# SDoH-Aware Approach to Prostate Cancer Screening: Addressing Overdiagnosis of Prostate Cancer using PSA

Ashley Lewis[† 1], Yash Samir Khandwala, MD[2], Tina Hernandez-Boussard, PhD[1], James D Brooks, MD[2]

[1]*Department of Biomedical Data Science, Stanford University, CA, USA*

[2]*Department of Urology, Stanford University, CA, USA*

*Email:* alewis23@stanford.edu, yashk@live.com, boussard@stanford.edu, jbrooks1@stanford.edu

This study investigates the potential of multimodal data for prostate cancer (PCa) risk prediction using the All of Us (AoU) research program dataset. By integrating polygenic risk scores (PRSs) with diverse clinical, survey, and genomic data, we developed a model that identifies established PCa risk factors, such as age and family history, and a novel factor: recent healthcare visits are linked to reduced risk. The model's performance, notably the false positive rate, is improved compared to traditional methods, despite the lack of Prostate-Specific Antigen (PSA) data. The findings demonstrate that incorporating comprehensive multimodal data from AoU can enhance PCa risk prediction and provide a robust framework for future clinical applications.

Code Available: https://github.com/ashlew23/pc_multimodal

*Keywords:* prostate cancer; multimodal data; false-positive results; social determinants of health.

© 2023 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

## 1. Introduction

Prostate cancer (PCa) is a complex and multifaceted disease characterized by significant heterogeneity and disparities among individuals.[1] To address this, novel methods that use a precision-medicine approach to improve traditional prostate-specific antigen (PSA) testing are needed to reduce overdiagnosis and identify cases that might be missed by conventional biological markers.[2] Integrating genetic information with clinical and lifestyle data presents a promising avenue for advancing disease prediction and precision medicine.

Polygenic risk scores (PRSs) quantify genetic predisposition by analyzing multiple genetic variants and have emerged as a promising tool for prostate cancer screening. However, their use can sometimes lead to overdiagnosis, and they may not perform as well as validated prostate cancer biomarkers.[3] However, cross-disciplinary research efforts have shown that integrating data of multiple modalities can significantly improve the predictive power of models.[4] A similar approach is possible for PCa screening models as we know that PRSs can be a powerful screening method when combined with clinical variables. We aim to extend this inquiry to other data sources that have not previously been used for this purpose.[5] The All of Us (AoU) research program offers a new opportunity to evaluate this hypothesis, providing access to both multimodal data and a large, diverse research population. Incorporating comprehensive survey data that capture social determinants of health (SDoH) and clinical history allows scores developed using AoU data to provide a holistic view of disease risk. However, the limited availability of PSA data in this dataset is a notable limitation, highlighting the potential for significant improvement when such data becomes accessible. Still, this framework can be applied to other disease types to inform SDoH-aware screening practices. Understanding the relative importance of each data type in predicting disease incidence can help refine predictive models and identify key risk factors.

## 2. Related Work

### 2.1. *Modern Methods for Prostate Cancer Incident Prediction*

PSA-only based screening has significantly contributed to the overdiagnosis of PCa, leading to controversy regarding its utility as a biomarker. A 2018 JAMA study by the US Preventive Services Task Force highlighted that periodic PSA-based screening could result in false positives,

---

[1] Mateo et al., "Accelerating Precision Medicine in Metastatic Prostate Cancer."
[2] Cuzick et al., "Prevention and Early Detection of Prostate Cancer."
[3] Schaffer et al., "A Polygenic Risk Score for Prostate Cancer Risk Prediction."
[4] Mohsen et al., "Artificial Intelligence-Based Methods for Fusion of Electronic Health Records and Imaging Data."
[5] Van Poppel et al., "Serum PSA-Based Early Detection of Prostate Cancer in Europe and Globally."

© 2023 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

necessitating further screening and potential biopsies, which carry inherent risks related to surgical treatment.[6] Beyond the immediate physical implications, overdiagnosis has been associated with both short-term and long-term anxiety.[7] These concerns underscore the need for careful and effective use of PSA in predictive models.

