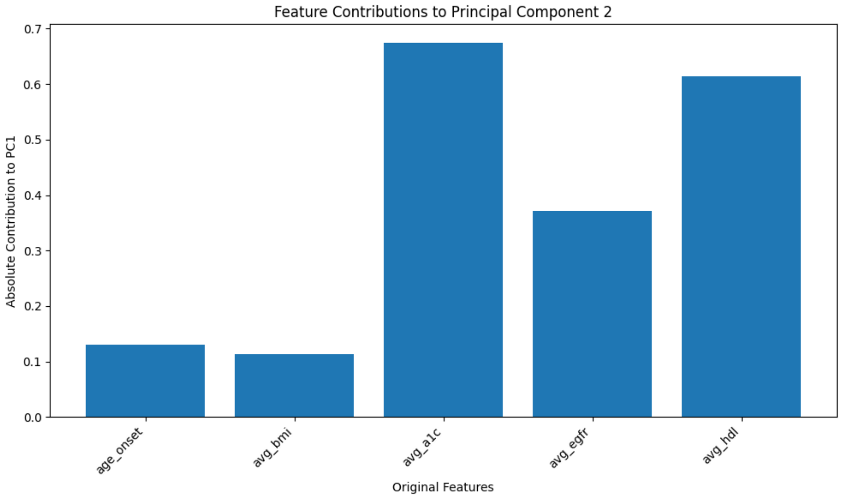
PC Contributions

First three Principal Components

- 33.00% + 23.84% + 17.99% = 74.83%





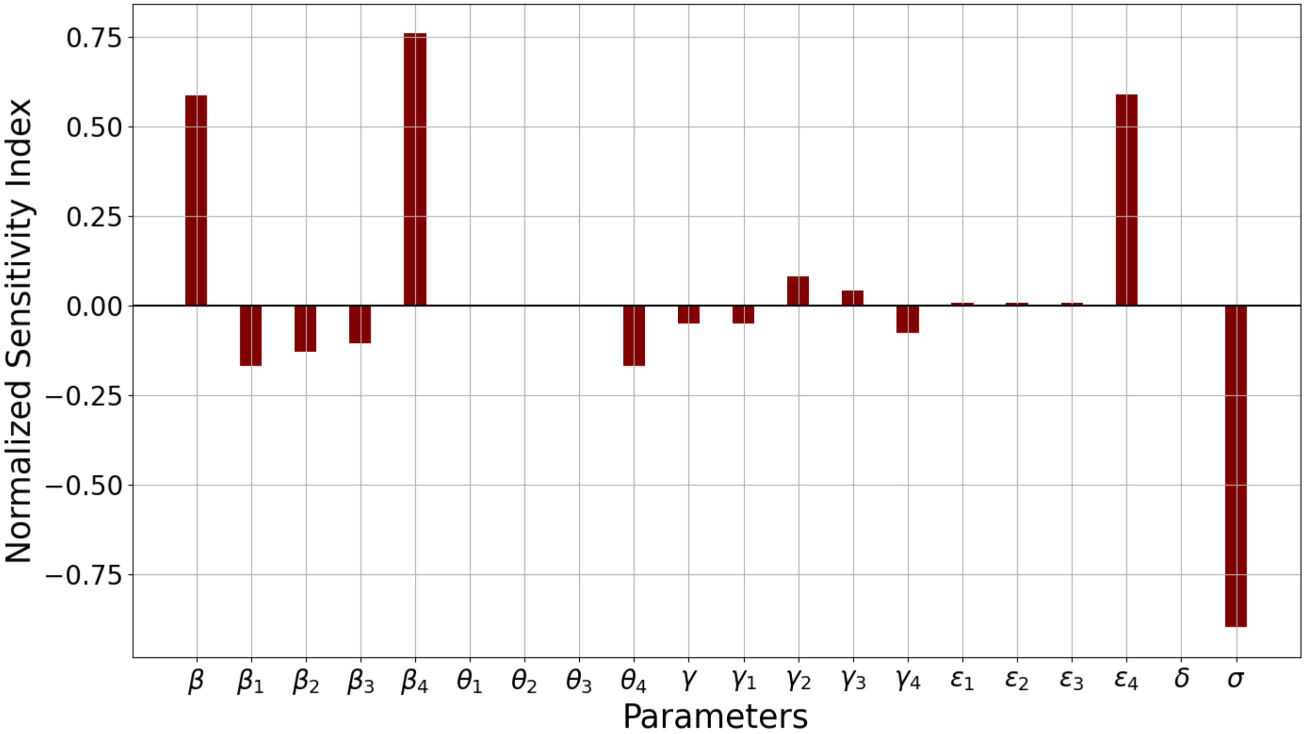


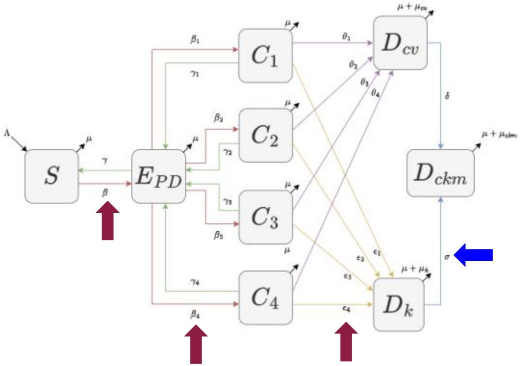
$$\boxed{\text{eGFR} = 142 \times \min(\text{SCr}/\kappa, 1)^{\alpha} \times \max(\text{SCr}/\kappa, 1)^{-1.200} \times 0.9938^{\text{Age}} \times 1.012 \text{ [if female]}}$$

Abbreviations/units

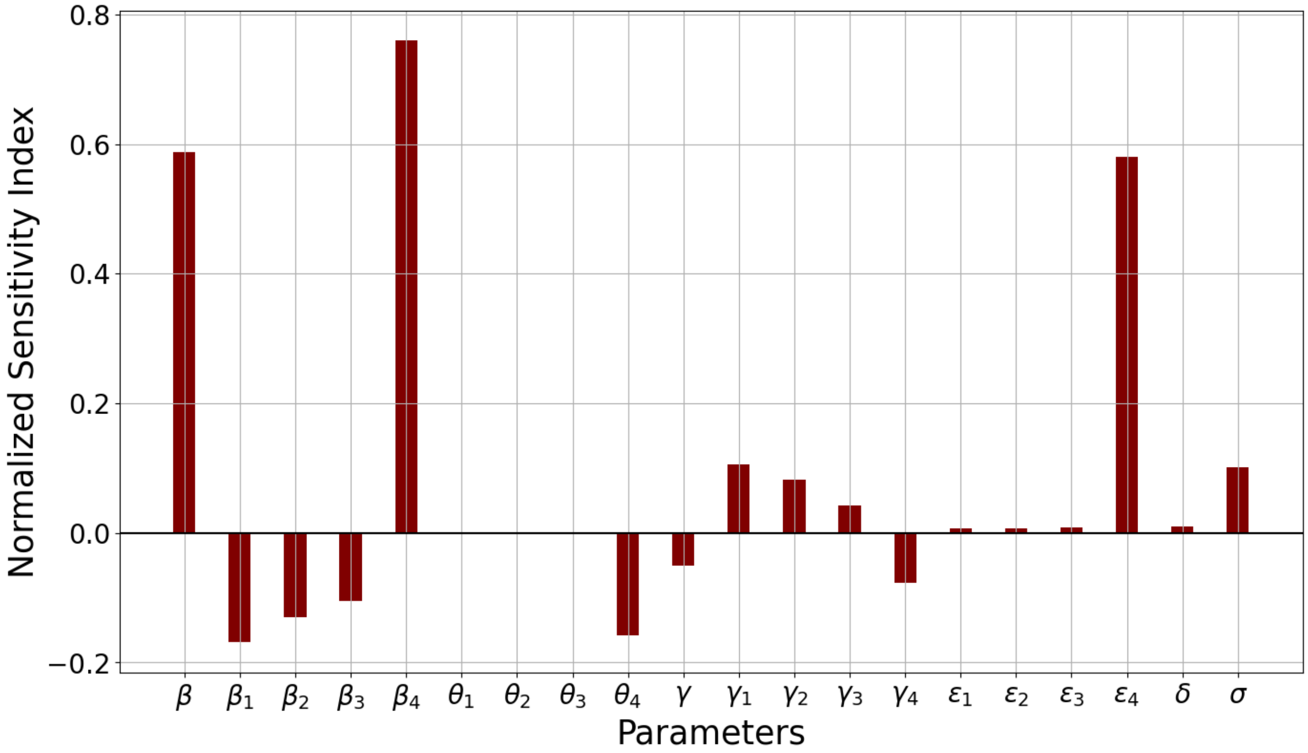
- eGFR = estimated GFR in mL/min/1.73 m^2
- SCr = standardized serum creatinine in mg/dL
- κ = 0.7 (females) or 0.9 (males)
- α = -0.241 (females) or -0.302 (males)
- min = indicates the minimum of SCr/κ or 1
- max = indicates the maximum of SCr/κ or 1
- age = years

