medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Discover overlooked comorbidities of preeclampsia using electronic health records (EHR)

Haoming Zhu[1], Xiaotong Yang[1], Ruowang Li[2], Lana X Gamire[1]*.

[1]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, 48105, United States of America

[2]Department of Computational Biomedicine, Cedars-Sinai Medical Center, West Hollywood, CA, 90069, United States of America

*Corresponding author

Abstract

Background

Preeclampsia (PE) is a severe pregnancy complication characterized by hypertension and proteinuria. PE poses a significant threat to women's health, including an increased risk of cardiovascular and renal diseases. Most previous studies have been hypothesis-based, potentially overlooking significant comorbidities. This study conducts a comprehensive, non-hypothesis-based analysis of PE comorbidities using large-scale electronic health records (EHR) data.

Method

From the University of Michigan (UM) Healthcare System, we collected 4,348 PE patients for the case cohort and 27,377 patients with pregnancies not complicated by PE or related conditions for the control cohort. We first conducted a non-hypothesis-based exploration to

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

identify any comorbidities associated with PE using logistic regression with adjustments to demographics, lifestyles, and medical history. We validated the results by performing logistic regression on the post-PE risks of the identified comorbidities from UKBiobank data. The validation data contain 443 PE cases and 14,954 non-PE controls from UKBiobank. We then conducted survival analysis on comorbidities that exhibited significance in more than 5 consecutive years post-PE. We further examined the racial disparities of identified comorbidities between Caucasian and African American patients.

Findings

Uncomplicated hypertension, complicated diabetes, congestive heart failure, renal failure, obesity, and hypothyroidism exhibited significantly increased risks in 5 consecutive years after PE in the UM discovery data. Caucasians were more sensitive to PE's impact on hypertension and hypothyroidism compared to African Americans. Validation using UKBiobank data confirmed increased risks in uncomplicated hypertension, complicated diabetes, congestive heart failure, renal failure, and obesity.

Interpretation

This study addresses the lack of a comprehensive examination of PE's long-term effects utilizing large-scale EHR and modern statistical methods. Our findings underscored the need for long-term monitoring and interventions for women with a history of PE, emphasizing the importance of personalized postpartum care. Notably, the racial disparities observed in the impact of PE on hypertension and hypothyroidism highlight the necessity of tailored aftercare based on ethnicity.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Preeclampsia (PE) is a severe pregnancy complication characterized by the onset of hypertension and proteinuria after the 20th week of pregnancy [1]. Its prevalence is estimated at approximately 4% in the United States and ranges from 2% to 8% globally [3, 4]. The immediate consequences of PE are substantial, impacting both maternal and fetal health severely. The hypertensive state induces heightened vascular resistance, potentially compromising blood flow to vital organs [2]. Resulting endothelial dysfunction contributes to multiple organ damage, affecting the kidneys, liver, and central nervous system of the mother [2]. Additionally, it can adversely affect placental perfusion, resulting in fetal growth restrictions and preterm births [2]. While childbirth often alleviates acute symptoms, the enduring impact of PE on maternal health can persist for years post-pregnancy, evidenced by chronic comorbidities [8].

Despite the recognition of PE's immediate consequences, there is a notable gap in comprehensive analyses of its long-term impacts using advanced statistical methods. Previous studies, including cohort studies and meta-analyses, have indicated chronic hypertension, cardiovascular, renal, and metabolic disorders post-PE [5-9, 11-13, 21-26]. However, these studies often adopted a hypothesis-based approach, focusing on confirming known associations and potentially overlooking important associations. Moreover, the majority of studies utilized basic statistical models (e.g., t-tests, chi-square tests, linear regression) [5-8, 11-13, 21-26] despite the need for more robust analytical methods to capture crucial time-to-event information and consider censorship, especially when dealing with longitudinal data over large time spans. In addition, few studies have paid attention to the racial disparities in PE's long-term effects, despite the known influence of race on disease risks. As statistical methods evolve, a more robust and comprehensive approach utilizing modern analytical methods is imperative to explore the long-term effects of PE and offer clinical insights.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

With the widespread use of electronic medical systems, the Electronic Health Record (EHR) has become an important source of patient data for medical research. EHR provides comprehensive, timestamped patient information over time, including demographics, diagnosis records, social history, laboratory results, etc., making it highly compatible with survival analysis. Survival analysis enables the examination of time until the occurrence of specific events [10], investigating both the risks and the trajectories concerning the time of specific conditions. Integrating survival analysis with traditional statistical models in EHR analysis offers a better understanding of PE's long-term consequences.

This study adopts a non-hypothesis-based exploration, leveraging EHR data to investigate connections between PE and subsequent health trajectories. Our goal is to first identify any long-term adverse effects significantly associated with PE, and then study the trajectory of these conditions and their racial disparities using survival analysis. Through this analysis, we aim to enhance personalized postpartum care and ultimately improve the prognosis for women with a PE history.

Methods

The analysis consists of two stages. Initially, a non-hypothesis-based exploratory analysis using logistic regression was conducted to identify any comorbidities significantly associated with PE. Subsequently, survival analysis was applied to these comorbidities to further examine their trajectories and potential racial disparities.

Data Source

Discovery data for this project was acquired from the *University of Michigan (UM) Medicine Healthcare System*, an academic health system catering to Ann Arbor and surrounding regions. The University of Michigan Medical School's Institutional Review Board (IRB) granted data

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

utilization approval under HUM#00168171. The EHR from the *UM Medicine Healthcare System* provided comprehensive features, including diagnoses, encounter information, demographics, medications, etc. Validation data were obtained from the *UKBiobank*, a long-term study with approximately 500,000 volunteers, providing similar features [31]. The regression models initially utilized features including diseases diagnosed, age at diagnosis, medical history, pre-existing comorbidities, race, and social history.

Original EHR data for case and control cohorts were obtained. The case cohort comprised patients with at least one PE diagnosis between 2003 and 2023. The control cohort comprised patients with at least one pregnancy during the same period, excluding those ever diagnosed with PE-related diseases (PE, eclampsia, pre-existing hypertension complicating pregnancy, gestational edema, and maternal hypertension). Inclusion and exclusion criteria were based on the International Classification of Diseases (ICD)-9 and -10 codes [14, 15]. Detailed case-control selection criteria based on ICD-9 and ICD-10 codes can be found in Supplementary Table 1.

EHR feature engineering and Non-hypothesis-based PE comorbidity Discovery

In this study, we defined a concept, the *underlying diagnosis*, which referred to the patient's first PE diagnosis in the case group and the patient's first pregnancy diagnosis in the control group. Diagnoses were grouped into ten different *year groups* based on the time since the patient's *underlying diagnosis*. Each *year group* contains diagnoses that 1) occurred within corresponding years from the patient's *underlying diagnosis*, and 2) came from a patient whose last diagnosis occurred corresponding years or more after their *underlying diagnosis*.

Diagnoses, initially recorded using a combination of ICD-9 and ICD-10 codes, were first standardized to ICD-10 using the "touch" package [28]. Diagnoses were then classified into 31 comorbidities according to the Elixhauser Comorbidity Index [16, 17]. A complete list of matches

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

between ICD-10 codes and Elixhauser Comorbidities is shown in Supplementary Table 2. For each patient, the pre-existing comorbidities were defined as those presented at or before the *underlying diagnosis*. The pre-existing conditions were encoded as binary entries (1 for presence, 0 for absence). Additional features, including PE history, gestational diabetes history, and living habits (alcohol and smoking status), were encoded as binary entries in the same manner. To enhance statistical power and reduce model complexity, features with more than 20% missing values or a p-value > 0.10 in the univariable test were removed. The feature selection process was applied independently to each *year group*. A complete list of final features in each *year group* is shown in Supplementary Table 3.

In each *year group,* logistic regression was applied to comorbidities that reached the effective sample size of regression. A binary variable was added to indicate the presence of comorbidities after the *underlying diagnosis* of each patient, which served as the dependent variable in logistic regression. Within each comorbidity's regression model, this variable was marked as 1 for patients who had the comorbidity after the *underlying diagnosis* and 0 for patients who did not have the comorbidity after the *underlying diagnosis*. Using package "glmnet" [18], a logistic model was fitted for each comorbidity in each *year group*. In each regression model, the regression coefficient, standard error, and p-value for feature "PE" were recorded. Odds ratios (OR) and corresponding 95% confidence intervals (CI) were calculated. OR is a measure of the association between the presence of a particular condition and a specific outcome, expressing the odds of the event occurring in the case group relative to the odds in the control group [36]. Comorbidities that showed significance in OR ($p < 0.05$) in more than 5 consecutive years were considered significantly associated with PE. Variance Inflation Factors (VIF) were calculated using the package "qacReg" [35] to check for potential collinearity among independent variables.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Validation using external data

To validate the significant comorbidities identified in the UM database, logistic regression was conducted on UKBiobank data following a similar process. A binary variable indicating the presence of comorbidity after *underlying diagnosis* served as the dependent variable, while PE history, pre-existing comorbidities, living habits (smoking and alcohol status), and race served as the regressors. ORs and CIs were calculated for the feature “PE".