The integration of additional relevant data has demonstrated potential in reducing the false-positive rates associated with PSA screening through the development of modern predictive tools. Established risk calculators, such as the Prostate Cancer Prevention Trial (PCPT), combine PSA levels with other clinical factors and demographic information to estimate prostate cancer risk more accurately. Recent studies have enhanced these models by incorporating genetic information in the form of PRSs. PRSs have shown promise in predicting both the risk of PCa incidence and the likelihood of mortality from the disease.[8] However, the utility of PRSs can be influenced by factors such as the selection of single nucleotide polymorphisms (SNPs) used in their development, the diversity of the genome-wide association study (GWAS) cohort, and the generalizability to the applied cohort.

Recent research has suggested that while PRSs for prostate cancer are associated with incident cancer and prostate cancer mortality, they do not enhance the prediction of aggressive cancers or outperform PSA testing.[9] It is important to note that these studies do not take into account any social factors that could enhance the accuracy and clinical utility of PRS-based assessments for prostate cancer.

### 2.2. *Importance of Understanding Social Determinants for PCa Outcomes*

While genetic differences can contribute to the development of certain diseases, there is a growing body of evidence suggesting that socioeconomic factors play a crucial role in disease outcomes. An analysis of temporal trends and key drivers of inequality in life expectancy revealed that behavioral and metabolic risk factors accounted for 74% of the county-level variation in life expectancy.[10] Access to healthcare is a critical factor in accurately assessing disease outcomes, as demonstrated by two separate studies. In a Veterans Health Administration (VA) cohort study, Black men presented at a younger age with higher PSA levels. When treated in an equal-access setting, the disparity in prostate cancer risk was significantly reduced, with Black men exhibiting

---

[6] US Preventive Services Task Force, "Screening for Prostate Cancer."
[7] Tosteson et al., "Consequences of False-Positive Screening Mammograms"; Lin et al., "Benefits and Harms of Prostate-Specific Antigen Screening for Prostate Cancer."
[8] Klein et al., "Prostate Cancer Polygenic Risk Score and Prediction of Lethal Prostate Cancer."
[9] Schaffer et al., "A Polygenic Risk Score for Prostate Cancer Risk Prediction"; Klein et al., "Prostate Cancer Polygenic Risk Score and Prediction of Lethal Prostate Cancer."
[10] Dwyer-Lindgren et al., "Inequalities in Life Expectancy Among US Counties, 1980 to 2014."

an 11% lower risk of developing metastases compared to White men.[11] Another study corroborated these findings by observing similar trends between Black and White men in the Surveillance, Epidemiology, and End Results (SEER) national cancer registry and the VA.[12] These studies underscore the importance of addressing socioeconomic factors and ensuring equitable access to healthcare to reduce disparities in prostate cancer outcomes and improve overall health equity.

### 2.3. *Multimodal Improvements in Prediction Across Disease Types*

The success of multimodal predictions relies on the availability of high-quality data, the ability to integrate different data types effectively, and the application of advanced machine learning techniques to extract meaningful patterns. As such, there have been unique methodological improvements to prediction of disease through the use of multiple data types available for disease prediction. For instance, in breast cancer screening (which has also historically suffered from the issue of misdiagnosis), the use of multimodal multiview breast-ultrasound images proved to be highly effective in reducing false positives.[13] Even in the non-cancer space such as in the screening for stroke, deep learning methods that integrate video and audio data with greater sensitivity than traditional emergency room triage.[14] Many other examples exist across the disease spectrum.

## 3. Materials and Methods

### 3.1 *Data Preparation*

The cases and control groups were assembled using the v7 controlled access tier in AoU. Participants were classified as cases if they had prostate cancer, as indicated by the presence of relevant SNOMED or ICD9/ICD10 vocabulary codes. Additionally, these participants provided survey responses to the Basics, Lifestyle, Health Care Access & Utilization, Personal and Family Health History, and Social Determinants of Health surveys, as well as short-read whole genome sequencing data. Controls were identified as participants who had the same relevant data modalities but without a diagnosis of PCa. This comprehensive approach ensured that both cases and controls were well-characterized across multiple data sources. Beta coefficients for prostate

---

[11] Yamoah et al., "Racial and Ethnic Disparities in Prostate Cancer Outcomes in the Veterans Affairs Health Care System."
[12] Klebaner, Courtney, and Rose, "Effect of Healthcare System on Prostate Cancer-Specific Mortality in African American and Non-Hispanic White Men."
[13] Qian et al., "Prospective Assessment of Breast Cancer Risk from Multimodal Multiview Ultrasound Images via Clinically Applicable Deep Learning."
[14] Cai et al., "*DeepStroke*."

cancer susceptibility loci were obtained from the trans-ancestry GWAS conducted by Conti et al.[15] These coefficients were utilized as the foundational weights for the construction of the PRSs.