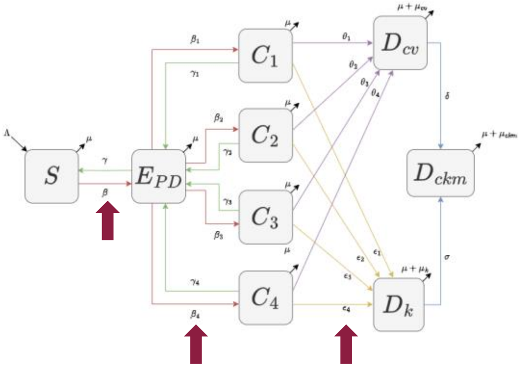
Sensitivity Analysis for D^*_K

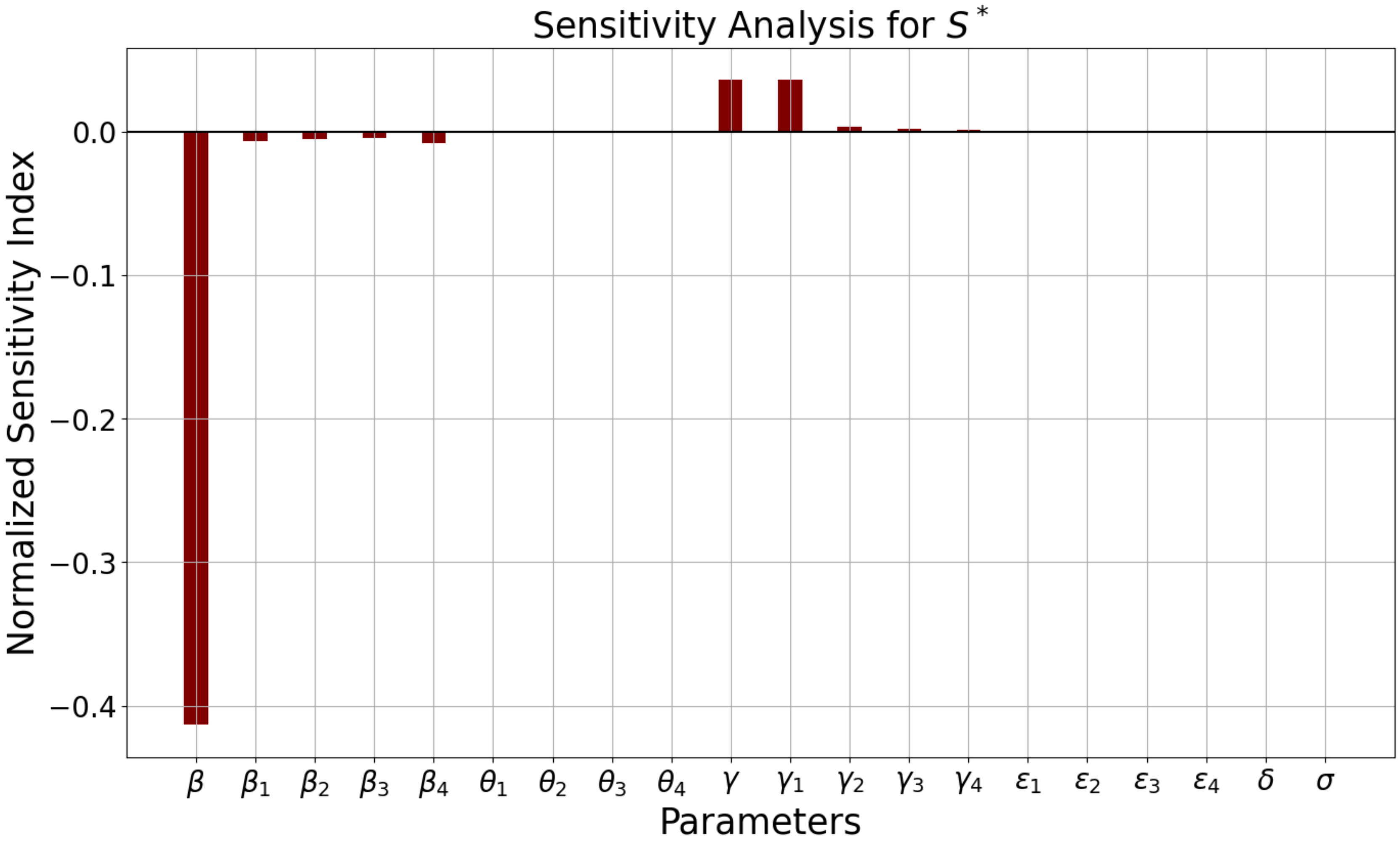


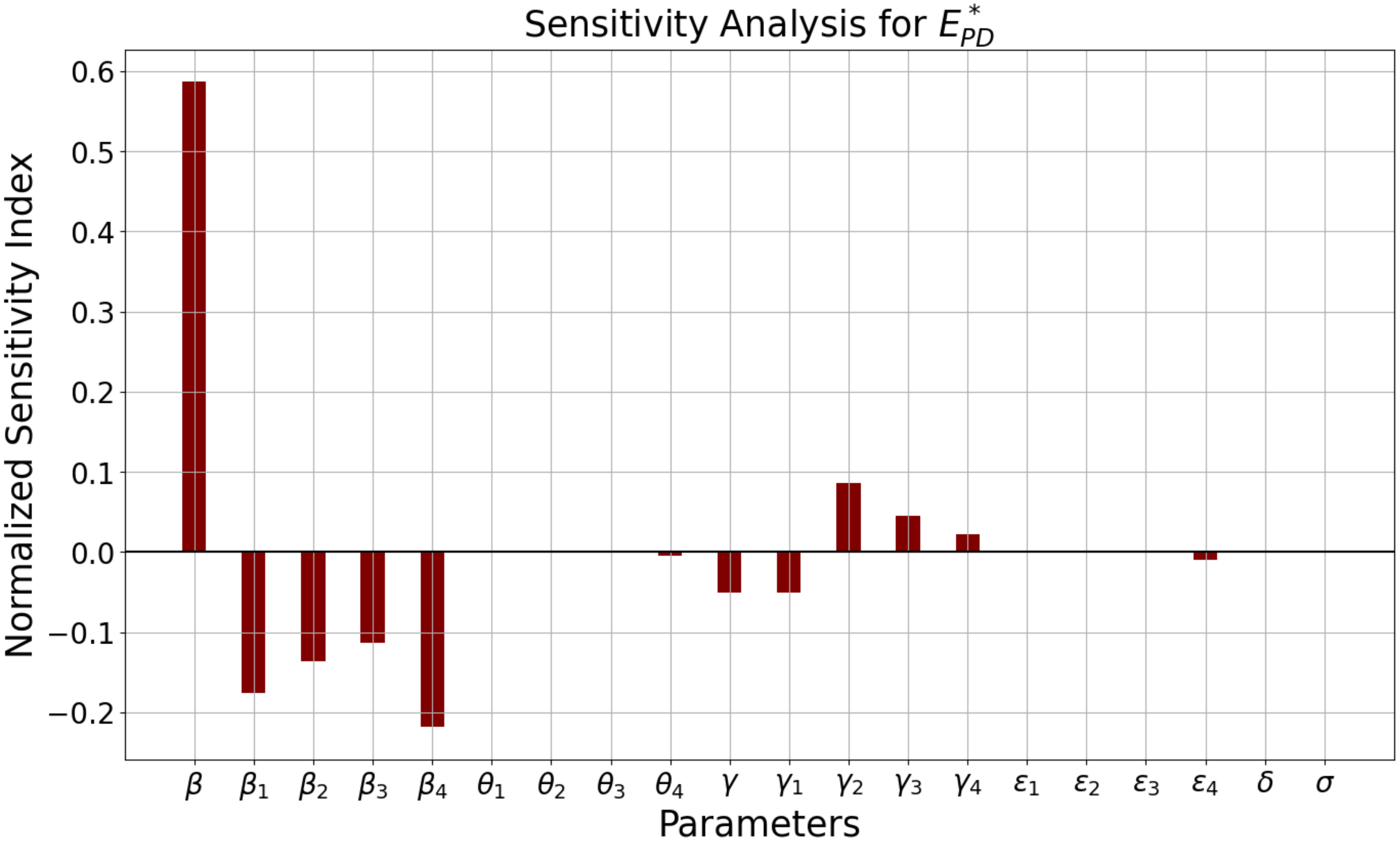


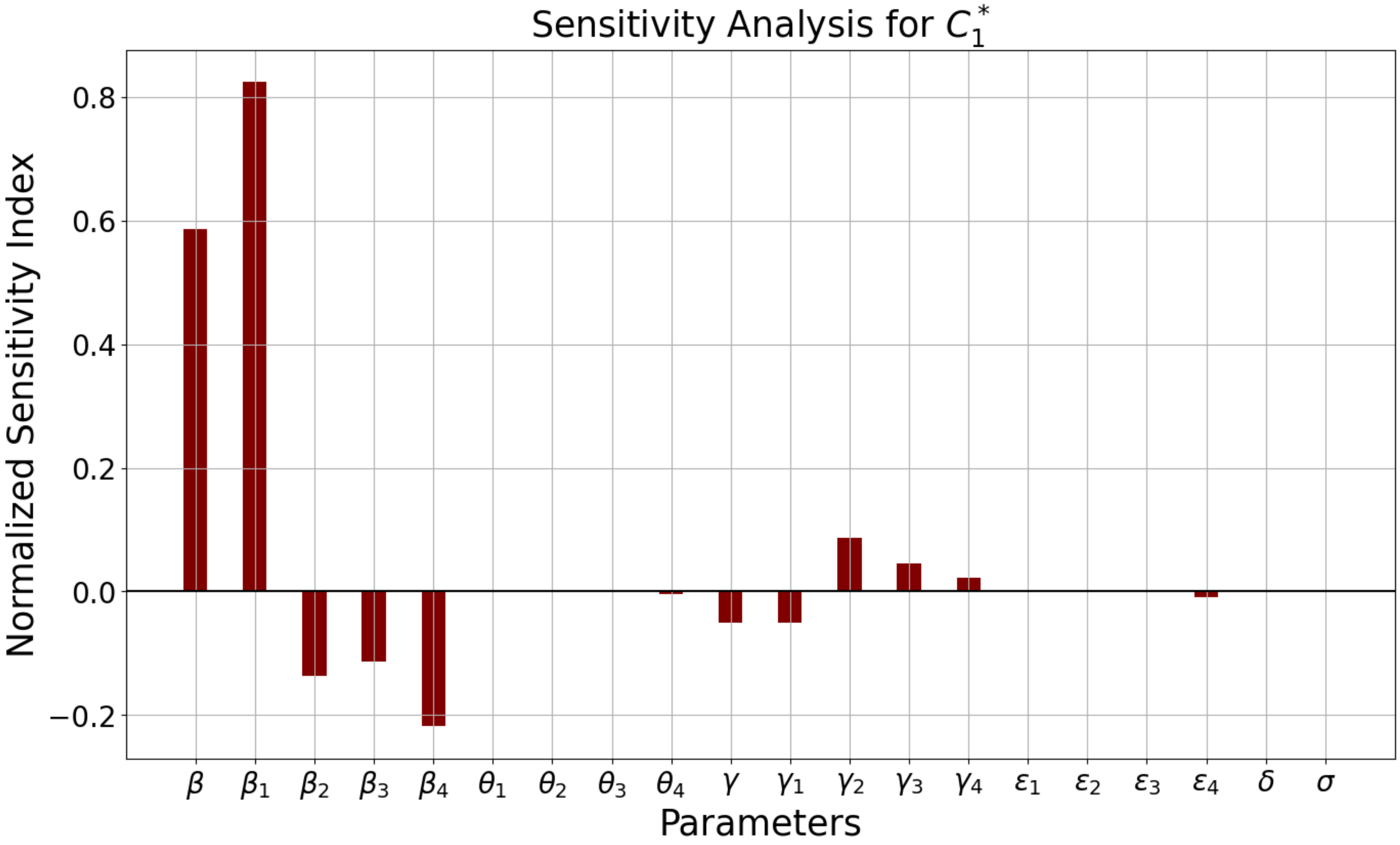
Sensitivity Analysis for D*$_{CKM}$



