Impact of Metformin on Immunotherapy Outcomes among Breast Cancer Patients with Type 2 Diabetes in the United States: A Retrospective Cohort Study

Zihui Chen[1], Lejing Li[1], Yael Yossefy[1], Kevin Kensler[1], Laura Pinheiro[1]

[1] Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

Group Members: Zihui Chen, Lejing Li, Yael Yossefy, Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

Faculty Advisor: Xi Kathy Zhou, PhD, Professor of Research in Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

Client: Laura Pinheiro, PhD, Associate Professor of Health Services Research in Medicine, and Kevin Kensler, ScD, Assistant Professor of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

1. Abstract

Introduction: Breast cancer is the second leading cause of cancer morbidity and mortality among women, with an estimated 316,950 new U.S. cases in 2025. Approximately 29% of women with breast cancer have type 2 diabetes (T2D). Metformin, the most common first-line therapy for T2D, may enhance immune responses and improve outcomes of immune checkpoint inhibitors such as pembrolizumab. However, evidence in breast cancer patients with T2D undergoing immunotherapy remains limited. We examined whether metformin use improves overall survival (OS) among patients with T2D and breast cancer treated with pembrolizumab.

Methods: We conducted a retrospective cohort study using de-identified TriNetX electronic health records. Eligible patients were women diagnosed with both T2D and breast cancer from 2019 onward who received pembrolizumab. Metformin exposure was defined as an active prescription prior to pembrolizumab initiation. OS, measured from pembrolizumab initiation, was analyzed using Kaplan–Meier and Cox proportional hazards models adjusting for demographics, treatments, and comorbidities. Secondary analyses examined whether baseline glycemic control and metastatic status modified associations.

Results: Among 483 patients: 128 (26.5%) were metformin users and 355 (73.5%) were non-users. The groups were generally similar, though metformin users had a higher prevalence of hypertension (84% vs 67%, $p<0.001$) and hyperlipidemia (70% vs 53%, $p=0.001$), and slightly higher median HbA1c (6.65 vs 6.15, $p=0.018$). Metformin users had higher OS than non-users

(log-rank p=0.028), especially with >6.6 years of metformin use (log-rank p=0.048). In adjusted Cox models, metformin use was not significantly associated with OS (HR=0.69, p=0.16), but each additional year of metformin use reduced death hazard by ~9%(HR=0.91, p=0.052), though this trend did not reach statistical significance. Among metastatic pateints, longer use was associated with better survival (HR=0.89, p=0.049). Glycemic control analyses were limited by missing data (68%).

Conclusion: Metformin use, particularly long-term, was associated with improved OS in unadjusted analyses but not significantly after adjustment. Within the subgroup of patients whose breast cancer metastasized, longer metformin use seemed to be associated with improved survival. Prospective studies are needed to clarify the role of metformin and glycemic control in immunotherapy outcomes.

2. Introduction

Breast cancer is the second most common cancer among women in the United States and the second leading cause of cancer-related mortality in this population [10]. In 2025, the estimated number of new breast cancer cases is 316,950 and the estimated number of deaths due to breast cancer is 42,170 [11]. The lifetime risk of a U.S. woman developing breast cancer is approximately **13%**, and incidence rates have increased by about **1%** per year, rising even faster (**1.4%**) in women younger than 50 years, during the most recent decade of data (2012-2021) [12].

Studies have shown that Type II Diabetes is associated with a significantly increased risk of overall and some site-specific cancers [5], including breast cancer [6], as well as an increased

risk in cancer mortality [7]. A meta-analysis of thirty-nine studies found that the risk of breast cancer in women with type 2 diabetes is increased by 27% compared to the general population (RR: 1.27, 95% CI: 1.16-1.39) [6]. A cross-sectional study of patients with invasive breast cancer diagnosed in 2000–2010, found that premenopausal women with diabetes tended to have more often PR-negative (OR: 2.44, 95% CI: 1.07–5.55) and HER2-negative (OR: 2.84, 95% CI: 1.11–7.22) tumors than the women without diabetes [17]. A retrospective case-control study of Louisiana Tumor Registry records of primary invasive breast cancer diagnosed in 2010–2015 found a significant association between diabetes and TNBC (OR: 1.82, 95% CI: 1.38-2.39) [18]. In addition, a U.S. population-based study of cancer mortality from 1988 to 2015 found that adults with diabetes were approximately 30% more likely to die from cancer than those without diabetes [15]. Metformin is currently the most widely prescribed drug for type 2 diabetes in the United States. Its main functions include lowering blood sugar levels by reducing the absorption of glucose in the gastrointestinal tract and inhibiting the production of glucose in the liver [9]. Preclinical studies further suggest that metformin may support anti-tumor immunity by strengthening the activity of immune effector cells [3], while reducing immunosuppressive influences in the tumor microenvironment. Importantly, metformin has also been shown to enhance the therapeutic effects of immune checkpoint inhibitors targeting the PD-1/PD-L1 axis [4].

Over the last decade, there has been growing interest in the potential anti-cancer properties of metformin [1]. Observational studies and meta-analyses suggest that metformin use is associated with improved cancer survival, with pooled hazard ratios for overall survival of approximately 0.81 (95% CI, 0.77–0.85) [2].

Given the increasing role of immunotherapy in breast cancer, particularly in patients with aggressive subtypes, pembrolizumab—a programmed cell death protein-1 (PD-1) immune checkpoint inhibitor—has become an established therapeutic option. It is approved for use in high-risk early-stage triple-negative breast cancer (TNBC) in combination with chemotherapy as neoadjuvant treatment, followed by adjuvant pembrolizumab monotherapy, and for locally recurrent unresectable or metastatic TNBC with PD-L1 expression (CPS ≥10) [13].

In early-stage TNBC, pembrolizumab is administered with chemotherapy for approximately 24 weeks (either 200 mg every 3 weeks or 400 mg every 6 weeks), followed by adjuvant pembrolizumab monotherapy for up to 27 weeks. In the advanced or metastatic setting, pembrolizumab may be continued until disease progression, unacceptable toxicity, or for up to 24 months [14]. By blocking the PD-1/PD-L1 pathway, pembrolizumab enhances anti-tumor immune responses, providing a significant therapeutic advance for this difficult-to-treat breast cancer subtype [14].