### 3.2 *PRS Development*

Using the acquired beta coefficients, we developed polygenic risk scores (PRS) for approximately 21,974 prostate cancer cases and 2,556 controls. This process involved calculating individual-level risk scores based on the sum of risk allele counts weighted by their respective beta coefficients. Weight files were converted to the GRCh38 reference assembly using LiftOver to ensure compatibility with the reference genome used in the AoU genomic dataset. Quality control checks such as for heritability, effect allele designation, and relatedness of the data were performed to remove unwanted samples. The ORs from 269 risk variants identified in the Conti et al study were used to calculate the raw PRSs. Raw PRSs were adjusted using residualization prior to being used in the modeling of PCa incidence.

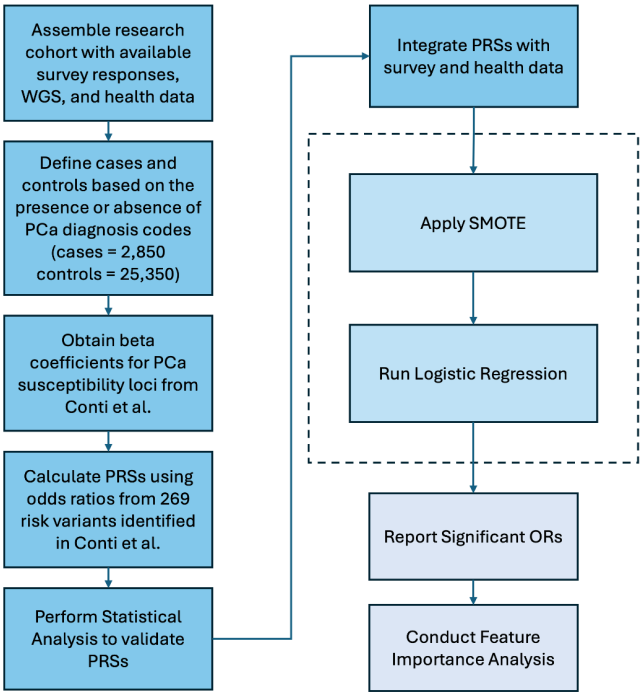


Figure 1. Data Processing and Model Workflow

### 3.3 *Data Integration and Processing*

Following the development of the PRSs, these scores were integrated with a comprehensive array of survey and health data to enhance predictive accuracy. We utilized surveys from the AoU,

[15] Conti et al., "Trans-Ancestry Genome-Wide Association Meta-Analysis of Prostate Cancer Identifies New Susceptibility Loci and Informs Genetic Risk Prediction."

including The Basics, Overall Health, Lifestyle, Personal and Family Health History, and Health Care Access and Utilization. In total, 11 questions from these surveys were employed to inform the relevant categories (Table 1). Additionally, age, zip code, and socioeconomic variables were derived from the relevant tables provided by AoU. Body mass index (BMI) was calculated for all participants based on the physical measurements data collected through AoU and included as a biometric in the analysis. Furthermore, the binary presence of comorbidities—such as hypertension, diabetes, stroke, and high cholesterol—was extracted using ICD9/10 source concept codes. The final cohort comprised 2,317 White cases and 19,056 controls, 189 Black or African American cases and 1,976 controls, and 23 Asian cases and 586 controls. Although the case counts for some race/ethnicity categories were low, participants from the Middle Eastern or North African (116 participants) and Multiracial (267 participants) groups were included in the analysis to enhance representativeness. Additionally, the cohort included 241 individuals who identified as Hispanic White, 26 as Hispanic Black, and 33 as Hispanic Other.