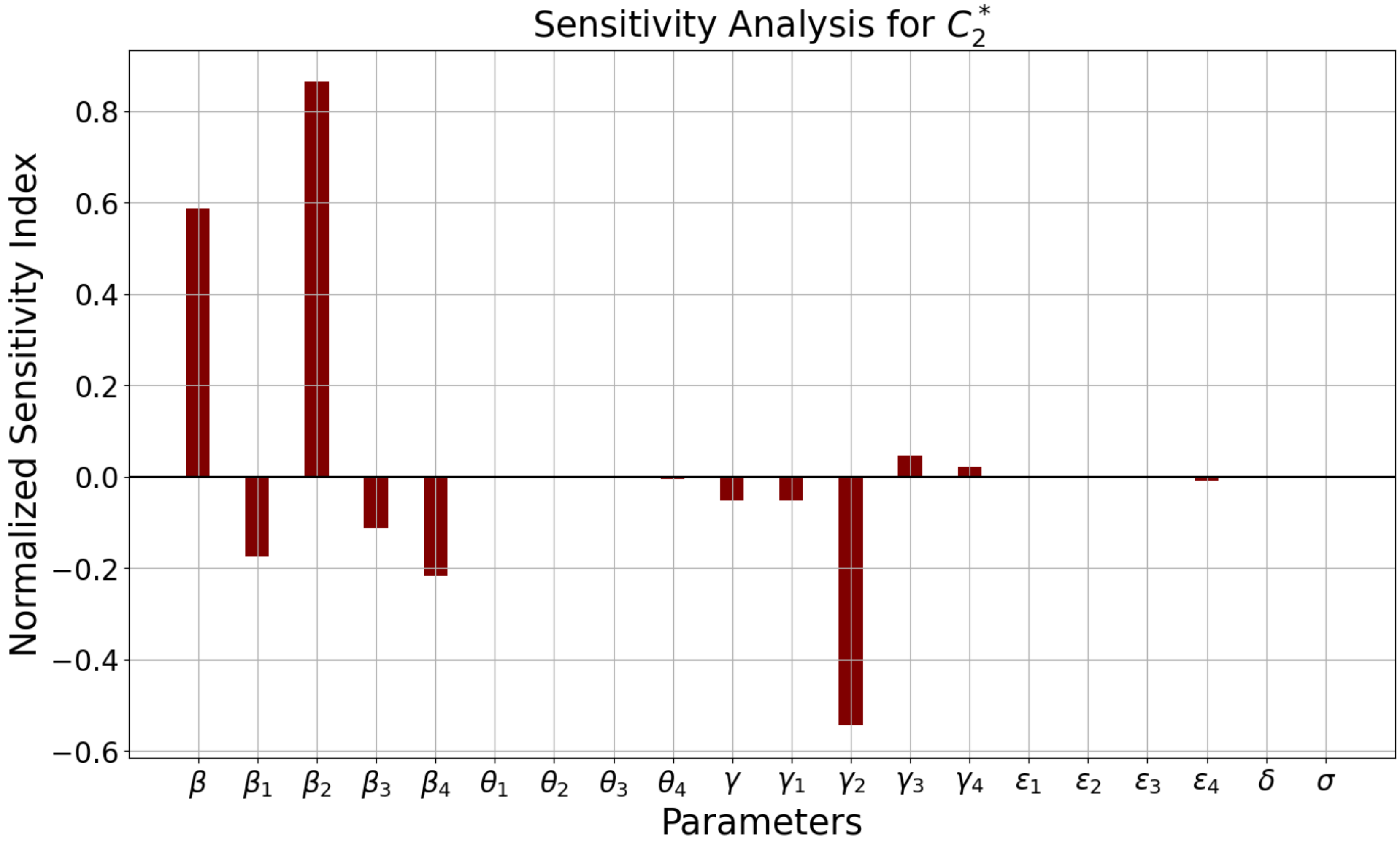


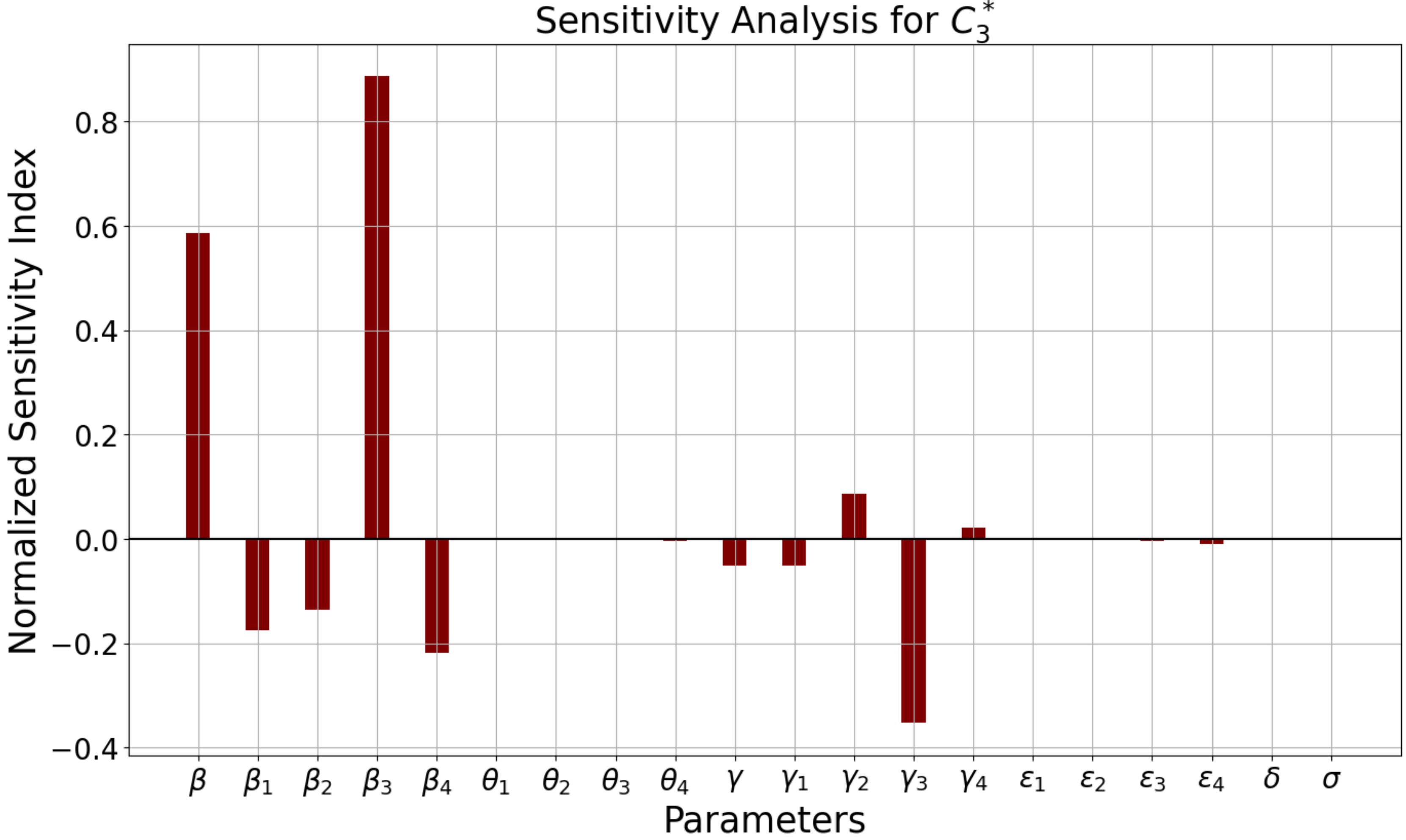

Sensitivity Analysis for S^*
Normalized Sensitivity Index
0.0
−0.1
−0.2
−0.3
−0.4
β β_1 β_2 β_3 β_4 θ_1 θ_2 θ_3 θ_4 γ γ_1 γ_2 γ_3 γ_4 ε_1 ε_2 ε_3 ε_4 δ σ
Parameters