Further inspection of identified PE comorbidities and their racial disparity using survival analysis

Comorbidities that were significantly associated with PE were selected for survival analysis to further explore the trajectories. Entire patient data (without grouping into *year groups*) was used for the survival analysis. Patients that died due to non-targeted diseases or lost follow-up were marked as censored. For each significant comorbidity, a Cox proportional-hazards (Cox-ph) regression model [19] was built. To further account for unrelated death’s effect on comorbidity risks, the Cox-ph models were adjusted for competing risks using the “tidycmprsk” package [20]. Features used in the Cox-ph models can be found in Supplementary Table 4. Hazard ratios (HR) of the factor “PE” were calculated from each survival model. HR is a measure used in survival analysis to compare the event rates at any given time point between two groups [37].

Patients in the case and control groups were further stratified according to their ethnicity (Caucasian, African American, Asian, American Indian or Alaska Native, Native Hawaiian and Other Pacific Islanders, Other Races, Unknown). Similar survival analysis on comorbidity risks was applied to each race. HR of the factor “PE” for a specific race was calculated from the survival models.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Analytical Tools

All analyses were conducted in R programming language v.4.0.3 [29], using packages "tidyverse" [33], "lubridate" [32], "glmnet" [18], "tidycmprsk" [20], "touch" [28], "survminer" [34], and "qacReg" [35].

Results

Study overview

Using two large EHR databases, we discovered and validated the overlooked post-PE comorbidities. We employed a non-hypothesis-based approach, utilizing EHR from the UM Medicine Healthcare System, incorporating logistic regression and survival analysis to comprehensively investigate the long-term effects of PE, identifying persistent comorbidities and exploring racial disparities, with subsequent validation using UKBiobank data. The overall workflow of the study is shown in Figure 1.

Cohort overview

The discovery data from *UM Medicine Healthcare System* included 4,348 patients for the case cohort and 27,377 patients for the control cohort. Patient characteristics are summarized in Table 1.

The validation data from *UKBiobank* included 443 patients for the case cohort and 14954 patients for the control cohort. Patient characteristics are summarized in Table 2. Due to data privacy restrictions, the characteristics overview was calculated based on a subset of the complete data (case = 443, control = 2000).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

In both discovery data and validation data, significant differences between the case and control cohorts were observed in demographic and clinical factors. To account for the differences, these factors were included in the regression models as confounders.

Logistic regression model discovered six comorbidities that were significantly related to PE's long-term effects

Data was pre-processed following the procedure in the Method section. The populations in each *year group* are shown in Supplementary Table 5. Case and control cohorts were then combined into the final dataset and Logistic regression was applied to each comorbidity in each *year group*. No significant collinearity (VIF > 5) was observed among the independent variables in all *year groups*.

OR and CI for each comorbidity in each *year group* were calculated and plotted against time to reveal the trajectory of PE's long-term effects (Figure 2). Within 1~10 years, uncomplicated hypertension (median OR = 8.85), renal failure (median OR = 1.84), complicated diabetes (median OR = 1.57), congestive heart failure (median OR = 1.47), obesity (median OR = 1.40), and hypothyroidism (median OR = 0.84) are the comorbidities that exhibited significance in more than 5 consecutive years after *underlying diagnosis*. Notably, the ORs of hypothyroidism are significantly lower than 1 in the first 5 years after the *underlying diagnosis*, though with an overall evaluating trend.

Validation confirmed the significance of five comorbidities

The six significant comorbidities were validated using logistic regressions on the UKBiobank data. The validation result is shown in Figure 3C. Renal failure (OR = 5.27, 95% CI: 3.35-8.29), congestive heart failure (OR = 3.33, 95% CI: 1.63-6.82), uncomplicated hypertension (OR = 3.30, 95% CI: 2.58-4.22), complicated diabetes (OR = 2.97, 95% CI: 2.08-4.23), and obesity

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(OR = 2.09, 95% CI: 1.53-2.86) exhibited significant increasing risks due to PE. Although hypothyroidism (OR = 1.34, 95% CI: 0.90-2.00) has a significance level slightly below the threshold, it suggests a potential increased risk caused by PE. The insignificance and larger CIs in the validation could be explained by the insufficiency of samples for cases in the validation data.

Survival analysis revealed significantly increasing risks in the six comorbidities and racial disparities in two comorbidities

The six significant comorbidities identified from the UM database were all kept for survival analysis. The survival curves for each comorbidity are shown in Figure 3A. All targeted comorbidities exhibited significance between the survival curves of the case and control groups (Log-Rank Test, $p < 0.05$).

Cox-ph models adjusted for confounders and competing risks were applied to each comorbidity. A complete table for the feature coefficients and significant levels can be found in Supplementary Table 4. HRs and CIs specific to PE were calculated and plotted (Figure 3B). Compared to women without a PE-related history, the HRs due to PE's effect are 5.70 (95% CI: 5.29-6.14) for uncomplicated hypertension, 2.22 (95% CI: 1.84-2.67) for complicated diabetes, 2.90 (CI: 2.32-3.62) for congestive heart failure, 1.90 (CI: 1.50-2.41) for renal disease, 1.68 (CI: 1.58-1.79) for obesity, and 1.22 (CI: 1.11-1.33) for hypothyroidism, all of which are significant.

Further, the data was grouped according to ethnicity. Due to insufficient samples, Asian, American Indian or Alaska Native, Native Hawaiian and Other Pacific Islanders, Other Races, and Unknown patients were dropped in racial analysis; Caucasian and African American patients were compared.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A similar process that compared the case and control groups within Caucasian and African American populations was applied. The race-specific survival curves are shown in Figure 4A. Uncomplicated hypertension, hypothyroidism, obesity, and congestive heart failure exhibited significance between the survival curves of African Americans and Caucasians within the case cohort (Log-Rank Test, $p < 0.05$)

Cox-ph models adjusted for confounders and competing risks were applied to each comorbidity. The race-specific HRs of PE are shown in Figure 4B. After adjusting for confounders and competing risks, significant racial disparity was only observed in the occurrence of hypothyroidism and uncomplicated hypertension. For hypothyroidism, PE's effect on increased hypothyroidism risk was only observed in Caucasians (HR = 1.30, CI: 1.18-1.44), while this effect was not significant in African Americans (HR = 0.81, CI: 0.60-1.11). For uncomplicated hypertension, Caucasians (HR = 6.15, CI: 5.62-6.73) had significantly higher HR due to PE than African Americans (HR = 4.12, CI: 3.50-4.85). Other comorbidities also exhibited racial differences; however, the disparities were not significant.

Discussion

This study employed a non-hypothesis-based approach and rigorous statistical analysis to comprehensively investigate and validate the long-term effects of PE using EHR from the *UM Medicine Healthcare System* and *UKBiobank* database. In addition to confirming previously known conditions, we discovered overlooked comorbidities of PE and their racial disparities. These findings can encourage better management and interventions that benefit post-PE patient care.

The logistic regression models identified six comorbidities from the *UM Medicine Healthcare* database—uncomplicated hypertension, complicated diabetes, congestive heart failure, renal

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

failure, obesity, and hypothyroidism—that exhibited sustained significance over 5 or more consecutive years following PE, as indicated by their ORs. Our findings were validated using data from the *UKBiobank*. The increased risks of uncomplicated hypertension, obesity, complicated diabetes, renal failure, and congestive heart failure due to PE were confirmed. We then conducted survival analysis on all six comorbidities, further taking time-to-event factors into account and confirming the enduring impact of PE. All six comorbidities showed significant increasing risks due to PE in the longitudinal analysis. The findings underscore the need for long-term monitoring and interventions for patients who have a PE history.

The increasing risks of hypertension [5, 8, 11, 12, 13], diabetes [8, 24, 25], congestive heart failure [5, 6, 8, 13, 23], renal failure [8, 21, 22], and obesity [8, 24, 25] after PE have been recognized by pre-existing studies on PE.

Specifically, previous studies indicated that while some women experience a return-to-normal blood pressure after the termination of a preeclamptic pregnancy, a proportion remains hypertensive [11, 12, 13]; even for those whose blood pressure normalizes, the risk of developing chronic hypertension in the future increases [5, 11, 13]. Acute gestational hypertension caused by PE is associated with an elevated risk of subsequent cardiovascular dysfunction, a significant factor in congestive heart failure [5, 6, 13, 23]. PE's typical symptom, proteinuria, is linked to endothelial damage and podocyte loss in the kidney, increasing the risk of subsequent chronic renal disease, and eventually leading to renal failure [21, 22]. The correlation between PE and metabolic syndrome contributes to higher risks of developing diabetes and obesity in women with a history of preeclampsia [24, 25]. Our research confirmed these findings via more robust and comprehensive statistical analyses using large-scale EHR data.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

One novel finding of our study was the discovery of the association between PE and hypothyroidism. Previous research suggests that PE might be related to an increased thyroid-stimulating hormone level [26], which is an indicator of hypothyroidism. Our study confirmed the relationship between the increasing risks of hypothyroidism and PE using more robust analytics and larger samples.