Table 1. Importance of Features by Survey Type

| Survey | Questions | Model Feature Importance |
|---|---|---|
| Basics | ● What is the highest grade or year of school you completed?<br>● Are you covered by health insurance or some other kind of health care plan?<br>● What is your current employment status? Please select 1 or more of these categories.<br>● What is your annual household income from all sources?<br>● In the past 6 months, have you been worried or concerned about NOT having a place to live? | Medium |
| Family Health History | ● Who in your family has had prostate cancer? Select all that apply. | High |
| Health Care Access and Utilization | ● Is there a place that you USUALLY go to when you are sick or need advice about your health?<br>● About how long has it been since you last saw or talked to a doctor or other health care provider about your own health? | Medium |
| Lifestyle | ● Have you smoked at least 100 cigarettes in your entire life? | Low |
| Social Determinants of Health | ● Within the past 12 months, we worried whether our food would run out before we got money to buy more. | Low |

### 3.3 *Model Training and Evaluation*

To address the substantial class imbalance between cases and controls, we applied the Synthetic Minority Over-sampling Technique (SMOTE) from the imblearn library to the training set. SMOTE generates synthetic samples to enhance the representation of the minority class, thereby improving model training outcomes. We used cross-validation with 5 folds to assess the model's performance consistently. The dataset was divided into an 80/20 ratio for training and test sets for each fold. This split was crucial for developing and validating the logistic regression (LR) model used to predict prostate cancer risk. We employed a logistic regression model to estimate the odds ratios (ORs) for each feature. The average false positive rate (FPR) across all folds was computed to provide a robust measure of the model's performance. The importance of each feature in the model was analyzed to understand its contribution to prostate cancer risk prediction (Figure 2).

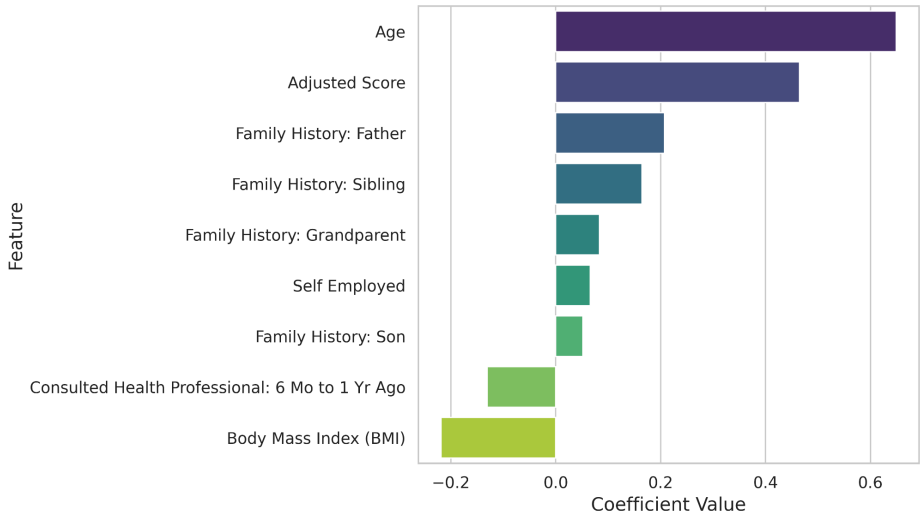


Figure 2. Feature Importance of Significant Variables from Logistic Regression

## 4. Results

### 4.1 Full Multimodal Model

In our analysis of PCa incidence, several variables demonstrated significant associations with the outcome. A family history of prostate cancer in a father (OR = 1.23, $p < 0.001$), grandparent (OR = 1.23, $p < 0.001$), sibling (OR = 1.18, $p < 0.001$), and son (OR = 1.05, $p = 0.001$) was strongly linked to increased risk. Age also emerged as a significant predictor, with each additional year increasing the odds of PCa by 1.92 ($p < 0.001$). A higher BMI was associated with a decreased risk (OR = 0.80, $p < 0.001$). The adjusted PRS score positively influenced the odds of PCa (OR =

1.59, $p < 0.001$). Additionally, being self-employed was associated with a marginal increase in risk (OR = 1.07, $p = 0.032$), while having consulted a health professional 6 months to 1 year ago was linked to a decreased risk (OR = 0.88, $p = 0.018$). These findings highlight the interplay between genetic, socio-economic, and health-related factors in influencing prostate cancer risk.

After applying SMOTE for resampling, the logistic regression model achieved an AUROC of 0.79. It correctly identified 403 prostate cancer cases, though 139 cases were misclassified. The model also accurately classified 3,428 non-cancer cases, with 1,406 false positives which represents a 29% FPR. Precision for cancer cases was 0.22, and recall was 0.74, indicating a strong ability to detect cancer cases but with room for improvement in precision.