Sensitivity Analysis for E_{PD}^*
Normalized Sensitivity Index
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
β β_1 β_2 β_3 β_4 θ_1 θ_2 θ_3 θ_4 γ γ_1 γ_2 γ_3 γ_4 ε_1 ε_2 ε_3 ε_4 δ σ
Parameters

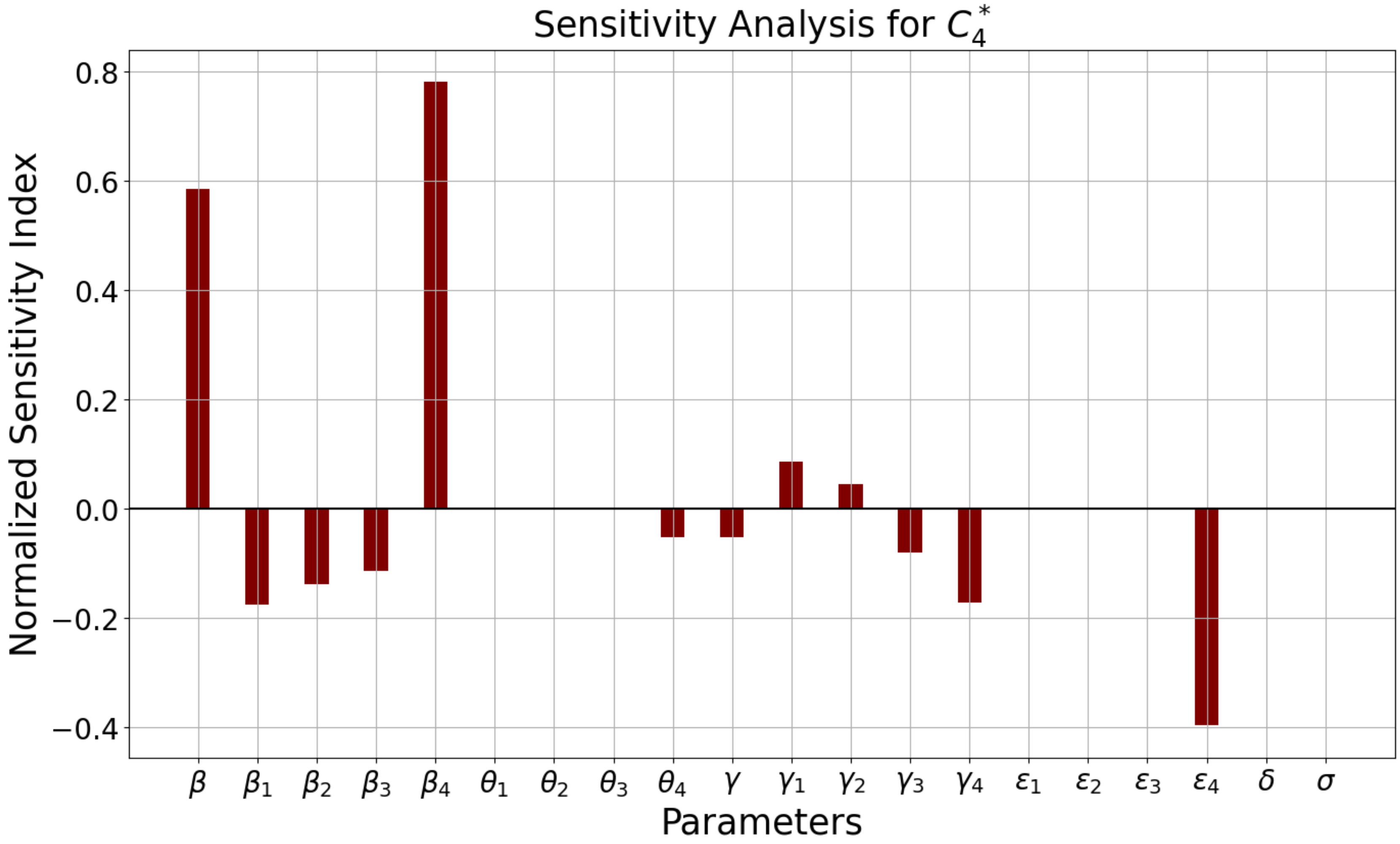


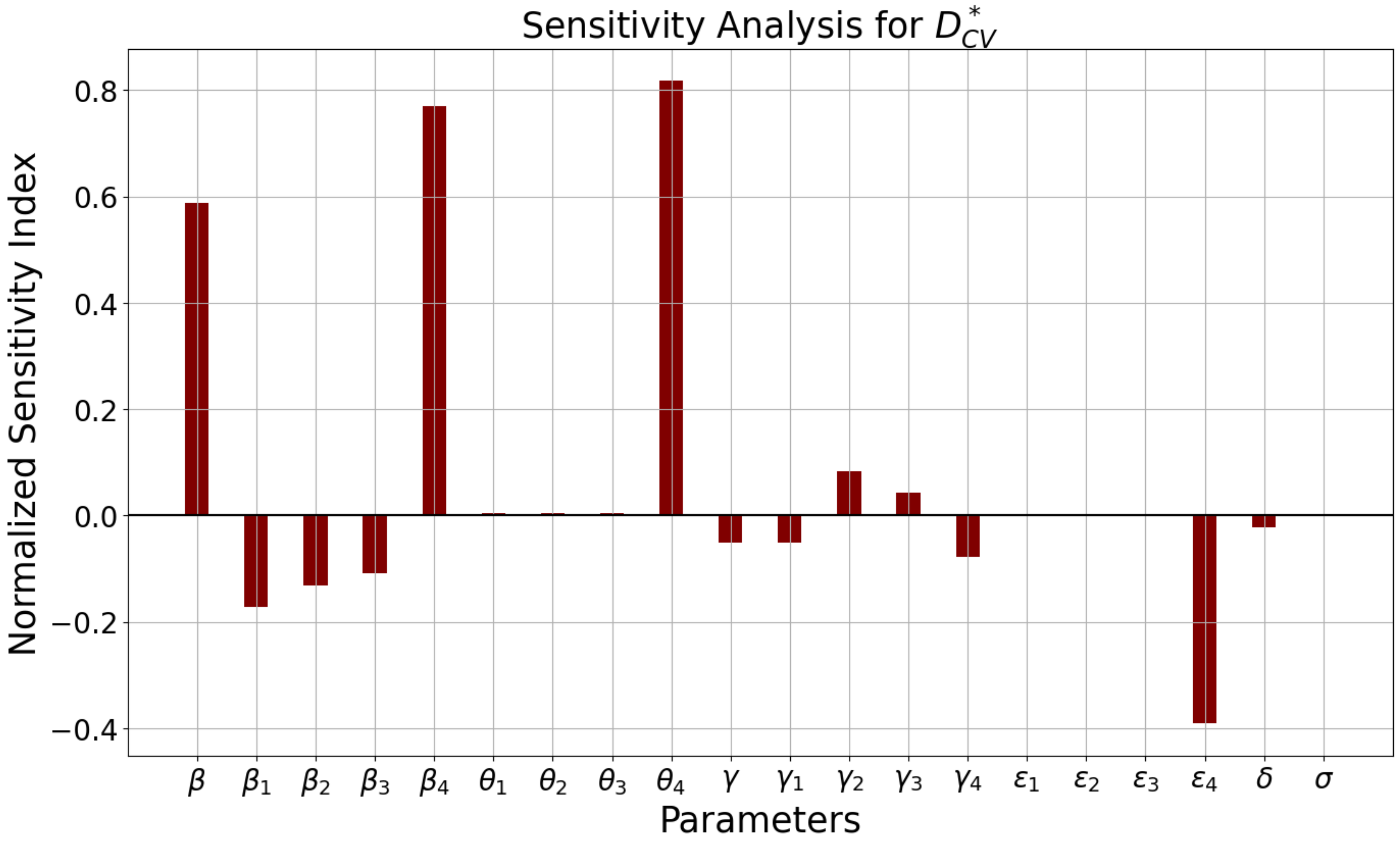
Sensitivity Analysis for C_1^*
Normalized Sensitivity Index
0.8
0.6
0.4
0.2
0.0
−0.2
β β_1 β_2 β_3 β_4 θ_1 θ_2 θ_3 θ_4 γ γ_1 γ_2 γ_3 γ_4 ε_1 ε_2 ε_3 ε_4 δ σ
Parameters


Sensitivity Analysis for C_2^*
Normalized Sensitivity Index
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
β β_1 β_2 β_3 β_4 θ_1 θ_2 θ_3 θ_4 γ γ_1 γ_2 γ_3 γ_4 ε_1 ε_2 ε_3 ε_4 δ σ
Parameters