Another novel finding in our study was the racial disparities in PE's adverse effects. Notably, our findings indicate that Caucasian women have significantly higher HR in uncomplicated hypertension than African American women, although both races exhibited increased risks in hypertension after PE. The significant difference in HR suggests that Caucasian women are more sensitive to PE's eliciting effect on hypertension than American African women.

Our study has suggested that PE would lead to an overall higher risk of hypothyroidism. Interestingly, after stratifying patients into Caucasians and African Americans, PE's eliciting effect on hypothyroidism was only observed among Caucasians, while African Americans did not respond significantly to PE's effect on hypothyroidism. One previous review study has suggested that white patients were generally referred for consultation for hypothyroidism at a younger age compared to black patients and that non-white patients have a higher probability of being underdiagnosed [27]. Therefore, it remains unclear whether the racial disparities in hypothyroidism after PE are primarily related to genetic disparities or bias in the diagnosis between the two races.

Our study had several notable strengths compared to previous studies on the long-term effects of PE. Methodologically, we utilized a non-hypothesis-based approach to identify any association between comorbidities and PE, allowing for a comprehensive investigation beyond preconceived hypotheses and the discovery of overlooked associations. Our analysis was adjusted for pre-PE medical histories, minimizing the bias caused by pre-existing conditions.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

This addressed the pitfall in traditional case-control matching studies [7, 9, 11-13]. In each step of our analysis, we conducted rigorous inference that ensured sufficient sample size and statistical power. We adopted a discovery-validation research strategy, findings from the UM database were validated using data from UKBiobank data, which strengthens the generalizability and reliability of our findings. There is a notable lack of previous studies about the racial disparities of PE's long-term adverse effects. Our study filled the research gap in this topic, revealing the association between ethnicity and PE's long-term effects. We extended the PE study to a novel dimension, hoping to enhance the understanding of this maternal disease from an ethnic perspective and promote precision aftercare of PE.

While our study provides valuable insights into the long-term effects of PE, it is important to acknowledge certain limitations that may impact the generalizability of our findings. One aspect pertains to the reliance on ICD codes for case-control identification. ICD codes may introduce biases and inaccuracies, as coding practices can vary between institutions. Furthermore, the complexity of maternal health, encompassing various genetic, environmental, and lifestyle factors, makes it challenging to delineate direct causal relationships between PE and long-term comorbidities. Future studies incorporating causal inference or randomized controlled trials may offer more insights into the causal pathways between PE and subsequent health outcomes. Our findings were purely based on EHR data; Other types of data that explain the molecular and pathological processes, such as genetics and genomics data, would be beneficial to deepen the mechanistic understanding of PE's long-term effects.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

[1]

M. A. Brown, M. D. Lindheimer, M. de Swiet, A. Van Assche, and J. M. Moutquin, "The classification and diagnosis of the hypertensive disorders of pregnancy: statement from the International Society for the Study of Hypertension in Pregnancy (ISSHP)," *Hypertens Pregnancy*, vol. 20, no. 1, p. IX–XIV, 2001, doi: 10.1081/PRG-100104165.

[2]

US Preventive Services Task Force, "Screening for Preeclampsia: US Preventive Services Task Force Recommendation Statement," *JAMA*, vol. 317, no. 16, pp. 1661–1667, Apr. 2017, doi: 10.1001/jama.2017.3439.

[3]

L. Duley, "The global impact of pre-eclampsia and eclampsia," *Semin Perinatol*, vol. 33, no. 3, pp. 130–137, Jun. 2009, doi: 10.1053/j.semperi.2009.02.010.

[4]

"ACOG Practice Bulletin No. 202: Gestational Hypertension and Preeclampsia," *Obstet Gynecol*, vol. 133, no. 1, p. 1, Jan. 2019, doi: 10.1097/AOG.0000000000003018.

[5]

L. Bellamy, J.-P. Casas, A. D. Hingorani, and D. J. Williams, "Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis," *BMJ*, vol. 335, no. 7627, p. 974, Nov. 2007, doi: 10.1136/bmj.39335.385301.BE.

[6]

S. D. McDonald, A. Malinowski, Q. Zhou, S. Yusuf, and P. J. Devereaux, "Cardiovascular sequelae of preeclampsia/eclampsia: A systematic review and meta-analyses," *American Heart Journal*, vol. 156, no. 5, pp. 918–930, Nov. 2008, doi: 10.1016/j.ahj.2008.06.042.

[7]

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

C. van Walraven, M. Mamdani, A. Cohn, Y. Katib, M. Walker, and M. A. Rodger, “Risk of subsequent thromboembolism for patients with pre-eclampsia,” *BMJ*, vol. 326, no. 7393, pp. 791–792, Apr. 2003, doi: 10.1136/bmj.326.7393.791.

[8]

D. Williams, “Long-term complications of preeclampsia,” *Semin Nephrol*, vol. 31, no. 1, pp. 111–122, Jan. 2011, doi: 10.1016/j.semnephrol.2010.10.010.

[9]

R. J. Levine *et al.*, “Pre-eclampsia, soluble fms-like tyrosine kinase 1, and the risk of reduced thyroid function: nested case-control and population based study,” *BMJ*, vol. 339, p. b4336, Nov. 2009, doi: 10.1136/bmj.b4336.

[10]

B. George, S. Seals, and I. Aban, “Survival analysis and regression models,” *J Nucl Cardiol*, vol. 21, no. 4, pp. 686–694, Aug. 2014, doi: 10.1007/s12350-014-9908-2.

[11]

V. D. Garovic and P. August, “Preeclampsia and the Future Risk of Hypertension: The Pregnant Evidence,” *Curr Hypertens Rep*, vol. 15, no. 2, p. 10.1007/s11906-013-0329–4, Apr. 2013, doi: 10.1007/s11906-013-0329-4.

[12]

L. C. Chappell, S. Enye, P. Seed, A. L. Briley, L. Poston, and A. H. Shennan, “Adverse perinatal outcomes and risk factors for preeclampsia in women with chronic hypertension: a prospective study,” *Hypertension*, vol. 51, no. 4, pp. 1002–1009, Apr. 2008, doi: 10.1161/HYPERTENSIONAHA.107.107565.

[13]

W. Hermes *et al.*, “Cardiovascular risk factors in women who had hypertensive disorders late in pregnancy: a cohort study,” *American Journal of Obstetrics and Gynecology*, vol. 208, no. 6, p. 474.e1-474.e8, Jun. 2013, doi: 10.1016/j.ajog.2013.02.016.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[14]

"ICD - ICD-9-CM - International Classification of Diseases, Ninth Revision, Clinical Modification." Accessed: Nov. 28, 2023. [Online]. Available: https://www.cdc.gov/nchs/icd/icd9cm.htm

[15]

"ICD- 10 - CM International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)." Accessed: Nov. 28, 2023. [Online]. Available: https://www.cdc.gov/nchs/icd/icd-10-cm.htm

[16]

A. Elixhauser, C. Steiner, D. R. Harris, and R. M. Coffey, "Comorbidity measures for use with administrative data," *Med Care*, vol. 36, no. 1, pp. 8–27, Jan. 1998, doi: 10.1097/00005650-199801000-00004.

[17]

"Elixhauser Comorbidity Software Refined for ICD-10-CM." Accessed: Nov. 28, 2023. [Online]. Available: https://hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp

[18]

J. Friedman *et al.*, "glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models." Aug. 22, 2023. Accessed: Nov. 28, 2023. [Online]. Available: https://cran.r-project.org/web/packages/glmnet/index.html

[19]

D. R. Cox and D. Oakes, "Analysis of Survival Data".

[20]

D. D. Sjoberg and T. Fei, "tidycmprsk: Competing Risks Estimation." Oct. 30, 2023. Accessed: Nov. 28, 2023. [Online]. Available: https://cran.r-project.org/web/packages/tidycmprsk/index.html

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[21]

A. G. Kattah *et al.*, “Preeclampsia and ESRD: The Role of Shared Risk Factors,” *Am J Kidney Dis*, vol. 69, no. 4, pp. 498–505, Apr. 2017, doi: 10.1053/j.ajkd.2016.07.034.

[22]

B. E. Vikse, L. M. Irgens, T. Leivestad, R. Skjaerven, and B. M. Iversen, “Preeclampsia and the risk of end-stage renal disease,” *N Engl J Med*, vol. 359, no. 8, pp. 800–809, Aug. 2008, doi: 10.1056/NEJMoa0706790.