### 4.2 Model Performance Across Different Feature Subsets

The performance of the logistic regression models varied significantly based on the features used. When utilizing only age and PRS, the model achieved an AUROC of 0.75 but exhibited low precision and recall for prostate cancer detection, identifying only a small fraction of positive cases. Incorporating additional health metrics alongside age and PRS improved the AUROC to 0.78 but still resulted in low precision and recall for cancer cases. Adding the race variable to the model further deteriorated precision and overall classification balance, leading to a higher number of false positives compared to the race-free model. Although non-cancer classifications were robust, all models faced significant challenges in accurately detecting prostate cancer, underscoring the need for further refinement to enhance sensitivity and reduce false positives.

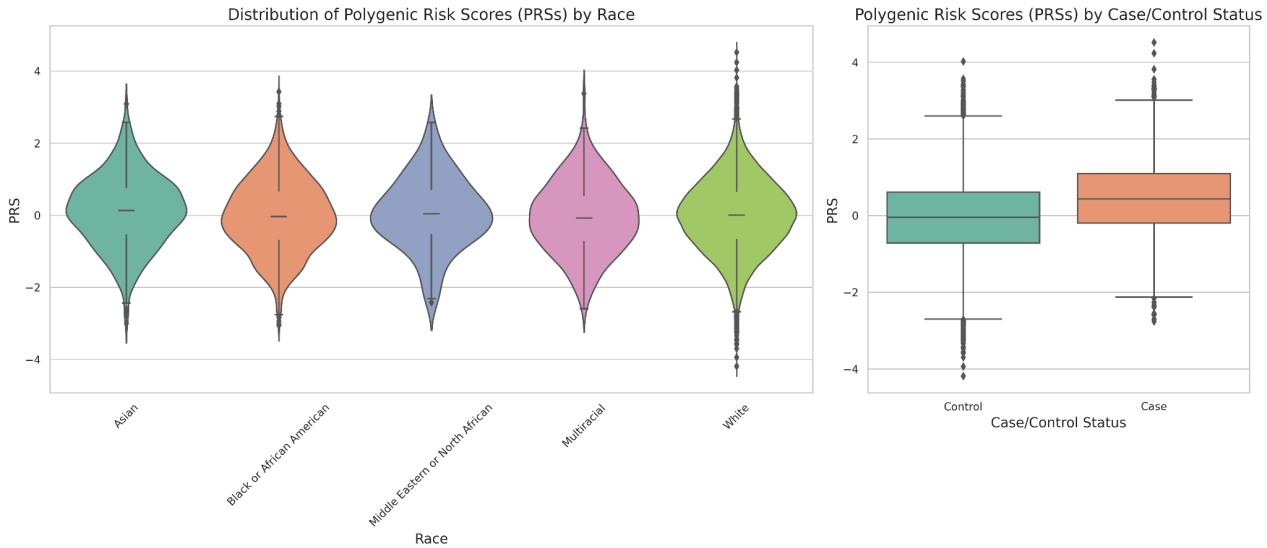


Figure 3. Distribution of Polygenic Risk Scores by different subgroups

Statistical analysis of PRSs between cases and controls revealed a significant difference. The independent t-test and Mann-Whitney U test both yielded p-values less than 0.001. This indicates a substantial disparity in PRSs between the PRS distribution for PCa cases and controls. An Analysis of Variance (ANOVA) was conducted to evaluate differences in PRSs across different racial groups. The results indicated that there was no statistically significant difference in PRS scores among the racial groups.

## 5. Discussion

In this study, we investigated the incidence of prostate cancer PCa using the national All of Us research cohort. This study identified several key factors influencing PCa risk, including increased age, family history, and higher PRSs. Additionally, we found that recent healthcare consultations were associated with a reduced risk of PCa, potentially highlighting the role of healthcare access and early detection. These findings underscore the importance of considering genetic, family health history, and healthcare utilization factors in PCa risk assessment. By advancing our understanding of these diverse factors, this research contributes to overcoming health disparities in precision medicine, emphasizing the need for tailored strategies to improve cancer prevention and treatment outcomes across different populations.