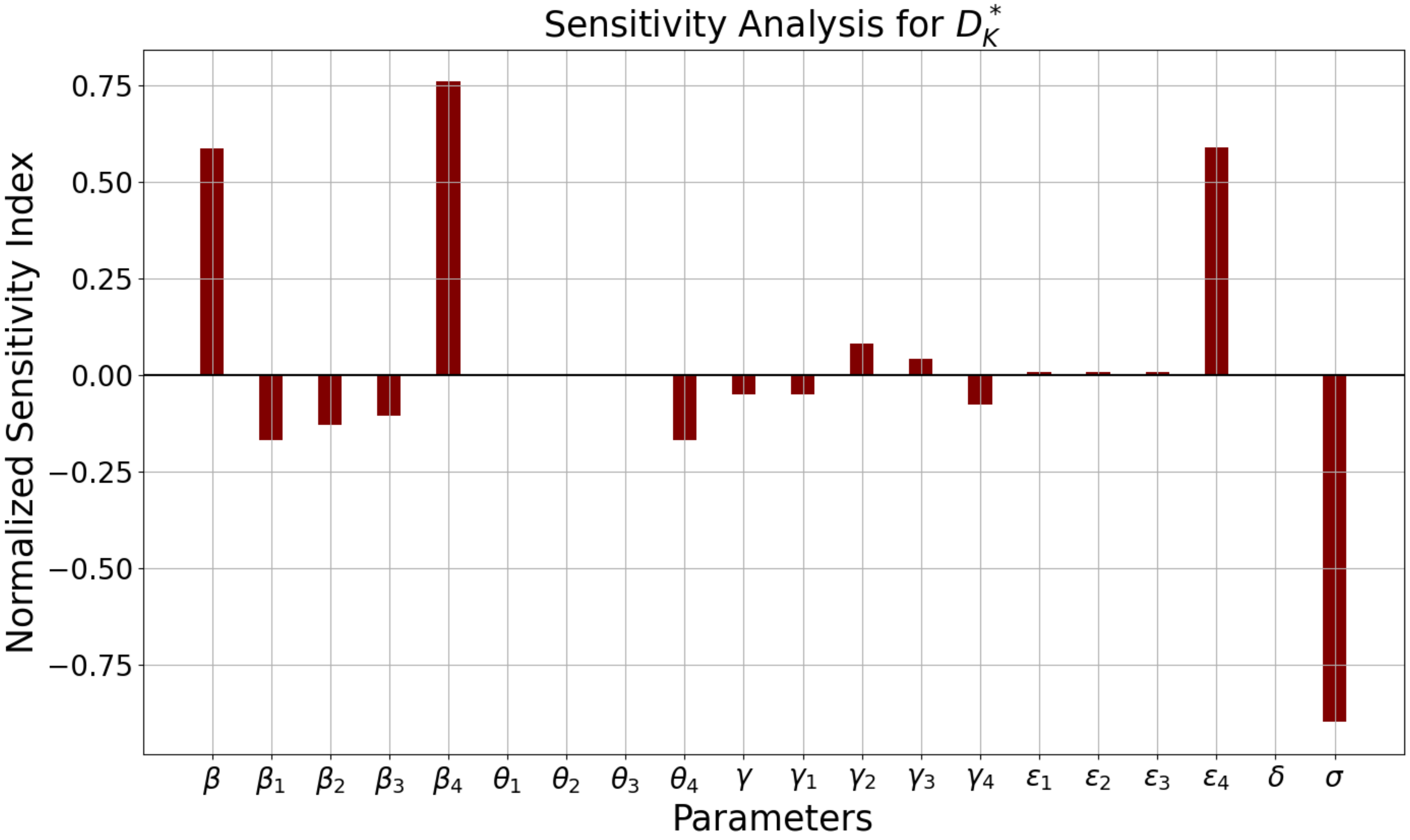


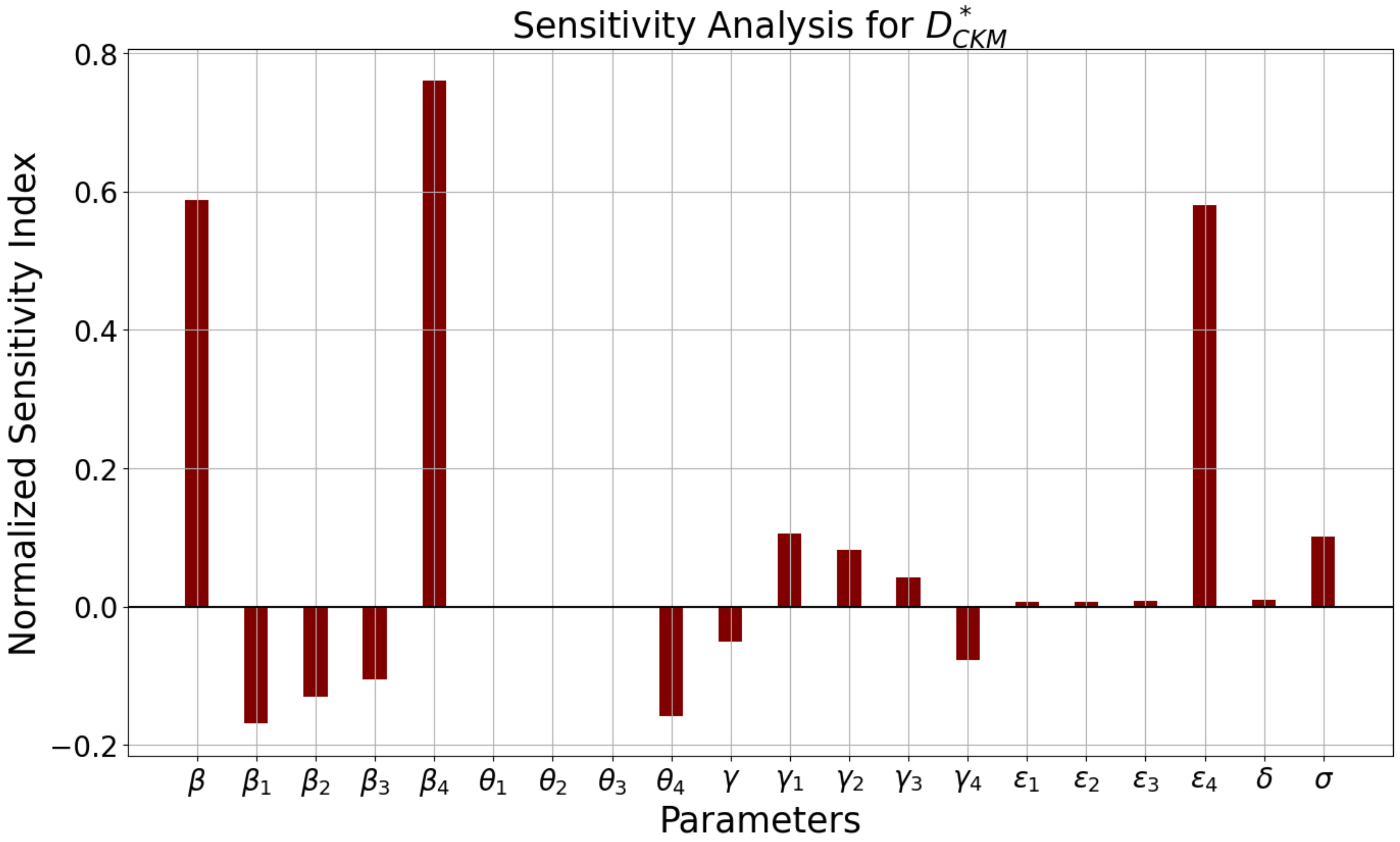
Sensitivity Analysis for C_3^*
Normalized Sensitivity Index
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
β β_1 β_2 β_3 β_4 θ_1 θ_2 θ_3 θ_4 γ γ_1 γ_2 γ_3 γ_4 ε_1 ε_2 ε_3 ε_4 δ σ
Parameters


Sensitivity Analysis for C_4^*
Normalized Sensitivity Index
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
β β_1 β_2 β_3 β_4 θ_1 θ_2 θ_3 θ_4 γ γ_1 γ_2 γ_3 γ_4 ε_1 ε_2 ε_3 ε_4 δ σ
Parameters



Sensitivity Analysis for D_{CV}^*
Normalized Sensitivity Index
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
β β_1 β_2 β_3 β_4 θ_1 θ_2 θ_3 θ_4 γ γ_1 γ_2 γ_3 γ_4 ε_1 ε_2 ε_3 ε_4 δ σ
Parameters


Sensitivity Analysis for D_K^*
Normalized Sensitivity Index
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
β β_1 β_2 β_3 β_4 θ_1 θ_2 θ_3 θ_4 γ γ_1 γ_2 γ_3 γ_4 ε_1 ε_2 ε_3 ε_4 δ σ
Parameters



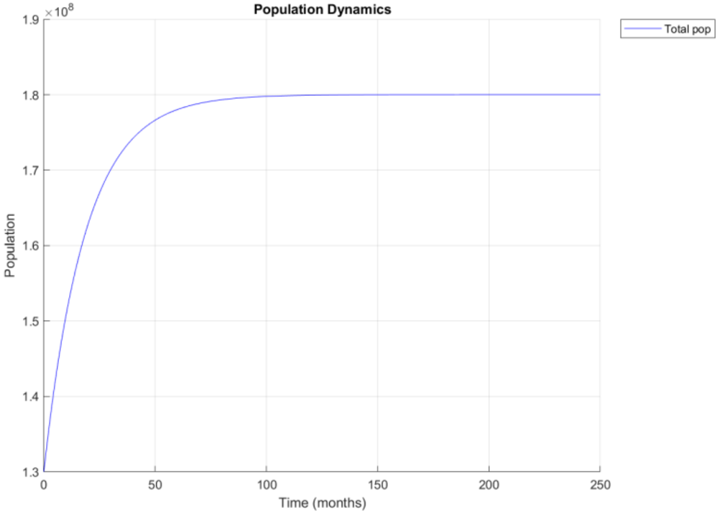
Sensitivity Analysis for D^*_{CKM}
Normalized Sensitivity Index
0.8
0.6
0.4
0.2
0.0
−0.2
β β_1 β_2 β_3 β_4 θ_1 θ_2 θ_3 θ_4 γ γ_1 γ_2 γ_3 γ_4 ε_1 ε_2 ε_3 ε_4 δ σ
Parameters