[23]

G. C. Smith, J. P. Pell, and D. Walsh, “Pregnancy complications and maternal risk of ischaemic heart disease: a retrospective cohort study of 129 290 births,” *The Lancet*, vol. 357, no. 9273, pp. 2002–2006, Jun. 2001, doi: 10.1016/S0140-6736(00)05112-6.

[24]

A. Pouta *et al.*, “Manifestations of Metabolic Syndrome After Hypertensive Pregnancy,” *Hypertension*, vol. 43, no. 4, pp. 825–831, Apr. 2004, doi: 10.1161/01.HYP.0000120122.39231.88.

[25]

D. B. Carr *et al.*, “Preeclampsia and Risk of Developing Subsequent Diabetes,” *Hypertension in Pregnancy*, vol. 28, no. 4, pp. 435–447, Nov. 2009, doi: 10.3109/10641950802629675.

[26]

J. M. Roberts and A. Rajakumar, “Preeclampsia and Soluble fms-Like Tyrosine Kinase 1,” *The Journal of Clinical Endocrinology & Metabolism*, vol. 94, no. 7, pp. 2252–2254, Jul. 2009, doi: 10.1210/jc.2009-0945.

[27]

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A. Gillis, H. Chen, T. S. Wang, and S. Dream, "Racial and Ethnic Disparities in the Diagnosis and Treatment of Thyroid Disease," *The Journal of Clinical Endocrinology & Metabolism*, p. dgad519, Aug. 2023, doi: 10.1210/clinem/dgad519.

[28]

W. Wang, Y. Li, and J. Yan, "touch: Tools of Utilization and Cost in Healthcare." Jul. 08, 2022. Accessed: Nov. 28, 2023. [Online]. Available: https://cran.r-project.org/web/packages/touch/index.html

[29]

"R: The R Project for Statistical Computing." Accessed: Nov. 28, 2023. [Online]. Available: https://www.r-project.org/

[30]

G. Grolemund and H. Wickham, "Dates and Times Made Easy with lubridate," *Journal of Statistical Software*, vol. 40, pp. 1–25, Apr. 2011, doi: 10.18637/jss.v040.i03.

[31]

"UK Biobank - UK Biobank." Accessed: Nov. 28, 2023. [Online]. Available: https://www.ukbiobank.ac.uk/

[32]

V. Spinu *et al.*, "lubridate: Make Dealing with Dates a Little Easier." Sep. 27, 2023. Accessed: Nov. 29, 2023. [Online]. Available: https://cran.r-project.org/web/packages/lubridate/index.html

[33]

H. Wickham and RStudio, "tidyverse: Easily Install and Load the 'Tidyverse.'" Feb. 22, 2023. Accessed: Nov. 29, 2023. [Online]. Available: https://cran.r-project.org/web/packages/tidyverse/index.html

[34]

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A. Kassambara, M. Kosinski, P. Biecek, and S. Fabian, “survminer: Drawing Survival Curves using ‘ggplot2.’” Mar. 09, 2021. Accessed: Nov. 29, 2023. [Online]. Available: https://cran.r-project.org/web/packages/survminer/index.html

[35]

R. Kabacoff, “qacReg.” Jun. 08, 2022. Accessed: Nov. 30, 2023. [Online]. Available: https://github.com/Rkabacoff/qacReg

[36]

M. Szumilas, “Explaining Odds Ratios,” *J Can Acad Child Adolesc Psychiatry*, vol. 19, no. 3, pp. 227–229, Aug. 2010. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938757/

[37]

A. Sashegyi and D. Ferry, “On the Interpretation of the Hazard Ratio and Communication of Survival Benefit,” *Oncologist*, vol. 22, no. 4, pp. 484–486, Apr. 2017, doi: 10.1634/theoncologist.2016-0198.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1. Patient Characteristics for UM Data.

Feature	Statistics		P-value
	Case (n = 4348)	Control (n = 27377)	
Age, Years	32.34	32.97	<0.001
Race, Rate			
African American	0.190	0.130	<0.001
Caucasian	0.708	0.714	0.435
Other	0.102	0.156	<0.001
Smoking History, Rate			
Smoking	0.343	0.304	<0.001
Alcohol	0.557	0.535	<0.001
Pre-existing Disease, Occurrence Rate			
Gestational Diabetes	0.067	0.055	0.002
Other Complicated Diabetes	0.039	0.005	<0.001
Hypertension Uncomplicated	0.2221	0.021	<0.001
Hypothyroidism	0.114	0.065	<0.001
Obesity	0.240	0.064	<0.001
RenalFailure	0.026	0.006	<0.001
Congestive Heart Failure	0.025	0.006	<0.001

Table 2. Patient Characteristics for UKBiobank Data.

Feature	Statistics		P-value
	Case (n = 443)	Control (n = 2000)	
Age, Years	37.63	36.42	<0.001

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Race, Rate			
African American	0.061	0.034	0.007
Caucasian	0.867	0.881	0.409
Other	0.072	0.085	0.377
Social History, Rate			
Smoking	0.474	0.576	<0.001
Alcohol	0.864	0.917	<0.001
Pre-existing Disease, Occurrence Rate			
Complicated Diabetes	0.007	0.009	0.645
Hypertension Uncomplicated	0.007	0.001	<0.001
Hypothyroidism	0.007	0.000	<0.001
Obesity	0.000	0.017	<0.001
RenalFailure	0.000	0.000	NA
Congestive Heart Failure	0.002	0.000	<0.001