Notably, we also discovered that individuals who visited a healthcare professional within the past 6 months to a year exhibited a reduced risk of PCa compared to those who consulted a professional at other frequencies. This finding differs from other major studies that primarily focus on genetic and biological factors.[16] Regular consultations likely facilitate early detection and intervention and may be related to insurance benefits, such as improved insurance coverage and affordability which could influence healthcare utilization and early detection.. Furthermore, addressing barriers to healthcare access can mitigate disparities in PCa outcomes, emphasizing the need for equitable healthcare services. These findings highlight the importance of healthcare utilization in cancer prevention and the broader implications for health equity.

Our study identified age, higher BMI (associated with decreased risk), employment status (being self-employed), and recent healthcare consultations as significant predictors of PCa risk. These findings are consistent with other major studies, underscoring the reproducibility of our work.[17]

---

[16] Schaffer et al., "A Polygenic Risk Score for Prostate Cancer Risk Prediction"; Black et al., "Validation of a Prostate Cancer Polygenic Risk Score"; Klein et al., "Prostate Cancer Polygenic Risk Score and Prediction of Lethal Prostate Cancer."

[17] Buschemeyer and Freedland, "Obesity and Prostate Cancer"; Porter and Stanford, "Obesity and the Risk of Prostate Cancer"; Penson et al., "The Association between Socioeconomic Status, Health Insurance Coverage, and Quality of

The similarity in results reinforces the reliability of these predictors in understanding PCa risk. Particularly, the association between higher BMI and decreased risk, as well as the impact of employment status and healthcare access, aligns with existing literature, highlighting the multifaceted nature of PCa risk factors.

Adjusting class instances in a model can significantly impact performance and outcomes. Ideally, with a sufficiently large dataset, the need for data generation processes like SMOTE could be eliminated, addressing concerns about the applicability of models trained on synthetic data. Nevertheless, developing a framework for testing with real-world data remains valuable. Our use of SMOTE led to a substantial improvement in model recall, but it also resulted in decreased performance regarding FPR. Further research is needed to evaluate alternative methods for addressing class imbalance and their effects on FPR.

We evaluated several models using different combinations of feature subsets, and found that the full multimodal model incorporating PRSs, health data, and surveys consistently outperformed the others. Our analysis revealed a progressive enhancement in model performance with the inclusion of data from diverse modalities. Models without social data showed limited accuracy in predicting the positive class. With the race variable included in the full model, we observed challenges with precision and overall classification balance, resulting in a higher number of false positives compared to the model excluding race. This suggests that the inclusion of race may have led the model to overly prioritize this factor, potentially overshadowing other critical features and contributing to increased misclassification rates. These findings underscore the importance of adopting a comprehensive approach that considers multiple variables beyond race to enhance both the robustness and fairness of predictive models.

In this analysis, we predicted incident PCa without utilizing race or PSA data and achieved performance comparable to traditional models. We introduced a novel approach by integrating survey, genomic, and health data from the AoU research program. This work establishes a methodological framework for incorporating SDoH data into clinical risk assessments for PCa and potentially other cancer types, enhancing the comprehensiveness and equity of predictive modeling in clinical settings.

---

Life in Men with Prostate Cancer"; Mahal et al., "The Association between Insurance Status and Prostate Cancer Outcomes."

## 6. Limitations

The analysis was limited due to the lack of PSA data, which is a crucial biomarker for prostate cancer diagnosis and management. The absence of comprehensive PSA data meant that it could not be incorporated into the model. Additionally, the study could not evaluate the model's ability to predict prostate cancer mortality because of insufficient mortality data and the unavailability of cancer Gleason scores. Gleason scores are assigned based on the histological features of the cancer tissues, and are vital for assessing tumor aggressiveness and clinical outcomes. The exclusion of these critical variables prevented the inclusion of mortality predictions and constrained the depth of the analysis regarding disease severity.