Preliminary results

Boundedness

From the assumption that μ_{cv}, μ_k, and μ_{ckm} are minimized, then $\frac{dN}{dt} = \Lambda - \mu N$. Thus,

$$\limsup_{t\to\infty} N(t) = \frac{\Lambda}{\mu}$$



[redacted]

Preliminary results

Positive Invariance

Starting at

$$\frac{d}{dt}S(t) = \Lambda + \gamma E_{pd}(t) - \beta S(t) - \mu S(t)$$

by integrating by integration factor, we get

$$\Psi(t)S(t) - \Psi(0)S(0) = \int_0^t \Psi(t')(\Lambda + \gamma E_{pd}(t')dt'$$

Then,

$$S(t) = \Psi(t)^{-1}\left(\int_0^t \Psi(t)(\Lambda + \gamma E_{pd}(t))dt + S(0)\right) > 0$$

where

$$\Psi(t) = e^{(\beta+\mu)t}$$

Preliminary results

Equilibrium Point

$$J(S^*, E^*_{PD}, C^*_i, D^*_{CV}, K^*, C^*_{KM}) = \begin{bmatrix} -\beta-\mu & \gamma & 0 & 0 & 0 & 0 \\ \beta & -\beta_i-\gamma-\mu & \gamma_i & 0 & 0 & 0 \\ 0 & \beta_i & -\gamma_i-\epsilon_i-\theta_i-\mu & 0 & 0 & 0 \\ 0 & 0 & \theta_i & -\delta-\mu-\mu_{CV} & 0 & 0 \\ 0 & 0 & \epsilon_i & 0 & -\mu-\mu_K-\sigma & 0 \\ 0 & 0 & 0 & \delta & \sigma & -\mu-\mu_{CKM} \end{bmatrix}. \tag{3}$$

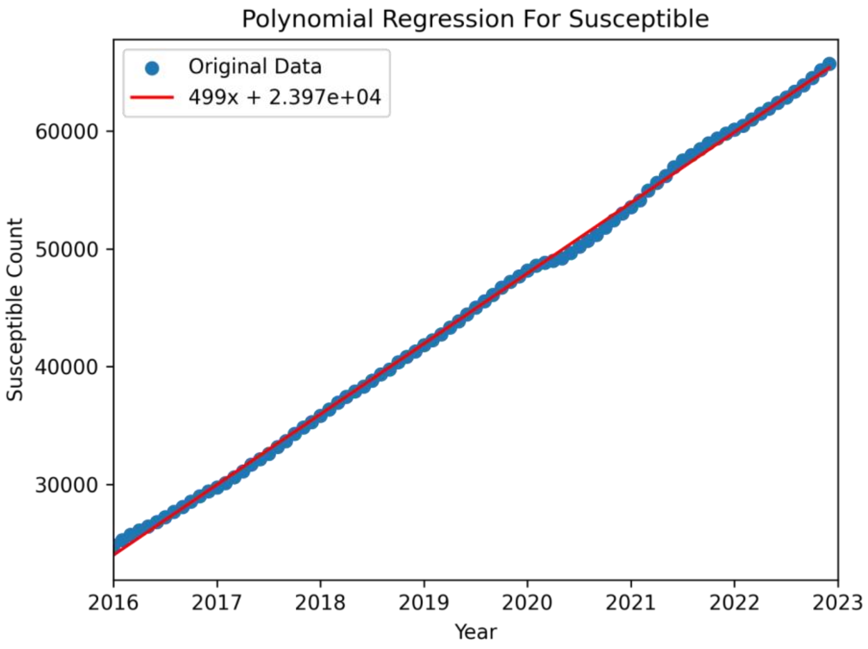
Preliminary results

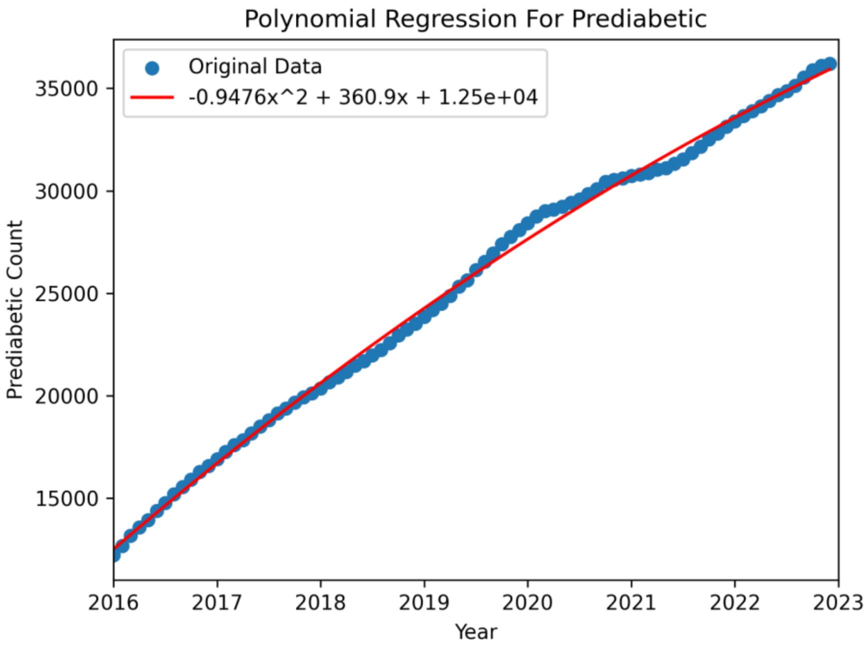
Equilibrium Point

$$J^* = \begin{bmatrix} -\beta - \mu & \gamma & 0 \\ \beta & -\beta_i - \gamma - \mu & \gamma_i \\ 0 & \beta_i & -\gamma_i - \epsilon_i - \theta_i - \mu \end{bmatrix}$$

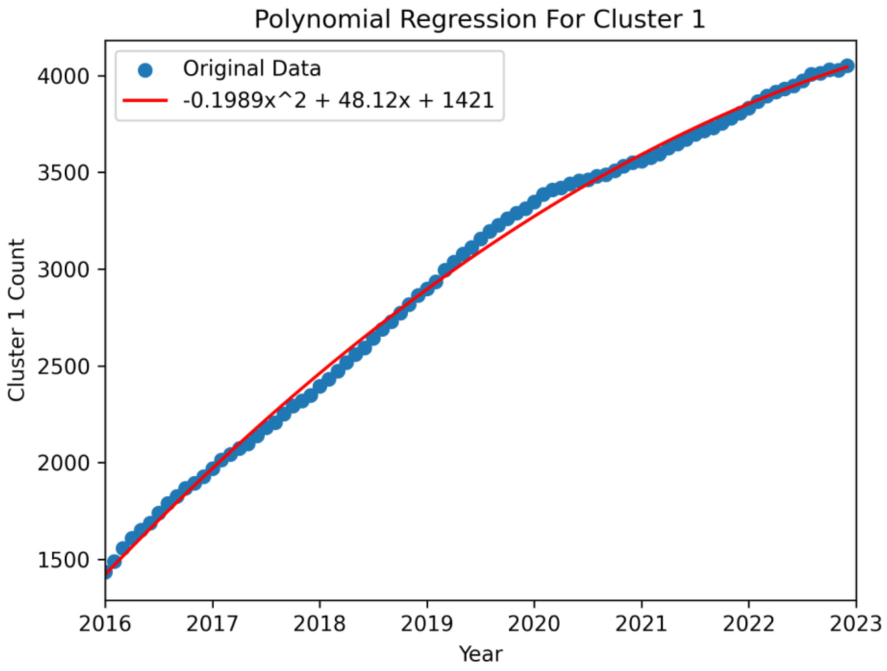
- Stability proven by Routh-Hurwitz theorem.