Existing preclinical and clinical data suggest that metformin may modulate the tumor microenvironment and enhance the activity of immune checkpoint inhibitors. Yet, robust evidence in large, real-world cohorts of breast cancer patients with T2DM is limited. This study addresses this gap by evaluating the association between metformin use and overall survival in female breast cancer patients with T2DM treated with pembrolizumab.

3. Methods

3.1 Data Source and Study Population

Data source: We used data from TriNetX, which is a global health research network of real-

world clinical data that is aggregated and harmonized directly from healthcare organizations on a continuous basis. TriNetX has partnered with healthcare organizations spanning 30 countries and thousands of sites, including a mix of hospital, primary care, and specialty treatment providers, providing de-identified, HIPAA and GDPR compliant data on 117 million U.S. patients. [8]

For our study, we received 24 relational TriNetX datasets, with data from U.S. EMR systems, cancer registries, insurance claims, mortality sources, and other sources. The data included tables called Patient Demographics, Patient Cohort, Procedure, Diagnosis, Medication Drug, Medication Ingredient, Encounter, Lab Results, Tumor, Chemo Lines, Claims Lines, and more. The datasets were provided to us on April 24th, 2025 and included data that had been regularly updated in 1, 2, or 4-week intervals starting from January 1st, 2019 up until April 24th, 2025.

Eligibility criteria: We initially identified **18,130 patients** in the United States who met the following conditions: (i) a documented diagnosis of type 2 diabetes and (ii) a subsequent diagnosis of breast cancer on or after January 1, 2019 and before April 24th, 2025. We then applied additional inclusion criteria: (iii) documented pembrolizumab treatment following the earliest breast cancer diagnosis, (iv) availability of birth year, and (v) a valid death date (only applicable for those patients that had a record of death). Specifically, if a patient had a recorded death date, it must have occured after the treatment and exposure dates.

Refining our sample: To capture receipt of immunotherapy, relevant codes containing the term

pembrolizumab were retrieved from the Standardized Terminology table. The primary codes were RxNorm 1547545 and HCPCS J9271 (Injection, pembrolizumab, 1 mg). These codes were applied to filter both the Procedures and Medication Ingredients tables for patients with a record of pembrolizumab treatment. This step yielded 536 distinct patients with documented pembrolizumab treatment in the Medication Ingredients table, 493 of which also had a record of pembrolizumab treatment in the Procedures table.

To investigate the discrepancy, we further explored the treatment histories of those 43 patients with only a record of pembrolizumab in the Medication Ingredients table but not in the Procedures table. Using the Chemo Lines table, we discovered that **all 43 patients** had chemotherapy record and had their pembrolizumab use date within ±7 days of a system-recorded chemotherapy line. This temporal overlap strongly suggests that pembrolizumab was part of the chemotherapy regimen, despite the absence of a corresponding HCPCS code (J9271) in the Procedure Table. Thus, we decided to include **all 536 patients** with a record of pembrolizumab in either the Medication Ingredient or Procedure table in our cohort. [Fig. 1]

Data management: To ensure our cohort only includes instances of pembrolizumab treatment for breast cancer as opposed to other cancer types for which pembrolizumab is an approved treatment, we restricted the cohort to patients whose pembrolizumab initiation occurred after their earliest breast cancer diagnosis. This reduced the cohort to **497 patients**. Patients without a recorded year of birth were excluded (n = 3), leaving **494 patients**. [Fig. 1]

We then conducted consistency checks on mortality data. TriNetX ascertains mortality data from EHR records and Death Registries, and the source of patients' death date is recorded in

the death_date_source_id variable of the patient demographic table. Patients with a recorded death date earlier than their last record of pembrolizumab or metformin treatment were considered implausible and excluded (n = 11). The final analytic sample consisted of **483 patients.** [Fig. 1]

Thus, the final dataset included **483 female patients** with **type 2 diabetes** and **breast cancer**, all of whom received **pembrolizumab treatment after 2019**, with reliable demographic and mortality data for survival analysis.

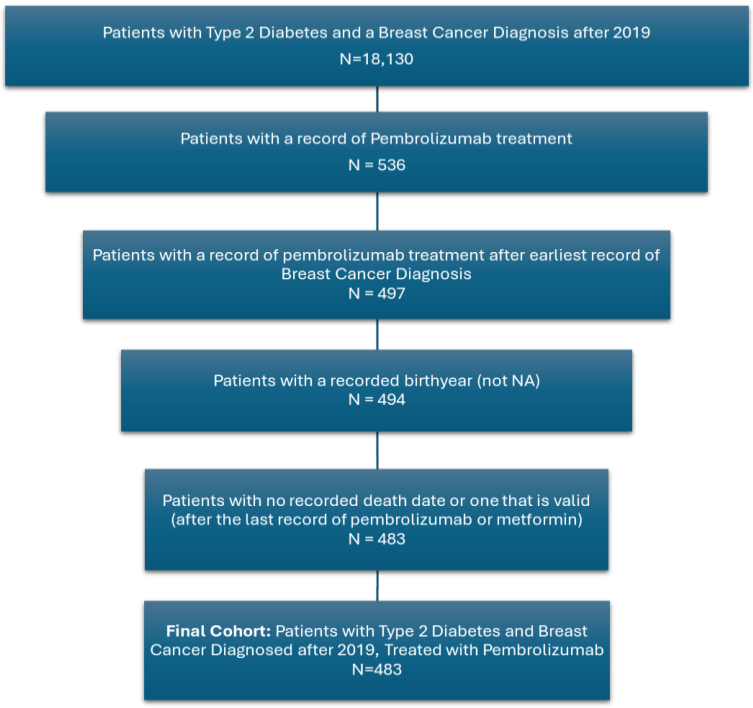

Figure 1: Exclusion cascade

3.2 Variables

Key explanatory variable: Metformin use was defined as the independent variable of our study. It was captured from the medication ingredients and procedure tables. All drug codes [**Table 6**] corresponding to metformin were identified through the standardized terminology table. For each patient, the earliest and latest prescription dates were recorded as the metformin start date and metformin end date. Using pembrolizumab initiation as the index date, patients were

classified as metformin users if the initiation date occurred between their first and last metformin prescription dates. A binary indicator (any use vs. no use) was then created to distinguish users from non-users. The exposure window assumed continuous use between the earliest and latest dates, acknowledging potential misclassification if treatment was interrupted.