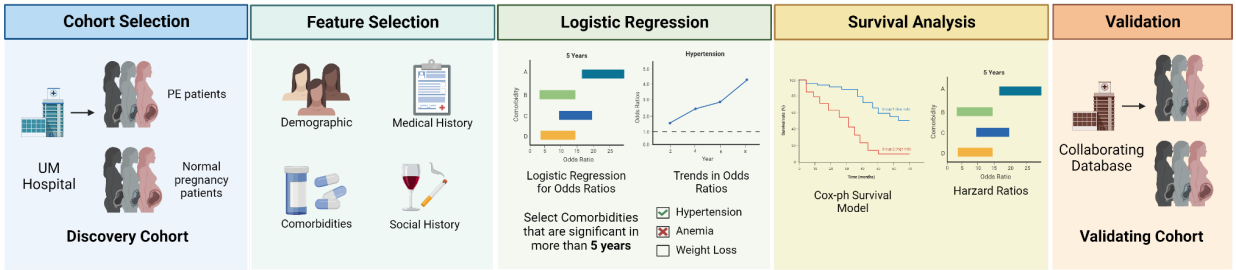


Figure 1. Project Workflow.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

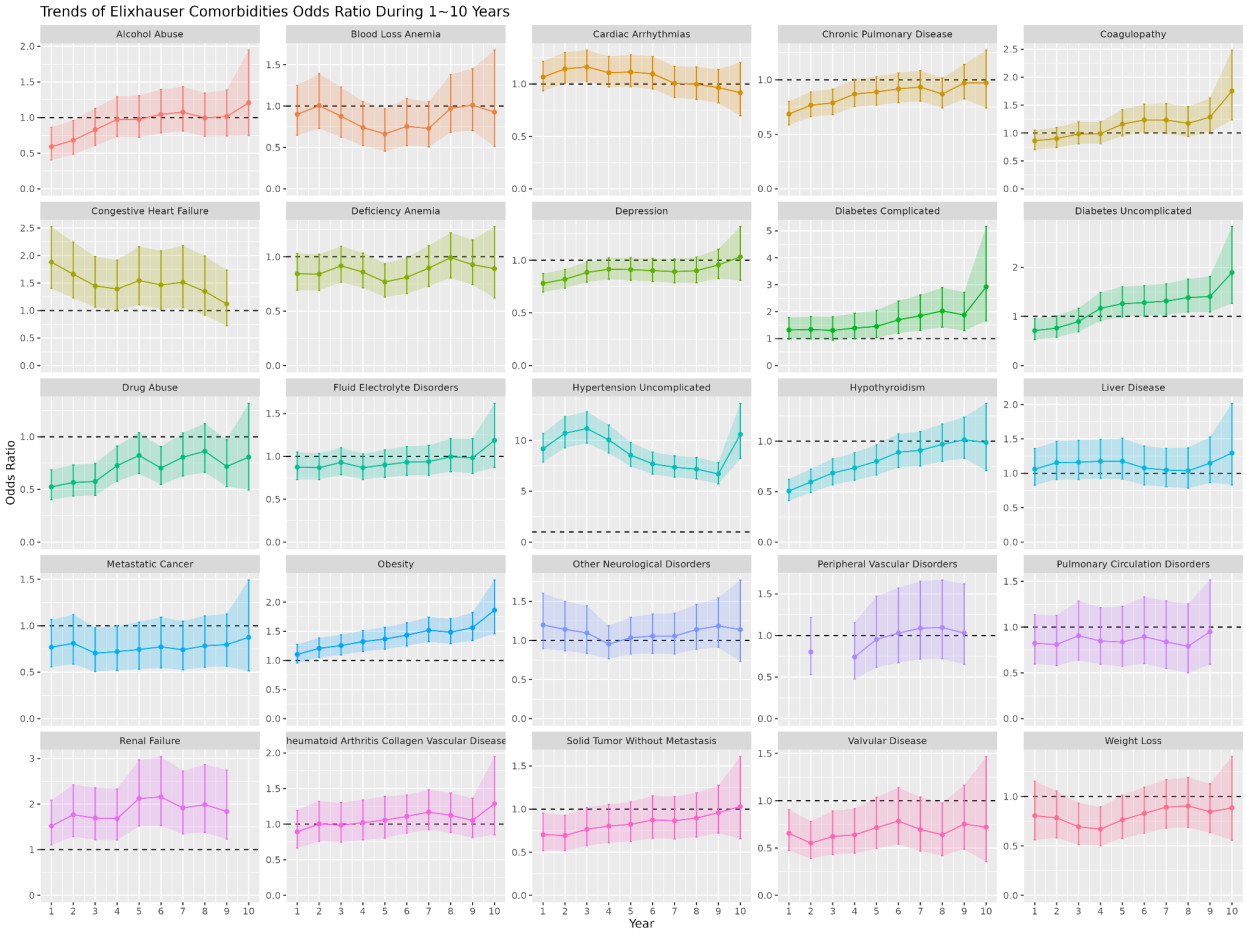


Figure 2. Odds Ratios (OR) Trends of All Elixhauser Comorbidities. This figure shows the change in OR of different comorbidities within 1~10 years after the *underlying diagnosis*. For some years certain comorbidities have inadequate effective size for analysis, causing missing points in the plot. The dashed lines are plotted at OR = 1, where ORs above the line indicate higher risks of comorbidities due to PE, while ORs below the line indicate lower risks of the comorbidities due to PE. Confidence intervals of the OR are indicated by the colored shadow areas.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

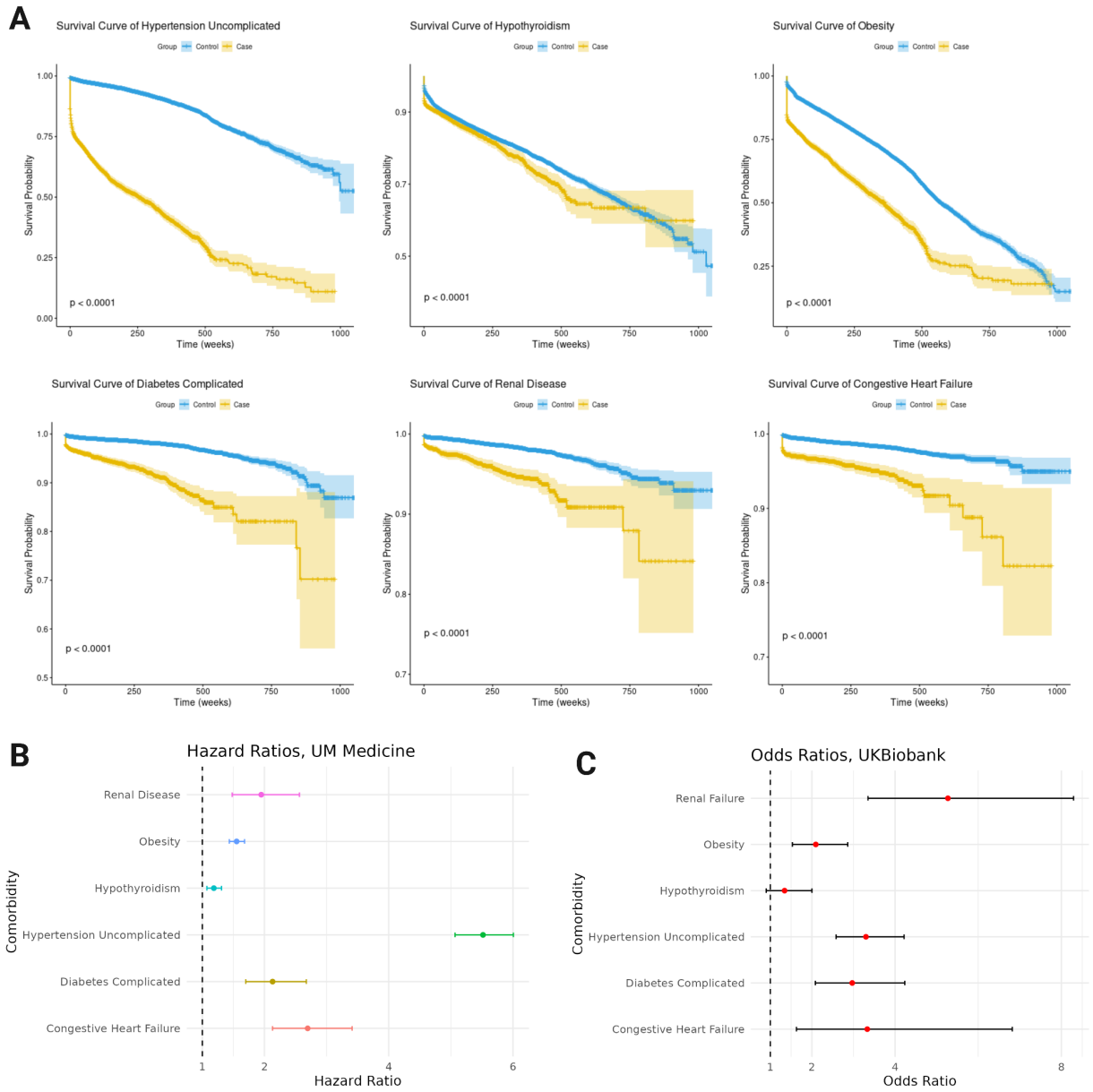


Figure 3 (A) Survival Curves of Significant Comorbidities of PE. The p-values were calculated from log-rank tests comparing the survival time of case and control groups. All p-values were smaller than 0.0001, demonstrating statistical significance. **(B) Hazard Ratios of Significant Comorbidities of PE.** Hazard ratios adjusted for competing risks were calculated using the "tidycmprsk" package. All hazard ratios were greater than 1, indicating increasing risks of comorbidities due to PE. Compared to women without a PE-related history, the HR due to PE's effect is 5.70 (95% CI: 5.29-6.14) for uncomplicated hypertension, 2.22 (95% CI: 1.84-2.67) for complicated diabetes, 2.90 (CI: 2.32-3.62) for congestive heart failure, 1.90 (CI: 1.50-2.41) for renal disease, 1.68 (CI: 1.58-1.79) for obesity, and 1.22 (CI: 1.11-1.33) for hypothyroidism. **(C) Validation of OR of Significant Comorbidities using UKBiobank Data.** The significant comorbidities identified in the UM database were validated using UKBiobank

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

data. Renal failure (OR = 5.27, 95% CI: 3.35-8.29), congestive heart failure (OR = 3.33, 95% CI: 1.63-6.82), uncomplicated hypertension (OR = 3.30, 95% CI: 2.58-4.22), complicated diabetes (OR = 2.97, 95% CI: 2.08-4.23), obesity (OR = 2.09, 95% CI: 1.53-2.86), hypothyroidism (OR = 1.34, 95% CI: 0.90-2.00).