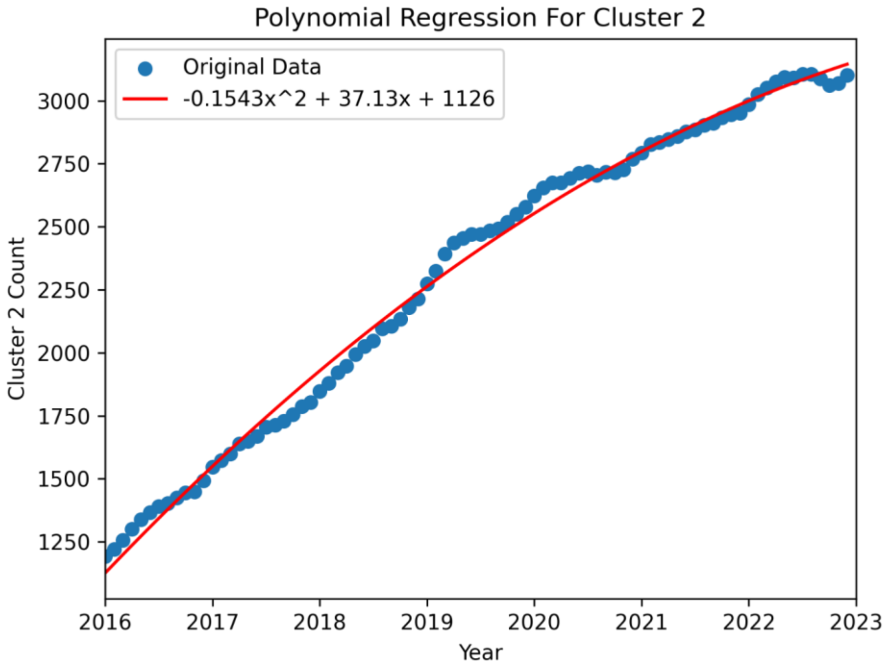
Polynomial Graphs



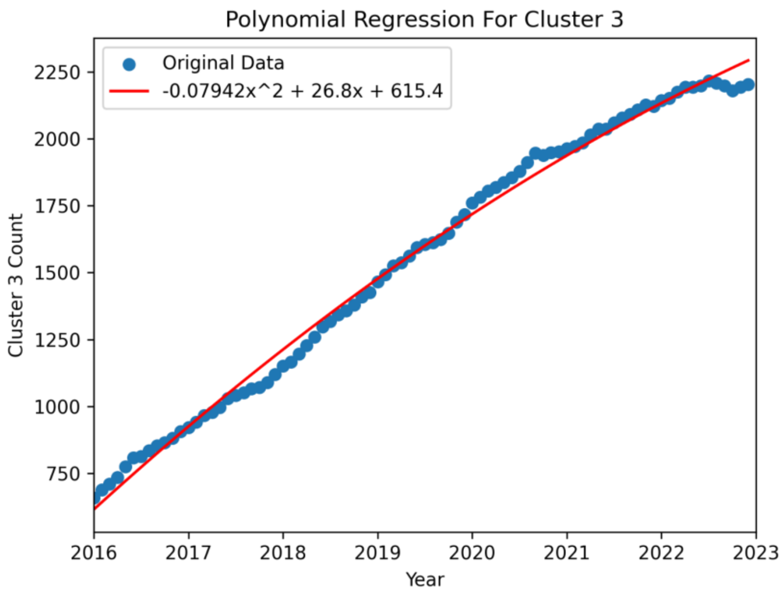


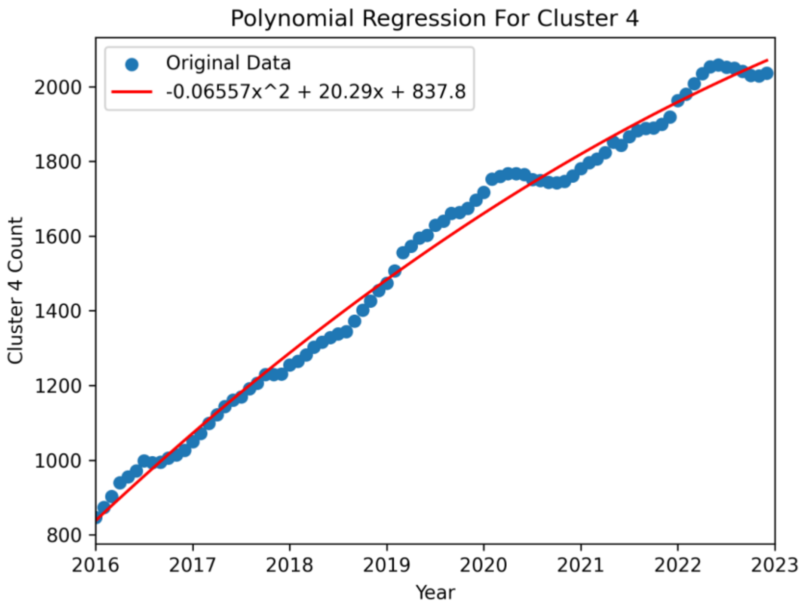
Polynomial Graphs



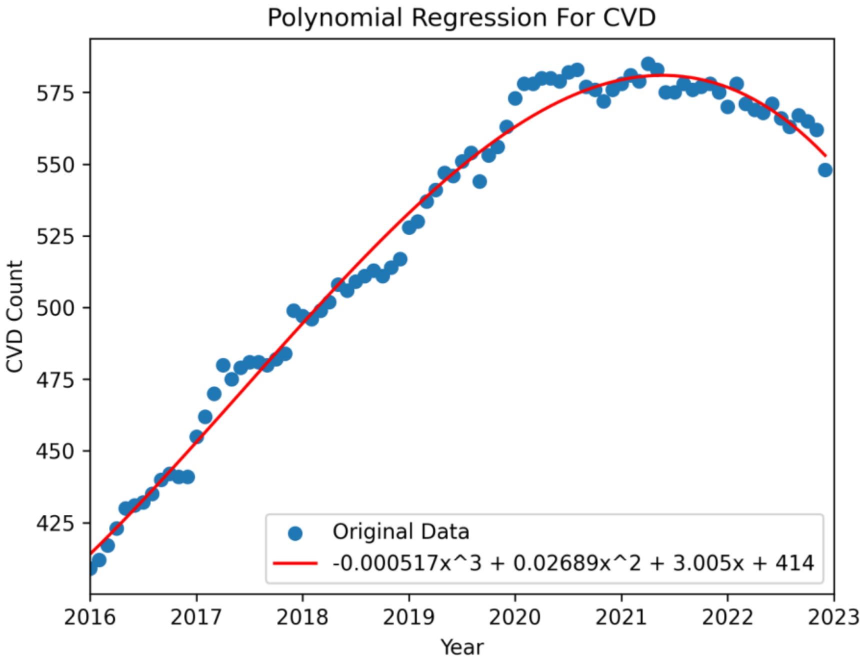


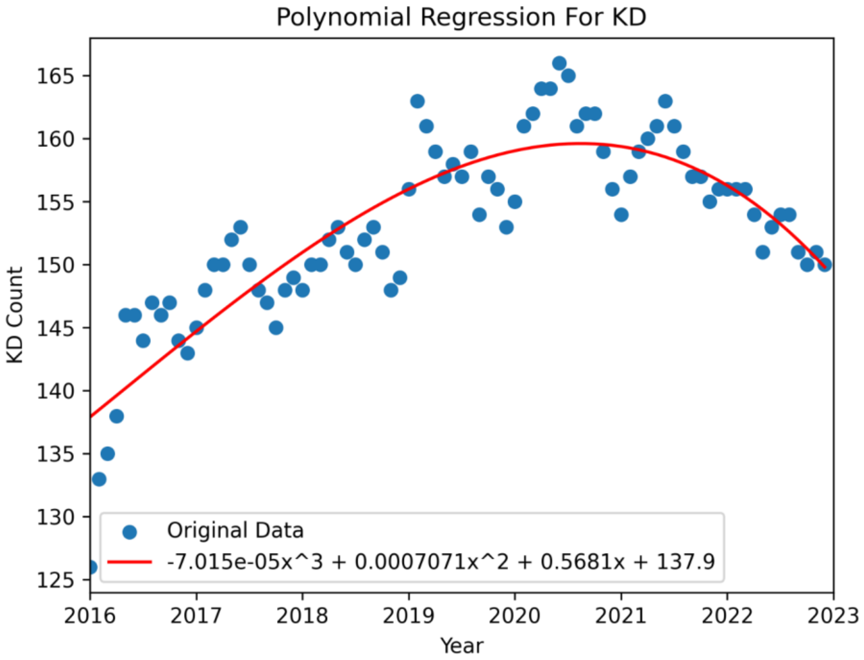
Polynomial Graphs



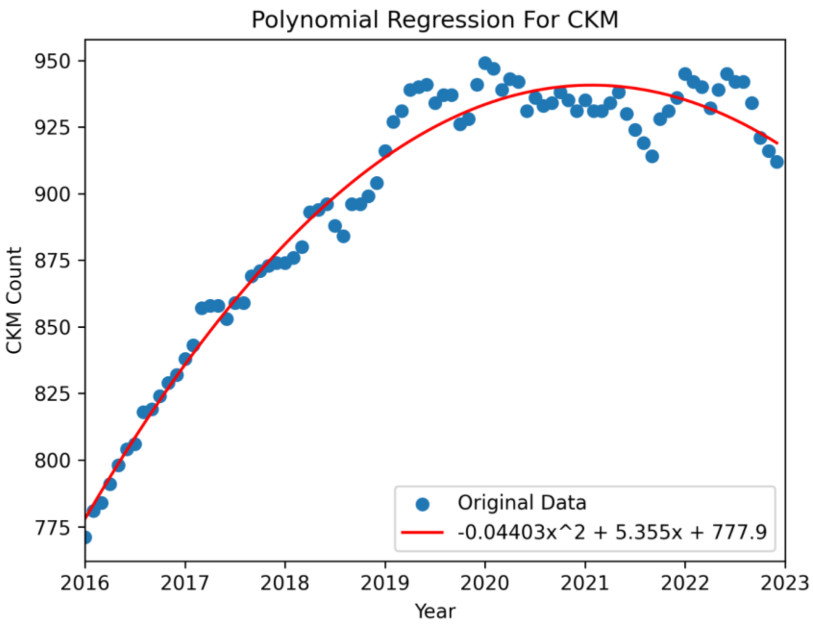


Polynomial Graphs





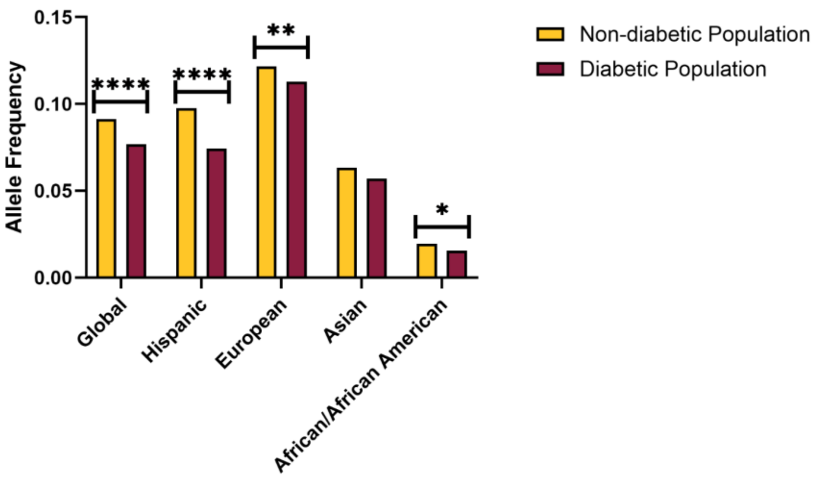
Polynomial Graphs



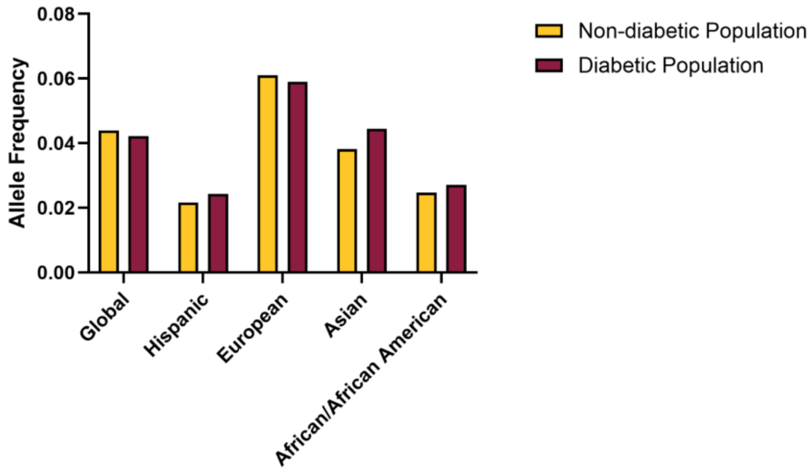
Allele Frequencies

PPARG

PPARG rs1801282 Allele Frequency by Race/Ethnicity



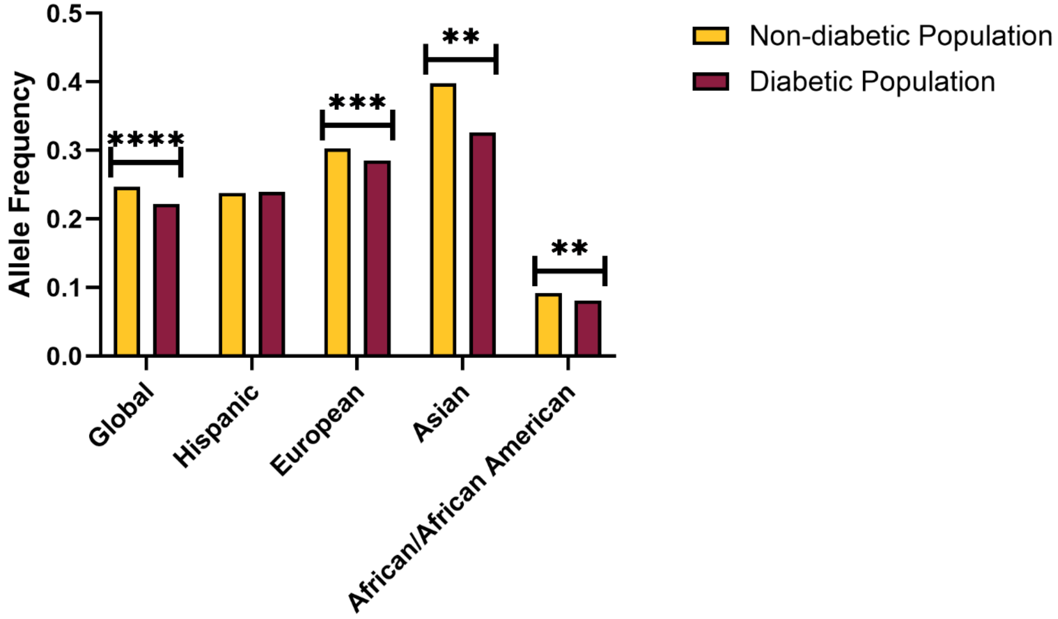
PPARG rs17036101 Allele Frequency by Race/Ethnicity



Allele Frequencies

SLC30A8

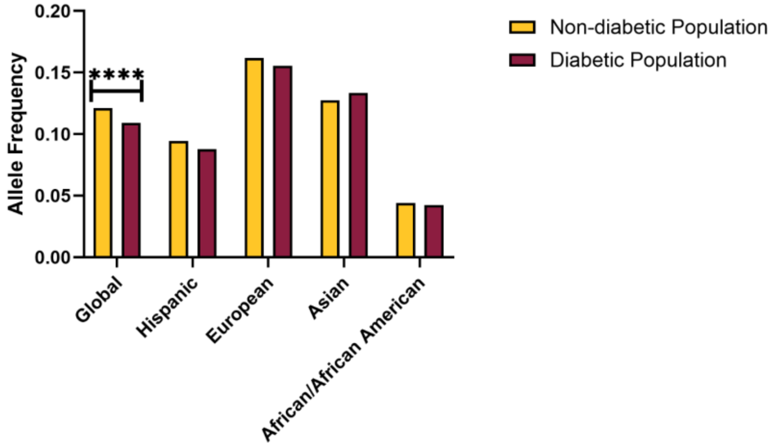
SLC30A8 rs13266634 Allele Frequency by Race/Ethnicity