Primary Outcome: Our dependent variable was overall survival. It was operationalized as the time from pembrolizumab initiation (time zero) to death from any cause. Patients without a recorded death were considered censored. Death events were obtained from cohort death records in the patient demographics table. For censored patients, the last observed medication activity date was used as the censoring time. Overall Survival was calculated as the number of days from pembrolizumab initiation to either death event or the last medication date, which was called the censoring date, whichever occurred first. This assumes that medication activity reflects continued follow-up, with potential risk of informative censoring.

Covariates: We adjusted for potential confounders selected on the basis of clinical relevance and prior literature. These included age at pembrolizumab initiation, year of initiation, and either metformin use (yes vs. no) or the duration of metformin use at pembrolizumab initiation. Clinical covariates comprised prior surgery, prior radiation, insulin use at initiation, and number of other diabetes medication classes used at pembrolizumab initiation. In addition, baseline health status was captured by the number of comorbidities diagnosed (including obesity, hypertension, hyperlipidemia, liver disease, and kidney disease) in the year prior to pembrolizumab initiation. All covariates were assessed prior to or at pembrolizumab initiation.

3.3 Statistical Analysis

- **Descriptive Statistics:**

Baseline demographic and clinical characteristics were summarized overall and stratified by metformin use (any vs. none). Continuous variables were reported as means with standard deviations or medians with interquartile ranges, depending on distribution. Categorical variables were summarized as counts and percentages. Between-group comparisons were performed using t-tests or Wilcoxon rank-sum tests for continuous variables and Chi-square or Fisher's exact tests for categorical variables, as appropriate.

- **Survival Analysis:**

Kaplan–Meier methods were used to estimate survival distributions, and log-rank tests were applied to compare OS between metformin users and non-users. Additional Kaplan–Meier analyses were conducted after stratifying metformin exposure by duration prior to pembrolizumab initiation into tertiles of nearly equivalent sizes.

Multivariable Cox proportional hazards regression models were constructed to estimate hazard ratios (HRs) with 95% confidence intervals (CIs) for the association between either metformin use or the duration of metformin use and OS. Models adjusted for demographic factors (age, year of initiation), treatment variables (surgery, radiation, chemotherapy, endocrine therapy, insulin use, other classes of diabetes medications), and clinical covariates (body mass index, comorbidities). The proportional hazards assumption was evaluated using

Schoenfeld residuals. No significant violations were observed overall (global test p = 0.30). All covariates satisfied the proportional hazards assumption, except for the number of comorbidities, which showed a minor deviation (p = 0.04) but was not considered to meaningfully affect model validity.

All statistical tests were two-sided, and a p-value < 0.05 was considered statistically significant.

- **Subgroup Analyses:**

The objective of this analysis was to evaluate whether baseline glycemic control modifies the association between metformin exposure and overall survival (OS). Baseline HbA1c values were ascertained using LOINC 4548-4, and for each patient the most recent value prior to pembrolizumab initiation was selected. Patients were stratified according to glycemic control into two categories: HbA1c < 7% and HbA1c ≥ 7%.

Missing values for HbA1c were defined as the absence of any qualifying LOINC 4548-4 record before pembrolizumab initiation, or the presence of non-numeric or implausible entries, including HbA1c values ≤ 3% or ≥ 20%. To assess the distribution of patients across the HbA1c strata and exposure groups, an overlap table (Figure 4) was constructed. This table summarized the cross-classification of HbA1c strata by metformin exposure, reporting the number of patients in each cell along with within-stratum percentages. Pre-specified positivity thresholds were applied, requiring a minimum of 20 patients and a minimum of 5% representation for each exposure level within each HbA1c stratum.

For the primary statistical test, the Wald test was used to evaluate whether the interaction term between metformin exposure and HbA1c stratum equaled zero. This tested the hypothesis that baseline blood glucose control modifies the relationship between metformin use and OS. As a confirmatory analysis, a likelihood ratio test (LRT) was conducted to compare nested Cox proportional hazards models, one including the interaction term 'metformin × HbA1c' and one without it.

The decision rule prespecified for this analysis was based on the interaction p-value. If the interaction p-value exceeded 0.05, it was interpreted as indicating some evidence that the association between metformin use and OS differed by HbA1c stratum. Results included estimates of the interaction term from the Cox model as well as Kaplan–Meier curves stratified by glycemic control, which were used to visually assess survival patterns across strata.

- **Additional Stratified Analysis by Metastatic Status:**

To further assess the robustness of findings, an additional stratified analysis was performed by metastatic status at pembrolizumab initiation (metastatic vs. non-metastatic). Separate Kaplan–Meier survival curves were generated for each subgroup, and log-rank tests were applied within each stratum to compare OS between metformin users and non-users.

This analysis aimed to evaluate whether disease stage modified the observed association between metformin use and OS. In addition, stratified Cox proportional hazards models were fitted within each subgroup, and an interaction term (metformin × metastatic) was tested in the overall multivariable Cox model to formally evaluate whether the association between

metformin use and OS differed by metastatic status.

4. Results

4.1 Descriptive Statistics

Baseline demographic and clinical characteristics of the 483 eligible patients (128 metformin users and 355 non-users) are summarized in **Table 1**. The median age at pembrolizumab initiation was 60 years in both groups (p=0.4). Age at breast cancer diagnosis was also similar (59 vs 60 years, p=0.4). Among metformin users, the median duration of metformin use prior to pembrolizumab initiation was 4.8 years (IQR 2.1–7.3).

Metformin users remained on pembrolizumab for a significantly longer period compared with non-users (median 268 vs 154 days, p=0.009). Rates of surgery, chemotherapy, and radiation did not differ significantly between groups. Metformin users had a higher prevalence of hypertension (84% vs 67%, p<0.001) and hyperlipidemia (70% vs 53%, p=0.001), while obesity rates were similar across groups (at 42%).

HbA1c measurements prior to pembrolizumab initiation were available for 156 of the 483 patients (32%). Among these, metformin users exhibited slightly higher HbA1c values compared with non-users (median 6.65% vs 6.15%, p=0.018). Use of 2+ additional diabetes medications of different classes was more common among metformin users (25% vs. 6%, p<0.001). Overall, 101 deaths (21%) occurred during the follow-up period, with a lower proportion observed among metformin users compared with non-users (16% vs 23%, p=0.086).