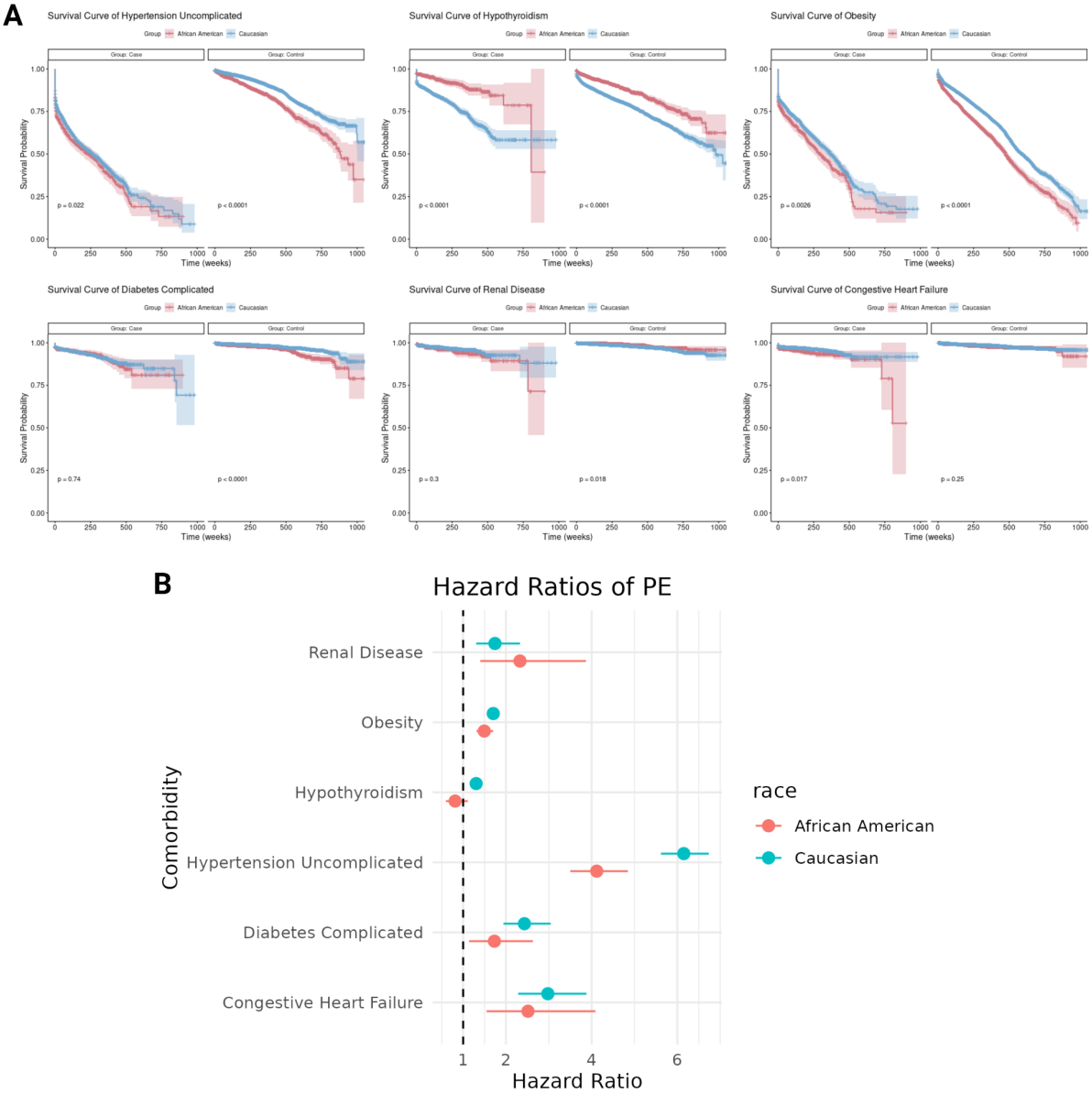
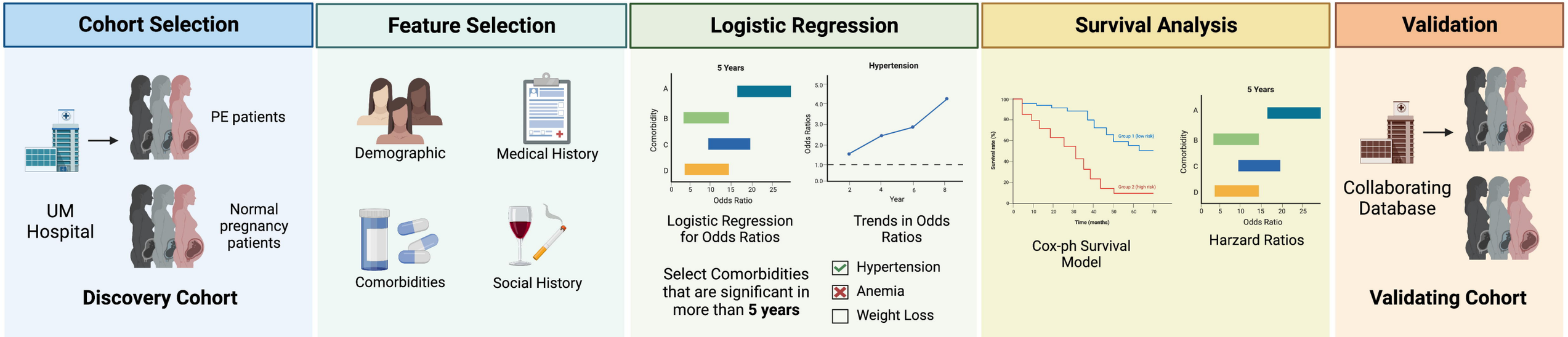
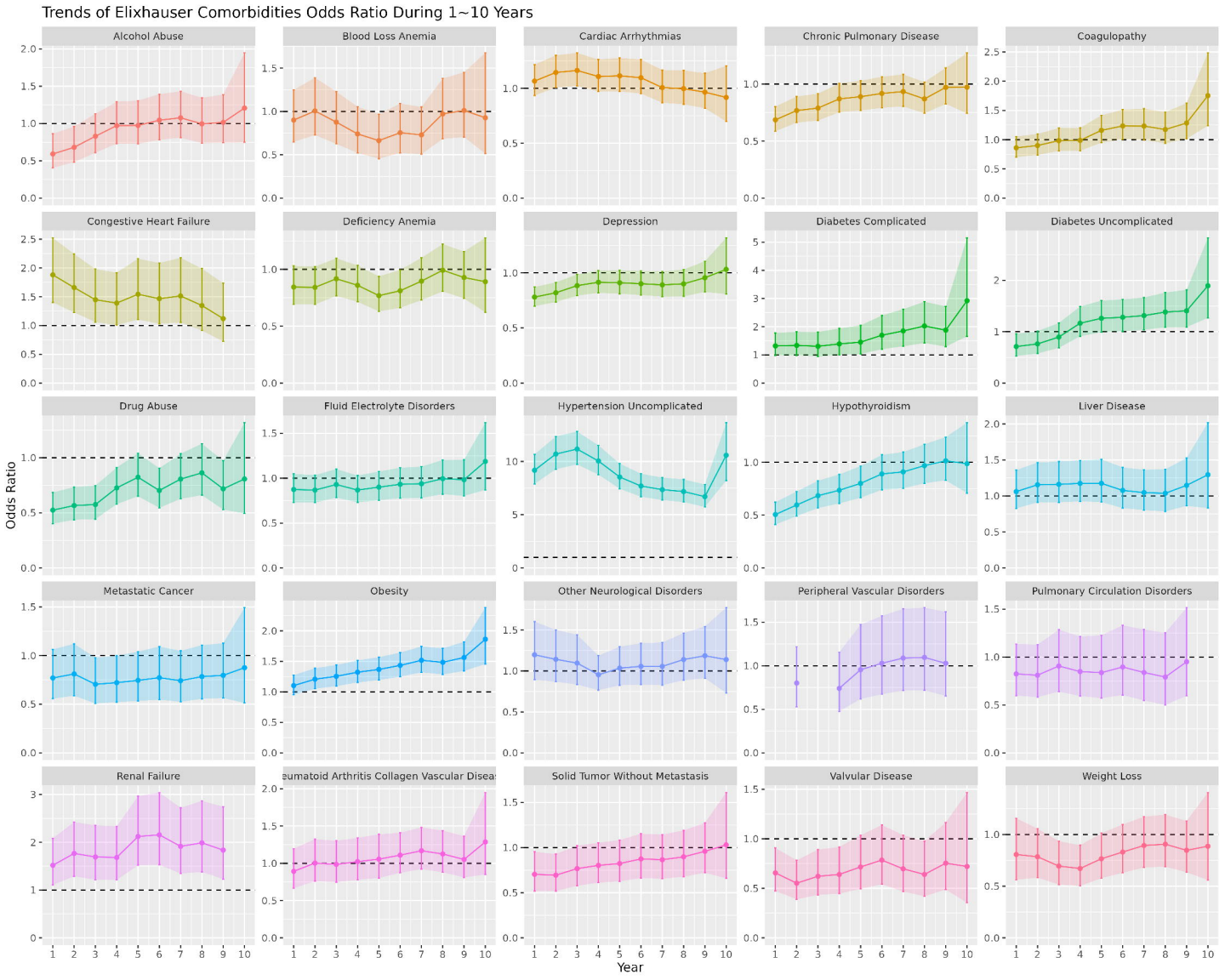