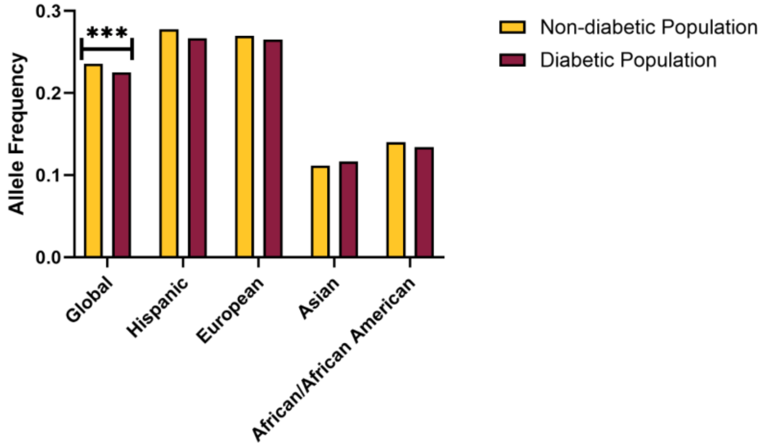
Allele Frequencies

CAPN10

CAPN10 rs2975760 Allele Frequency by Race/Ethnicity



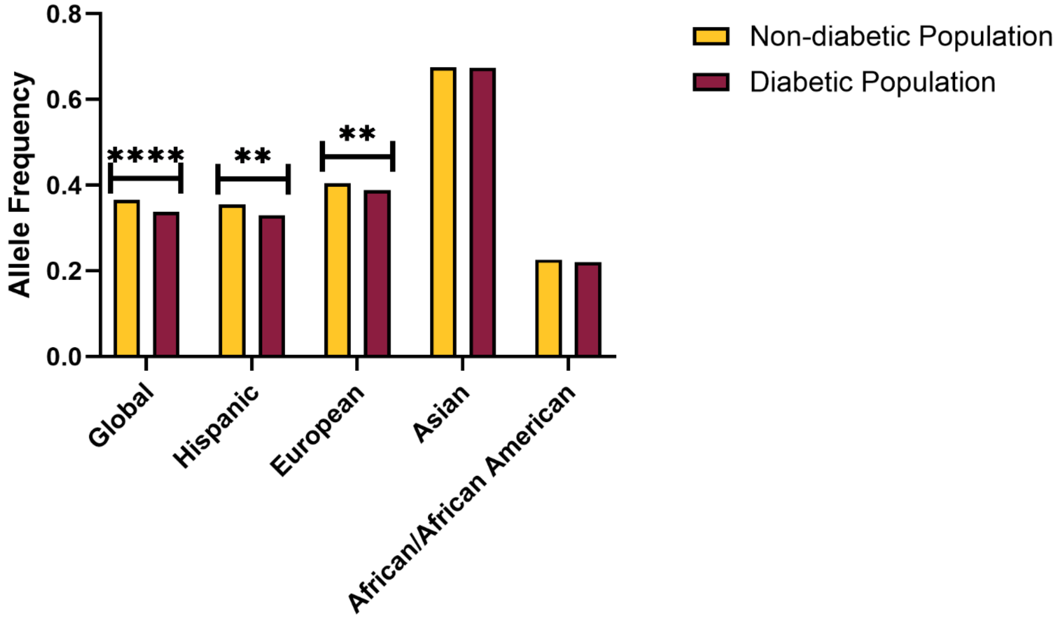
CAPN10 rs3792267 Allele Frequency by Race/Ethnicity



Allele Frequencies

HHEX/IDE

HHEX/IDE rs1111875 Allele Frequency by Race/Ethnicity



Relative Error

Per Cluster

	error_c1	error_c2	error_c3	error_c4
0	0.036609	0.051199	0.074788	0.000433
1	0.036478	0.050548	0.073827	0.008182
2	0.037020	0.048663	0.073920	0.019918
3	0.036751	0.045114	0.073013	0.021411
4	0.035935	0.050381	0.070196	0.022008
5	0.040320	0.056263	0.077616	0.008893
6	0.044905	0.071182	0.099879	0.003380
7	0.050198	0.078415	0.117229	0.006519
8	0.055528	0.083127	0.133252	0.012917