Table 1. Patient Characteristics by Metformin Use

Characteristic	Overall N = 483[1]	No N = 355[1]	Yes N = 128[1]	p-value[2]
Age at Pembrolizumab Initiation	60 (53, 66)	60 (52, 67)	60 (53, 64)	0.4
Age at Diagnosis	59 (52, 66)	60 (51, 66)	59 (53, 64)	0.4
Age at Metformin Initiation	54 (48, 60)	NA (NA, NA)	54 (48, 60)	
Unknown	355	355	0	
Year of Pembrolizumab Initiation				0.12
2019	9 (1.9%)	7 (2.0%)	2 (1.6%)	
2020	17 (3.5%)	10 (2.8%)	7 (5.5%)	
2021	57 (12%)	46 (13%)	11 (8.6%)	
2022	158 (33%)	106 (30%)	52 (41%)	
2023	156 (32%)	116 (33%)	40 (31%)	
2024	85 (18%)	69 (19%)	16 (13%)	
2025	1 (0.2%)	1 (0.3%)	0 (0%)	
Duration of metformin use before pembro (Years)	4.8 (2.1, 7.3)	NA (NA, NA)	4.8 (2.1, 7.3)	
Unknown	355	355	0	
Time on Pembrolizumab (Days)	175 (51, 343)	154 (49, 336)	268 (84, 364)	0.009
Surgery (yes/no)	245 (51%)	183 (52%)	62 (48%)	0.5
Chemotherapy (yes/no)	483(100%)	355 (100%)	128 (100%)	0.00
HbA1c Prior to Pembro (%)	6.30 (5.70, 7.30)	6.15 (5.70, 7.10)	6.65 (6.10, 7.80)	0.018
Unknown	327 (67.7%)	235 (66.2%)	92(71.9%)	
Radiation Therapy (yes/no)	83 (17%)	63 (18%)	20 (16%)	0.6
Latest BMI	30 (26, 36)	30 (25, 36)	31 (27, 37)	0.4

Table 1. Patient Characteristics by Metformin Use

Characteristic	Overall N = 483[1]	No N = 355[1]	Yes N = 128[1]	p-value[2]
Unknown	214	153	61	
Tamoxifen/AI Status				0.4
Aromatase inhibitor only	18 (3.7%)	12 (3.4%)	6 (4.7%)	
Both	3 (0.6%)	2 (0.6%)	1 (0.8%)	
No endocrine therapy	456 (94%)	338 (95%)	118 (92%)	
Tamoxifen only	6 (1.2%)	3 (0.8%)	3 (2.3%)	
Insulin Use	62 (13%)	41 (12%)	21 (16%)	0.2
Other Diabetes Medication Classes				<0.001
0	350 (72%)	289 (81%)	61 (48%)	
1	80 (17%)	45 (13%)	35 (27%)	
2+	53 (11%)	21 (5.9%)	32 (25%)	
Obesity	203 (42%)	149 (42%)	54 (42%)	>0.9
Hypertension	347 (72%)	239 (67%)	108 (84%)	<0.001
Hyperlipidemia	278 (58%)	189 (53%)	89 (70%)	0.001
Liver Disease	27 (5.6%)	26 (7.3%)	1 (0.8%)	0.006
Kidney Disease	73 (15%)	58 (16%)	15 (12%)	0.2
Comorbidities Total				
0	71 (15%)	61 (17%)	10 (7.8%)	
1	102 (21%)	78 (22%)	24 (19%)	
2	152 (31%)	106 (30%)	46 (36%)	
3	114 (24%)	73 (21%)	41 (32%)	

Table 1. Patient Characteristics by Metformin Use

Characteristic	Overall N = 483[1]	No N = 355[1]	Yes N = 128[1]	p-value[2]
4	40 (8.3%)	33 (9.3%)	7 (5.5%)	
5	4 (0.8%)	4 (1.1%)	0 (0%)	
Metastatic	253 (52%)	188 (53%)	65 (63%)	0.7
Death Event	101 (21%)	81 (23%)	20 (16%)	0.086

[1] Median (Q1, Q3); n (%)

[2] Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

4.2 Overall Survival

Kaplan–Meier survival analysis revealed that metformin use was associated with statistically significantly improved overall survival compared with non-use. At 500 days, estimated survival probabilities were 82% among metformin users and 68% among non-users (log-rank p=0.028; **Figure 2**).

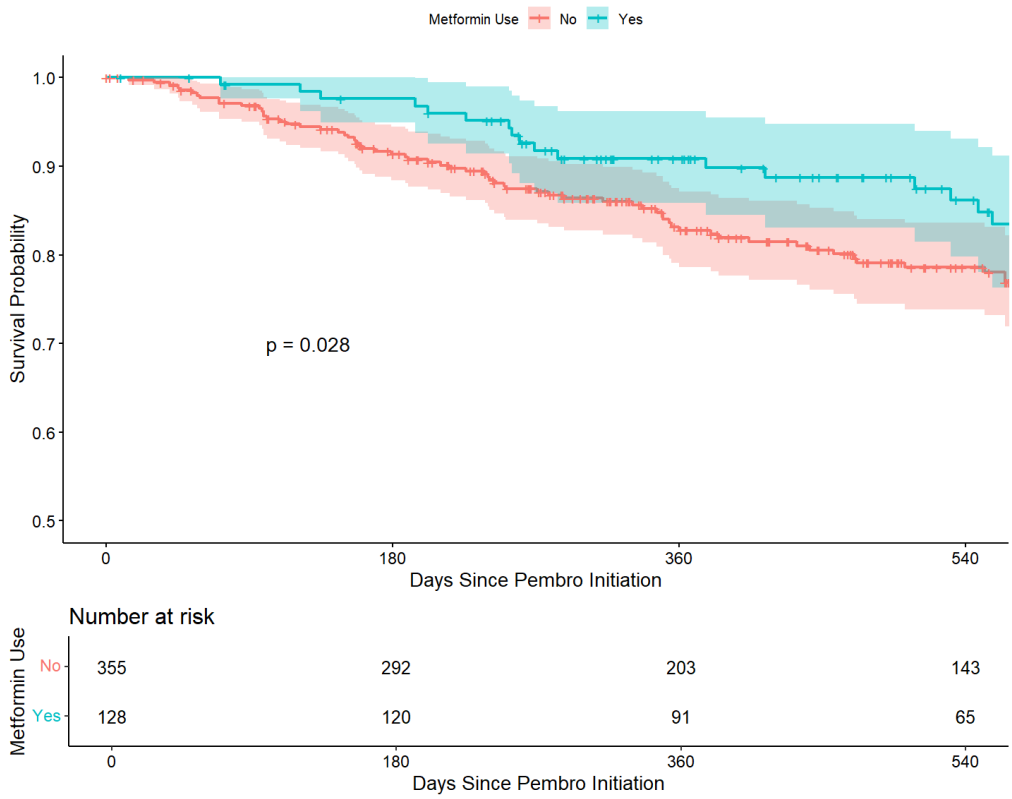

Figure 2: Overall survival by metformin use (Yes vs No)