One of the limitations of the AoU surveys as they are currently structured is the ability to establish temporality between the survey responses and the outcome of interest. Participants take surveys upon registration and regularly thereafter. For certain questions, such as those about family history of disease, establishing the timing of the family member's disease incidence may not be crucial for predicting the individual's risk. However, questions related to employment status only capture the participant's current situation, which may not reflect their status prior to a prostate cancer (PCa) diagnosis and for such cases, timing is more important. The only reliable way to establish the temporality of predictors with PCa incidence would be to limit the analysis to participants who developed PCa since the enrollment into the program. This approach would significantly reduce the study's power by decreasing the number of cases, especially among underrepresented racial categories, which are already fairly low in the positive class. More research is needed to address this limitation. One suggestion to the AoU program is to link questions from the Personal and Family Health History survey to questions in other survey questionnaires, providing detail on the participant's conditions not only at recruitment but also during the critical period before their diagnosis. This additional detail would allow for a more precise and less confounded analysis of the association between identified variables and outcomes. The availability of these surveys represents a significant advance in promoting equity in clinical algorithms. However, further work is needed to improve data annotation and refine methods for utilizing the data effectively.

## 7. Acknowledgements

This work was supported in part by the Blavatnik Fellowship. We thank the Department of Biomedical Data Science for resources to support this effort. 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 examined in this study.

## References


Black, Mary H., Shuwei Li, Holly LaDuca, Min-Tzu Lo, Jefferey Chen, Robert Hoiness, Stephanie Gutierrez, et al. "Validation of a Prostate Cancer Polygenic Risk Score." *The Prostate* 80, no. 15 (2020): 1314–21. https://doi.org/10.1002/pros.24058.

Buschemeyer, W. Cooper, and Stephen J. Freedland. "Obesity and Prostate Cancer: Epidemiology and Clinical Implications." *European Urology* 52, no. 2 (August 1, 2007): 331–43. https://doi.org/10.1016/j.eururo.2007.04.069.

Cai, Tongan, Haomiao Ni, Mingli Yu, Xiaolei Huang, Kelvin Wong, John Volpi, James Z. Wang, and Stephen T. C. Wong. "*DeepStroke*: An Efficient Stroke Screening Framework for Emergency Rooms with Multimodal Adversarial Deep Learning." *Medical Image Analysis* 80 (August 1, 2022): 102522. https://doi.org/10.1016/j.media.2022.102522.

Conti, David V., Burcu F. Darst, Lilit C. Moss, Edward J. Saunders, Xin Sheng, Alisha Chou, Fredrick R. Schumacher, et al. "Trans-Ancestry Genome-Wide Association Meta-Analysis of Prostate Cancer Identifies New Susceptibility Loci and Informs Genetic Risk Prediction." *Nature Genetics* 53, no. 1 (January 2021): 65–75. https://doi.org/10.1038/s41588-020-00748-0.

Cuzick, Jack, Mangesh A. Thorat, Gerald Andriole, Otis W. Brawley, Powel H. Brown, Zoran Culig, Rosalind A. Eeles, et al. "Prevention and Early Detection of Prostate Cancer." *The Lancet Oncology* 15, no. 11 (October 1, 2014): e484–92. https://doi.org/10.1016/S1470-2045(14)70211-6.

Dwyer-Lindgren, Laura, Amelia Bertozzi-Villa, Rebecca W. Stubbs, Chloe Morozoff, Johan P. Mackenbach, Frank J. Van Lenthe, Ali H. Mokdad, and Christopher J. L. Murray. "Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers." *JAMA Internal Medicine* 177, no. 7 (July 1, 2017): 1003. https://doi.org/10.1001/jamainternmed.2017.0918.

Klebaner, Daniella, Patrick Travis Courtney, and Brent S. Rose. "Effect of Healthcare System on Prostate Cancer-Specific Mortality in African American and Non-Hispanic White Men." *Journal of Clinical Oncology* 39, no. 6_suppl (February 20, 2021): 23–23. https://doi.org/10.1200/JCO.2021.39.6_suppl.23.

Klein, Robert J., Emily Vertosick, Dan Sjoberg, David Ulmert, Ann-Charlotte Rönn, Christel Häggström, Elin Thysell, et al. "Prostate Cancer Polygenic Risk Score and Prediction of Lethal Prostate Cancer." *Npj Precision Oncology* 6, no. 1 (April 8, 2022): 1–8. https://doi.org/10.1038/s41698-022-00266-8.