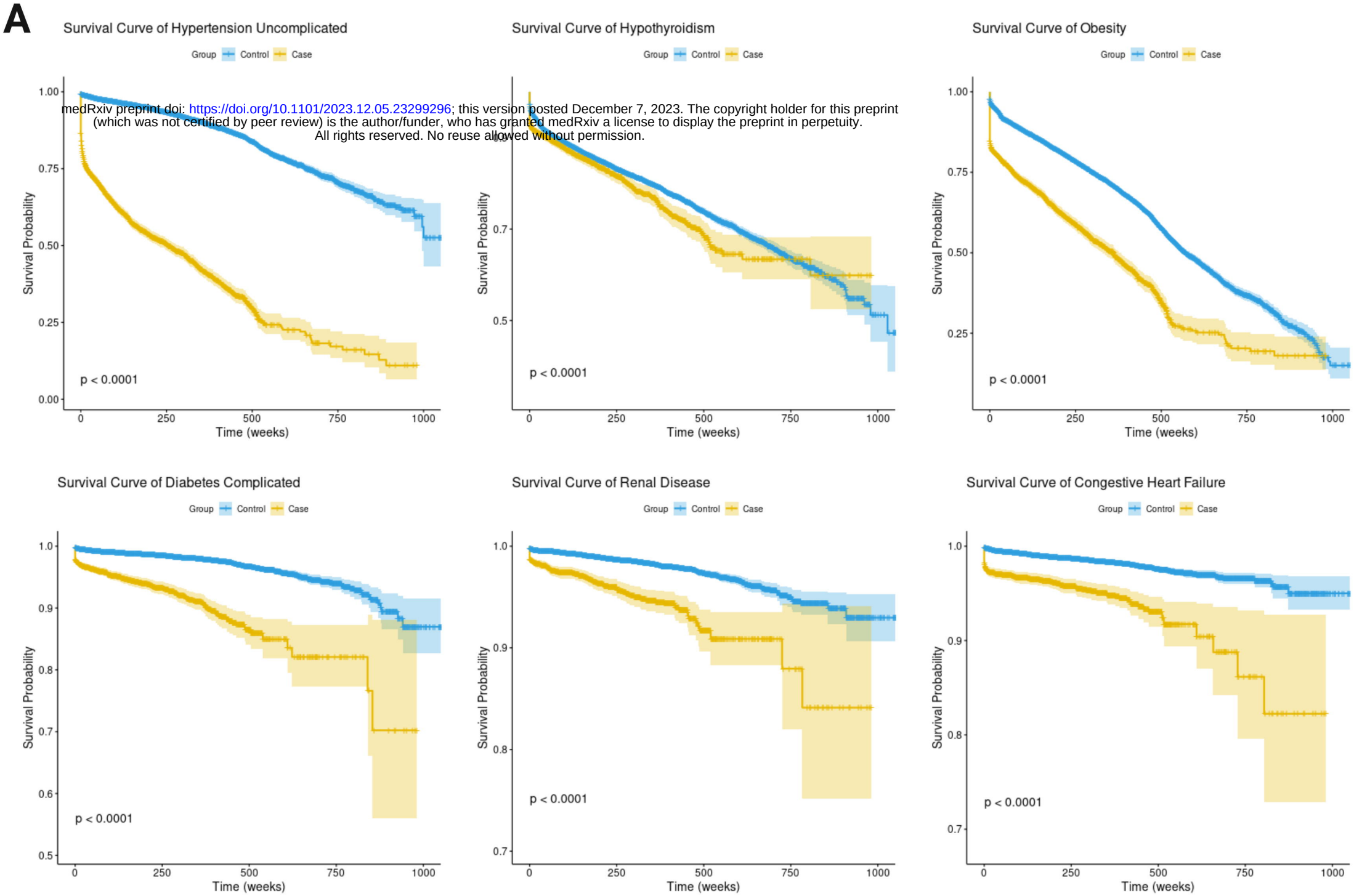
Figure 4 (A). Survival Curves of Comorbidities w.r.t. African American and Caucasian. The p-values were calculated from log-rank tests comparing the survival time of Caucasians and African Americans in case and control groups, respectively. **(B). Hazard Ratios of Comorbidities due to PE w.r.t. African American and Caucasian.** Hazard ratios adjusted for competing risk were calculated using the “tidycmprsk” package.


Cohort Selection
PE patients
UM Hospital
Normal pregnancy patients
Discovery Cohort
Feature Selection
Demographic
Medical History
Comorbidities
Social History
Logistic Regression
5 Years
Comorbidity
A
B
C
D
0 5 10 15 20 25
Odds Ratio
Hypertension
Odds Ratios
5.0
4.0
3.0
2.0
1.0
0.0
2 4 6 8
Year
Logistic Regression for Odds Ratios
Trends in Odds Ratios
Select Comorbidities that are significant in more than 5 years
Hypertension
Anemia
Weight Loss
Survival Analysis
Survival rate (%)
100
80
60
40
20
0
0 10 20 30 40 50 60 70
Time (months)
Group 1 (low risk)
Group 2 (high risk)
5 Years
Comorbidity
A
B
C
D
0 5 10 15 20 25
Odds Ratio
Cox-ph Survival Model
Harzard Ratios
Validation
Collaborating Database
Validating Cohort

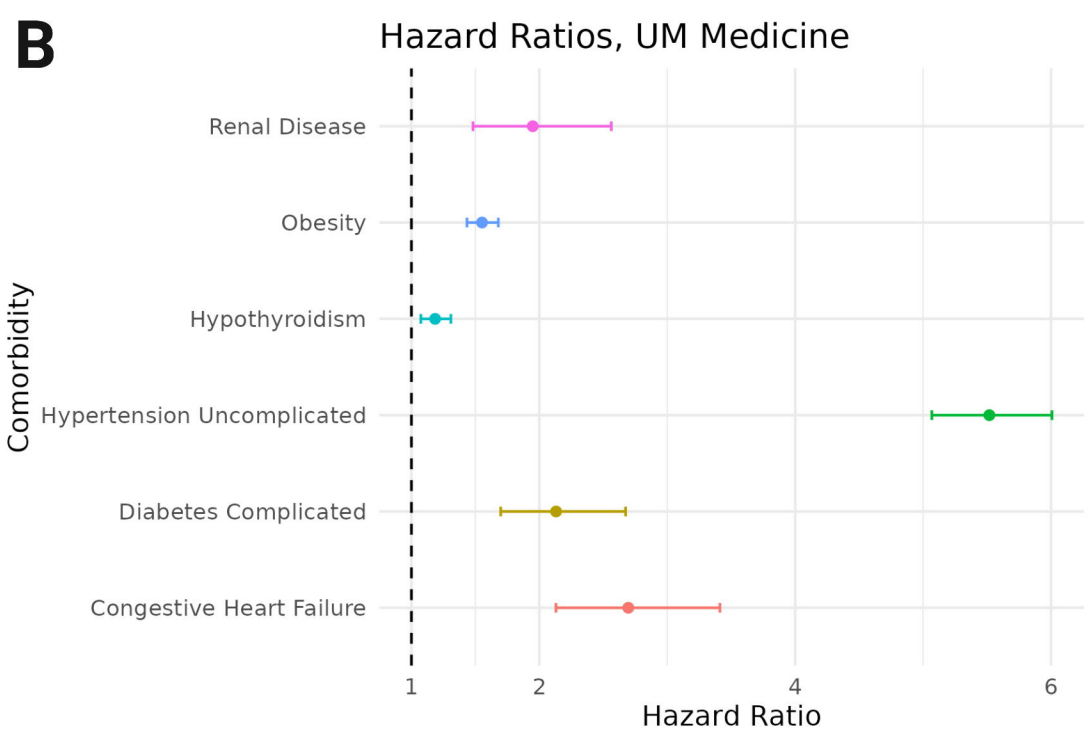

Trends of Elixhauser Comorbidities Odds Ratio During 1~10 Years
Alcohol Abuse
Blood Loss Anemia
Cardiac Arrhythmias
Chronic Pulmonary Disease
Coagulopathy
Congestive Heart Failure
Deficiency Anemia
Depression
Diabetes Complicated
Diabetes Uncomplicated
Drug Abuse
Fluid Electrolyte Disorders
Hypertension Uncomplicated
Hypothyroidism
Liver Disease
Metastatic Cancer
Obesity
Other Neurological Disorders
Peripheral Vascular Disorders
Pulmonary Circulation Disorders
Renal Failure
eumatoid Arthritis Collagen Vascular Disea
Solid Tumor Without Metastasis
Valvular Disease
Weight Loss
Odds Ratio
Year

medRxiv preprint doi: https://doi.org/10.1101/2023.12.05.23299296; this version posted December 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