When stratified by duration of metformin exposure, overall survival curves differed significantly across exposure groups (log-rank p=0.048; **Figure 3**). Patients with longer metformin use (>6.6 years and 2.61-6.6 years) appeared to have higher survival probabilities compared to non-users, although this global comparison does not establish a dose–response relationship. At 6 months, survival probabilities were **100%** for those with 2.61–6.6 and >6.6 years of prior use, compared with **91.4%** for non-users and **92.9%** for those with 0.1–2.6 years of metformin use. By 1 year, survival declined to **82.8%** in non-users, **88.1%** among short-duration users, and remained higher at **92.4%** and **92.0%** for the 2.61–6.6 and > 6.6 year groups, respectively. Sample sizes of the metformin exposure groups (n≈40) **[Figure 8]** limit the conclusions that can be drawn from these analyses.

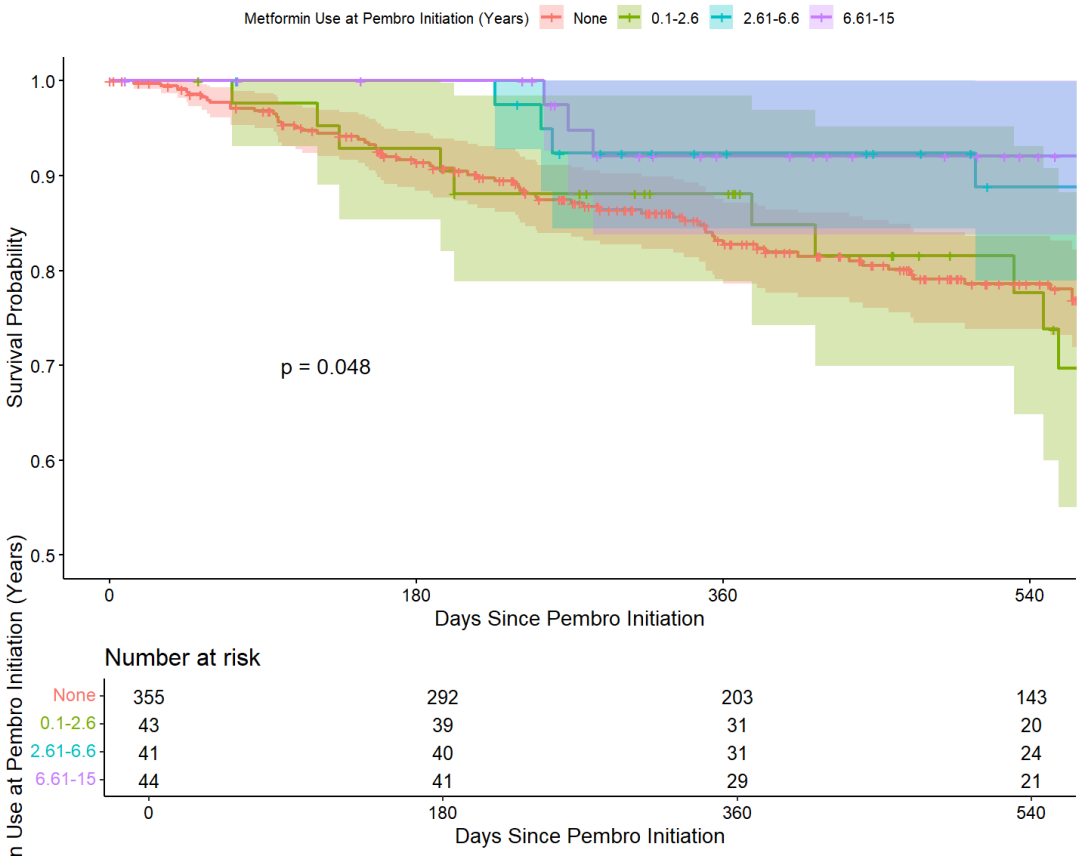

Figure 3: Overall survival by duration of metformin use

Table 2. Multivariable Cox proportional hazards model of survival among breast cancer patients with type 2 diabetes, with (Y/N) Metformin Use at Pembrolizumab Initiation (n = 483, events = 101).

Variable	Hazard Ratio	95% CI	p-value
Metformin use (yes vs. no)	0.69	0.41 – 1.16	0.16
Age at pembrolizumab initiation	1.02	1.00 – 1.04	0.13
Year of pembrolizumab initiation	0.94	0.78 – 1.14	0.53
Surgery (yes vs. no)	0.55	0.37 – 0.83	0.004**
Radiation (yes vs. no)	2.12	1.35 – 3.33	0.001**
Insulin use at pembrolizumab	1.97	0.88 – 4.40	0.10.
Number of comorbidities	1.12	0.94 – 1.33	0.20
Other diabetes medication classes (1)	0.91	0.46 – 1.78	0.77
Other diabetes medication classes (2+)	0.22	0.08 – 0.65	0.006**

Note: HR (hazard ratios); 95% CI (95% confidence intervals)

Among 483 patients (101 events), metformin use was not significantly associated with survival outcomes (HR = 0.69, 95% CI: 0.41–1.16, p = 0.16). Age at pembrolizumab initiation,

year of pembrolizumab initiation, and number of comorbidities were also not significantly associated with survival outcomes.

Patients who had undergone surgery experienced a 45% reduction in hazard of death compared with those who did not (HR = 0.55, 95% CI: 0.37–0.83, p = 0.004). In contrast, receipt of radiation therapy was associated with a more than two-fold increase in the hazard of death (HR = 2.12, 95% CI: 1.35–3.33, p = 0.001).

Insulin use at pembrolizumab initiation was associated with a trend toward higher hazard of death, though not statistically significant (HR = 1.97, 95% CI: 0.88–4.40, p = 0.10). Patients prescribed two or more other diabetes medication classes had significantly lower hazard compared with those on none (HR = 0.22, 95% CI: 0.08–0.65, p = 0.006).