Lin, Kenneth, Robert Lipsitz, Therese Miller, and Supriya Janakiraman. "Benefits and Harms of Prostate-Specific Antigen Screening for Prostate Cancer: An Evidence Update for the U.S. Preventive Services Task Force." *Annals of Internal Medicine* 149, no. 3 (August 5, 2008): 192–99. https://doi.org/10.7326/0003-4819-149-3-200808050-00009.

Mahal, B. A., A. A. Aizer, D. R. Ziehr, A. S. Hyatt, C. Lago-Hernandez, Y.-W. Chen, T. K. Choueiri, et al. "The Association between Insurance Status and Prostate Cancer Outcomes: Implications for the Affordable Care Act." *Prostate Cancer and Prostatic Diseases* 17, no. 3 (September 2014): 273–79. https://doi.org/10.1038/pcan.2014.23.

Mateo, Joaquin, Rana McKay, Wassim Abida, Rahul Aggarwal, Joshi Alumkal, Ajjai Alva, Felix Feng, et al. "Accelerating Precision Medicine in Metastatic Prostate Cancer." *Nature Cancer* 1, no. 11 (November 2020): 1041–53. https://doi.org/10.1038/s43018-020-00141-0.

Mohsen, Farida, Hazrat Ali, Nady El Hajj, and Zubair Shah. "Artificial Intelligence-Based Methods for Fusion of Electronic Health Records and Imaging Data." *Scientific Reports* 12, no. 1 (October 26, 2022): 17981. https://doi.org/10.1038/s41598-022-22514-4.

Penson, David F, Marcia L Stoddard, David J Pasta, Deborah P Lubeck, Scott C Flanders, and Mark S Litwin. "The Association between Socioeconomic Status, Health Insurance Coverage, and Quality of Life in Men with Prostate Cancer." *Journal of Clinical Epidemiology* 54, no. 4 (April 1, 2001): 350–58. https://doi.org/10.1016/S0895-4356(00)00312-7.

Porter, Michael P., and Janet L. Stanford. "Obesity and the Risk of Prostate Cancer." *The Prostate* 62, no. 4 (2005): 316–21. https://doi.org/10.1002/pros.20121.

Qian, Xuejun, Jing Pei, Hui Zheng, Xinxin Xie, Lin Yan, Hao Zhang, Chunguang Han, et al. "Prospective Assessment of Breast Cancer Risk from Multimodal Multiview Ultrasound Images via Clinically Applicable Deep Learning." *Nature Biomedical Engineering* 5, no. 6 (June 2021): 522–32. https://doi.org/10.1038/s41551-021-00711-2.

Schaffer, Kerry R., Mingjian Shi, John P. Shelley, Jeffrey J. Tosoian, Linda Kachuri, John S. Witte, and Jonathan D.

Mosley. “A Polygenic Risk Score for Prostate Cancer Risk Prediction.” *JAMA Internal Medicine* 183, no. 4 (April 1, 2023): 386–88. https://doi.org/10.1001/jamainternmed.2022.6795.

Tosteson, Anna N. A., Dennis G. Fryback, Cristina S. Hammond, Lucy G. Hanna, Margaret R. Grove, Mary Brown, Qianfei Wang, Karen Lindfors, and Etta D. Pisano. “Consequences of False-Positive Screening Mammograms.” *JAMA Internal Medicine* 174, no. 6 (June 1, 2014): 954–61. https://doi.org/10.1001/jamainternmed.2014.981.

US Preventive Services Task Force. “Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement.” *JAMA* 319, no. 18 (May 8, 2018): 1901–13. https://doi.org/10.1001/jama.2018.3710.

Van Poppel, Hendrik, Tit Albreht, Partha Basu, Renée Hogenhout, Sarah Collen, and Monique Roobol. “Serum PSA-Based Early Detection of Prostate Cancer in Europe and Globally: Past, Present and Future.” *Nature Reviews Urology* 19, no. 9 (September 2022): 562–72. https://doi.org/10.1038/s41585-022-00638-6.

Yamoah, Kosj, Kyung Min Lee, Shivanshu Awasthi, Patrick R. Alba, Cristina Perez, Tori R. Anglin-Foote, Brian Robison, et al. “Racial and Ethnic Disparities in Prostate Cancer Outcomes in the Veterans Affairs Health Care System.” *JAMA Network Open* 5, no. 1 (January 18, 2022): e2144027. https://doi.org/10.1001/jamanetworkopen.2021.44027.