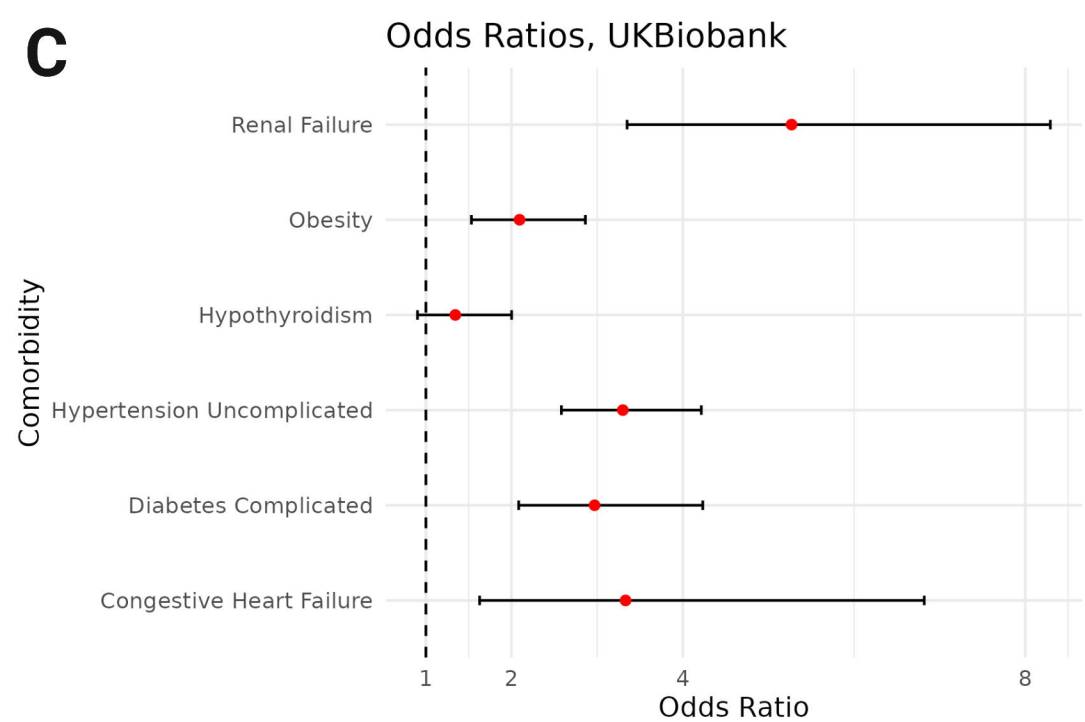
A

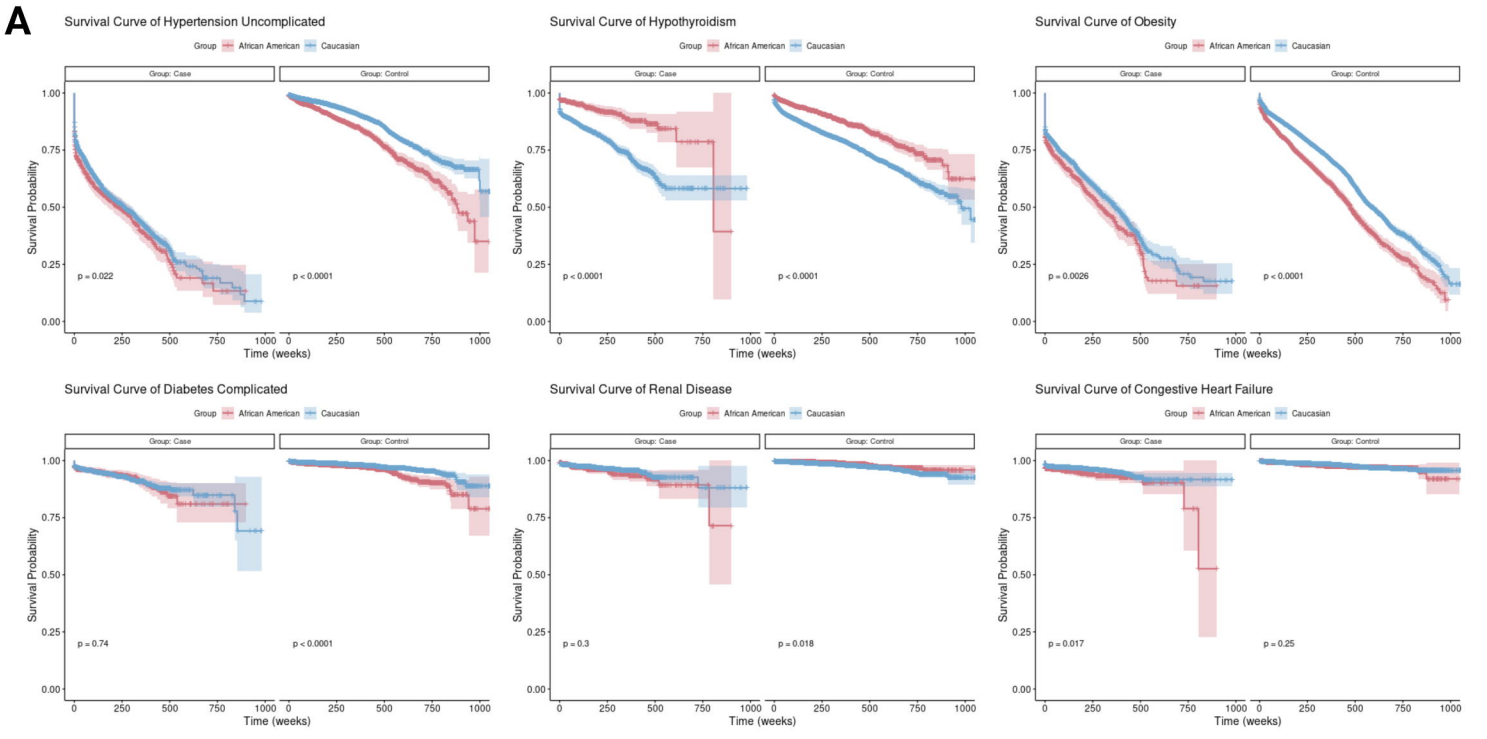


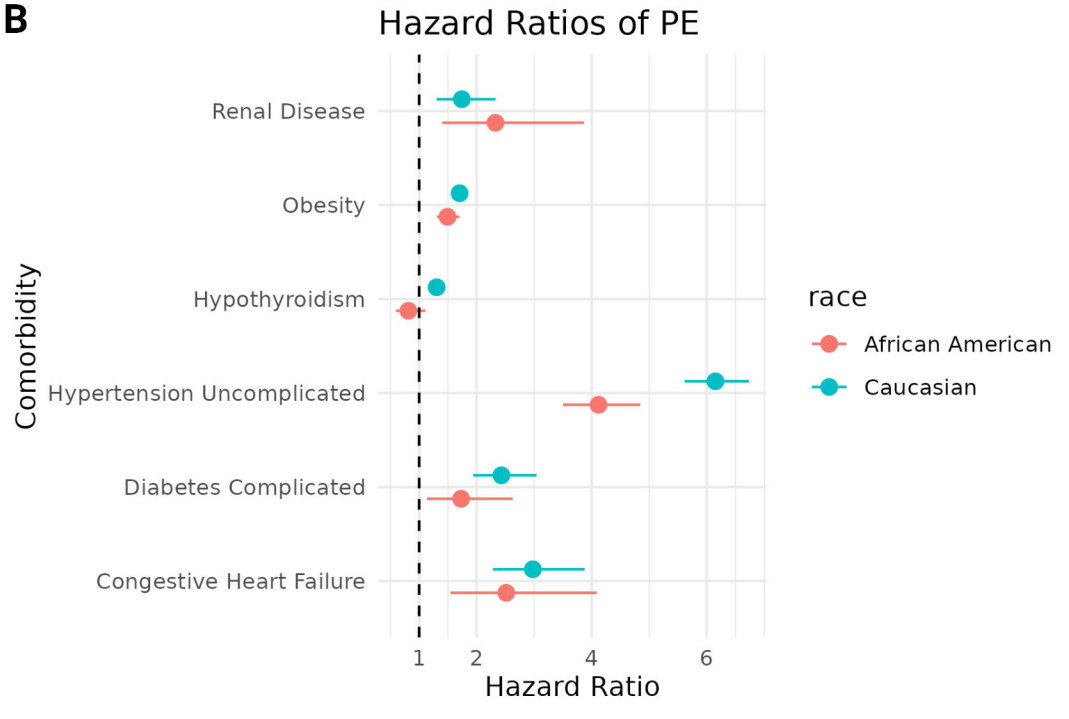
B



C




A
Survival Curve of Hypertension Uncomplicated
Survival Curve of Hypothyroidism
Survival Curve of Obesity
Survival Curve of Diabetes Complicated
Survival Curve of Renal Disease
Survival Curve of Congestive Heart Failure
Group
African American
Caucasian
Group: Case
Group: Control
Survival Probability
Time (weeks)
p = 0.022
p < 0.0001
p < 0.0001
p < 0.0001
p = 0.0026
p < 0.0001
p = 0.74
p < 0.0001
p = 0.3
p = 0.018
p = 0.017
p = 0.25


B
Hazard Ratios of PE
Renal Disease
Obesity
Hypothyroidism
Hypertension Uncomplicated
Diabetes Complicated
Congestive Heart Failure
Comorbidity
1
2
4
6
Hazard Ratio
race
African American
Caucasian