Next, we'll look at a model with the continuous variable of metformin duration, in place of the binary metformin variable, including all the same covariates.

Table 3. Multivariable Cox proportional hazards model of survival among breast cancer patients with type 2 diabetes, with Duration of Metformin Use at Pembrolizumab Initiation (n = 483, events = 101).

Variable	Hazard Ratio	95% CI	p-value
Metformin Duration (Years)	0.91	0.83 – 1.00	0.052
Age at pembrolizumab initiation	1.02	1.00 – 1.04	0.12
Year of pembrolizumab initiation	0.95	0.78 – 1.14	0.57
Surgery (yes vs. no)	0.55	0.36 – 0.82	0.004
Radiation (yes vs. no)	2.19	1.40 – 3.43	<0.001*
Insulin use at pembrolizumab	1.97	0.89 – 4.33	0.09
Number of comorbidities	1.12	0.94 – 1.32	0.21
Other diabetes medication classes (1)	0.91	0.48 – 1.86	0.78
Other diabetes medication classes (2+)	0.23	0.08 – 0.68	0.007*

Note: HR (hazard ratios); 95% CI (95% confidence intervals)

In this model, we still have 483 patients and 101 events. Similarly to the last model, age at pembrolizumab initiation, year of pembrolizumab initiation, and number of comorbidities were also not significantly associated with survival outcomes. The results for surgery, radiation, insulin use, and number of other diabetes medication classes were also similar to the last model. Again, patients who had undergone surgery experienced a 45% reduction in hazard of death compared with those who did not (HR 0.55, 95% CI 0.36–0.82, $p = 0.004$), while radiation therapy was associated with a significantly increased hazard (HR 2.19, 95% CI 1.40–3.43, $p < 0.001$). Patients receiving two or more other diabetes medication classes also demonstrated a markedly lower hazard (HR 0.23, 95% CI 0.08–0.68, $p = 0.007$) than those receiving no other diabetes medication classes.

In this model, each additional year of metformin use prior to pembrolizumab initiation was associated with an approximately 9% reduction in hazard (HR = 0.91 per year, 95% CI: 0.83–1.00, $p = 0.052$), though this trend is not quite statistically significant.

The proportional hazards assumptions held for both models, as Schoenfeld residual tests showed no evidence of proportional hazards violations for individual covariates (all $p > 0.05$) or globally (global $p = 0.069$).

4.3 Effect Modification by Glycemic Control

Evaluating the potential effect modification by glycemic control, we observed that of the 483 patients included in the analytic cohort, only 156 (32.3%) had a valid pre-index HbA1c measurement, while 327 (67.7%) were missing this information. Among those with available data, 108 patients had HbA1c < 7%, of whom 22 (20.4%) were metformin users and 86 (79.6%)

were non-users. In contrast, 48 patients had HbA1c ≥ 7%, with 14 (29.2%) exposed to metformin and 34 (70.8%) were not exposed (Figure 4). Although both strata contained users and non-users, the relatively small number of patients in the HbA1c ≥ 7% and metformin group indicated that the analysis of the interaction term would likely be underpowered and unstable.

In the Cox proportional hazards model, the interaction term between metformin uses and HbA1c group did not reach statistical significance. The Wald test yielded a p-value of 0.66, while the likelihood ratio test provided a consistent result ($p = 0.65$). These findings suggest no evidence that baseline glycemic control modifies the association between metformin exposure and overall survival.

Nevertheless, stratum-specific analyses indicated that among participants with HbA1c < 7%, metformin use was associated with a hazard ratio (HR) of 0.85 (95% CI, 0.35–2.10; $p = 0.73$), while among those with HbA1c ≥ 7%, the HR was 0.51 (95% CI, 0.06–4.33; $p = 0.53$). Although these estimates suggest a possible trend toward improved survival with metformin use in both subgroups, the wide confidence intervals reflect limited precision due to small sample sizes.

Given the limited number of patients within key strata and the lack of statistical support for a formal interaction, additional stratified Kaplan–Meier analyses were not pursued, as they would be underpowered and unlikely to yield reliable conclusions.

4.4 Effect Modification by Metastatic Disease Status

To explore whether the association between metformin use and overall survival (OS)

differed by disease stage, patients were stratified by metastatic status at pembrolizumab initiation.

Kaplan–Meier survival curves are displayed in **Figure 5**, comparing metformin users and non-users within the metastatic and non-metastatic subgroups. Among metastatic patients (n = 253; 76 deaths), survival curves showed modest separation between groups, with higher estimated OS among metformin users (log-rank p = 0.011; **Figure 5**). 2 years after pembrolizumab initiation, the estimated survival probability was around 78.9% for metformin users versus 59.5% for non-users.

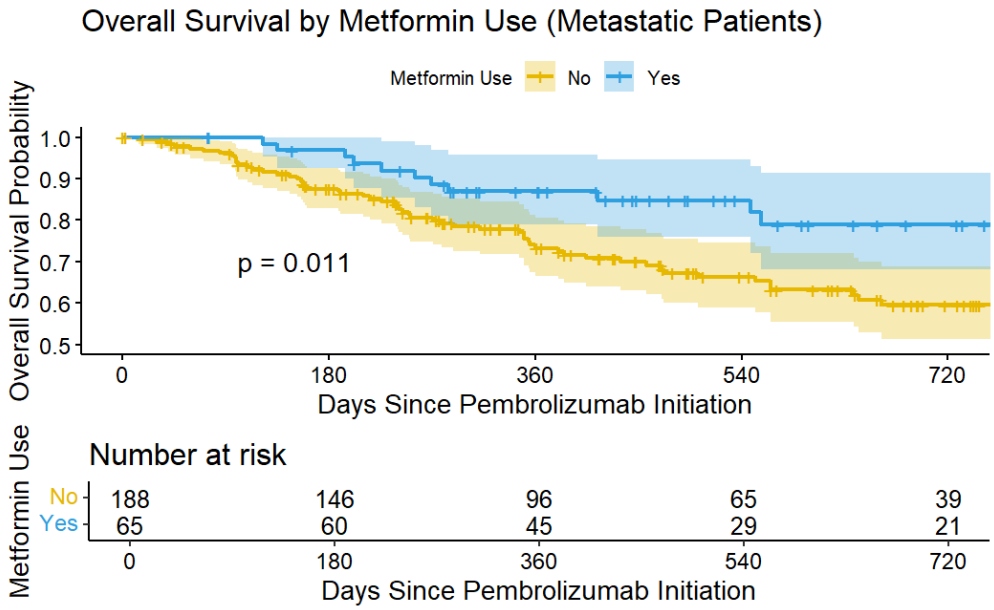

Figure 5: Overall Survival by Metformin Use Among Metastatic Patients

However, in the multivariable Cox proportional hazards model adjusting for age, year of pembrolizumab initiation, time between metformin and pembrolizumab initiation, surgery, radiation, insulin use, and comorbidities, metformin use was not significantly associated with OS (HR = 0.98; 95% CI 0.40–2.43; p = 0.97). In contrast, receipt of radiation therapy was significantly associated with worse survival (HR = 1.89; 95% CI 1.19–2.98; p = 0.006).

Metastatic disease itself remained a strong predictor of increased mortality in the full model (HR = 5.18; 95% CI 2.36–11.35; p < 0.001).

Among non-metastatic patients (n = 230; 25 deaths), Kaplan–Meier curves overlapped closely between exposure groups, with no discernible difference in OS (log-rank p = 0.86) (**Figure 6**). The adjusted hazard ratio for metformin use was 2.54 (95% CI 0.29–21.84; p = 0.40). The wide confidence interval reflects limited statistical power due to the small number of events in this subgroup.

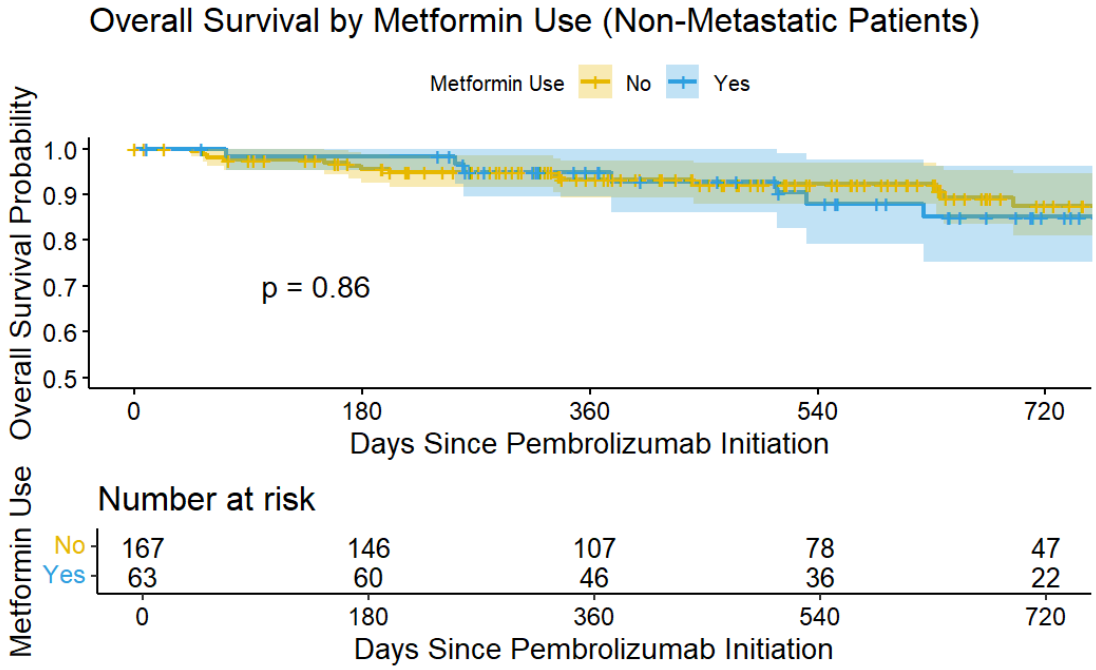

Figure 6: Overall Survival by Metformin Use Among Non-Metastatic Patients

To evaluate whether the survival association of metformin varies by disease stage, we examined the interaction between metformin (both duration and use) and metastatic status in multivariable Cox models. In the full models including interaction terms, the metformin duration × metastatic interaction was not statistically significant (p = 0.25), while the metformin use × metastatic interaction was borderline significant (HR = 0.35, 95% CI: 0.12–

1.00, p = 0.05), suggesting a possible disease stage–specific effect.

Stratified analyses further supported this pattern. Among metastatic patients, both **longer metformin use (HR = 0.89, p = 0.049)** and metformin exposure (HR = 0.53, p = 0.051) were associated with improved overall survival, whereas no associations were observed among non-metastatic patients (HRs ≈ 1.0, all p > 0.2).

Table 4. Multivariable Cox Proportional Hazards Model for Metastatic Patients

Variable	Hazard Ratio	95% CI	p-value
Metformin Duration (Years)	0.89	0.79 - 1.00	0.049
Age at Pembro Initiation	1.01	0.98 - 1.03	0.6
Year of Pembro Initiation	0.96	0.78 - 1.18	0.7
Surgery (Yes vs. No)	0.68	0.41 - 1.13	0.14
Radiation (Yes vs. No)	1.94	1.19 - 3.15	0.007
Insulin use at pembrolizumab	2.58	0.91 - 7.33	0.075
Number of comorbidities	1.12	0.9 - 1.39	0.3
Other diabetes medication classes (1)	0.67	0.26 - 1.73	0.4
Other diabetes medication classes (2+)	0.21	0.06 - 0.76	0.017

Note: HR (hazard ratios); 95% CI (95% confidence intervals)

These findings suggest that the potential survival benefit of metformin may be confined to patients with metastatic breast cancer, while its effect appears minimal in the non-metastatic setting. Clinical factors such as metastasis, surgery, and radiation remained independent predictors of survival across models.

5. Discussion

Principal findings: In this retrospective cohort of women with type 2 diabetes (T2D) and breast cancer receiving pembrolizumab, metformin use was associated with improved overall survival (OS) in unadjusted Kaplan–Meier analysis, but this association was not significant after multivariable adjustment. Long-term metformin use (>6.6 years) showed the strongest association with survival benefit, while analyses of glycemic control were inconclusive due to substantial missing HbA1c data. However, in stratified analyses by metastatic status, longer metformin duration was significantly associated with improved OS among metastatic patients (HR = 0.89, 95% CI 0.79–1.00, p = 0.049), whereas no survival benefit was observed in non-metastatic patients. These findings suggest that metformin's potential survival effect may be more pronounced in advanced disease contexts.

Comparison with prior studies: The present findings align with prior clinical and real-world studies reporting heterogeneous and context-dependent effects of metformin among cancer patients treated with immune checkpoint inhibitors (ICIs). While the overall cohort showed no independent association between metformin use and survival, the subgroup analysis suggested a potential benefit in metastatic breast cancer patients, consistent with evidence that metformin's antitumor activity may depend on disease stage and metabolic context. Wang et al. (2024, Frontiers in Pharmacology) found that metformin exposure was not significantly associated with improved survival across solid tumors treated with ICIs [19], and Afzal et al. (2018, Journal for ImmunoTherapy of Cancer) similarly reported no clear efficacy

enhancement from combining metformin with ICIs [20]. A 2023 meta-analysis (ScienceDirect) further emphasized substantial heterogeneity across tumor types and diabetic subgroups [21]. Together, these findings indicate that while metformin's overall clinical benefit in ICI-treated populations remains limited, its potential survival advantage may be more pronounced in patients with advanced or metastatic disease.

Clinical implications: While exploratory Kaplan–Meier analyses suggested a possible survival advantage with longer metformin exposure, this pattern was not confirmed in multivariable Cox regression after adjusting for key clinical covariates. Nevertheless, the significant association between metformin duration and improved survival among metastatic patients highlights a potentially stage-specific therapeutic interaction that deserves further study. Therefore, no strong evidence was found to support an independent association between metformin use and improved overall survival in the entire cohort, but the metastatic subgroup findings suggest that disease burden and treatment context may modify metformin's clinical relevance. Future studies should assess whether metformin's metabolic and immunologic effects could enhance checkpoint inhibitor efficacy specifically in advanced disease.

Strengths and limitations: Strengths of this study include the use of a large multi-institutional real-world dataset and the adjustment for demographic, treatment, and comorbidity variables. But this study has several limitations. As a retrospective real-world analysis using TriNetX data, it is subject to residual confounding and possible misclassification despite covariate adjustment. Key tumor-level factors such as PD-L1 expression, molecular subtype, and tumor mutational

burden were unavailable, potentially influencing both treatment assignment and outcomes. Metformin exposure was inferred from prescription records without dose or adherence data, and HbA1c values were missing for many patients, limiting the assessment of glycemic control effects. In addition, overall survival may not capture earlier treatment effects, and inter-institutional variability in coding and follow-up may have introduced measurement bias. These limitations underscore the exploratory nature of this analysis and the need for prospective, biomarker-integrated studies to better define the role of metformin in modulating immunotherapy outcomes in breast cancer.

6. Conclusion

In this retrospective cohort study of 483 women in the U.S. with type 2 diabetes and breast cancer treated with pembrolizumab, several analyses were run to explore the relationship between metformin use and overall survival. In preliminary Kaplan-Meier plots, it appeared as though metformin use, and especially long-term metformin use, was associated with improved overall survival. However, after adjusting for potential confounders such as age, year of treatment, surgery, radiation, number of comorbidities, and use of other diabetes medications in a cox proportional hazards model, metformin was no longer statistically significant in its association to overall survival. This suggests that this difference in survival outcomes between the group on metformin and the control group may be partially explained by confounding variables.

In stratified analyses, however, longer metformin exposure was significantly associated with improved survival among patients with metastatic breast cancer (HR = 0.89, 95% CI 0.79–1.00, p = 0.049), whereas no survival benefit was observed among non-metastatic patients.

This stage-specific trend suggests that the potential protective effect of metformin may be more relevant in advanced disease contexts.

Subgroup analyses showed no clear evidence that baseline glycemic control modified the association between metformin use and survival outcomes, though the significant amount of missing data limits the generalizability of this result. While metformin survival curves appeared to very much diverge in the Kaplan-Meier plot for patients with metastatic cancer, further analysis showed that metastatic status did not significantly modify the association between metformin use and survival outcomes. However, these subgroup analyses results should be interpreted with caution due to small subgroup sample sizes and missing data, which limits the precision and interpretability of estimates.

Despite these limitations, the observed association between longer metformin exposure and improved outcomes in metastatic patients demonstrate a potential link between sustained metformin exposure and enhanced immunotherapy response, aligning with preclinical evidence of metformin's immunomodulatory properties. Future prospective studies with larger, stage-balanced samples are warranted to confirm whether metformin could serve as an effective adjunct to immune checkpoint inhibitors, particularly in metastatic breast cancer patients with diabetes.

7. References

[1] Corleto KA, et al. Metformin and breast cancer: Current findings and future directions. *Pharmaceutics.* 2024;17(3):396. doi:10.3390/pharmaceutics17030396

[2] Yang J, Li H, Chen Y, et al. Prognostic value of metformin in cancers: An updated meta-analysis based on 80 cohort studies. *Front Oncol.* 2022;12:989215. doi:10.3389/fonc.2022.989215

[3] Amato L, et al. Addition of metformin to anti-PD-1/PD-L1 drugs activates anti-tumor immunity via the cGAS-STING pathway. *Cell Death Dis.* 2025; (online ahead of print). doi:10.1038/s41419-025-07636-7

[4] Jin M, et al. PD-1/PD-L1 immune checkpoint blockade in breast cancer. *Mol Cancer.* 2024;23:21. doi:10.1186/s12943-024-02176-8

[5] Qi J, He P, Yao H, et al. Cancer risk among patients with type 2 diabetes: A real-world study in Shanghai, China. *J Diabetes.* 2019;11(11):878-883. doi:10.1111/1753-0407.12926

[6] Boyle P, Boniol M, Koechlin A, et al. Diabetes and breast cancer risk: A meta-analysis. *Br J Cancer.* 2012;107(9):1608-1617. doi:10.1038/bjc.2012.414

[7] Chen Y, Wu F, Saito E, et al. Association between type 2 diabetes and risk of cancer mortality: A pooled analysis of over 771,000 individuals in the Asia Cohort Consortium. *Diabetologia.* 2017;60(6):1022-1032. doi:10.1007/s00125-017-4229-z

[8] TriNetX. Datasets & Analytics [Internet]. Cambridge, MA: TriNetX; 2025 [cited 2025 Sep 23]. Available from: https://trinetx.com/data-sets-analytics/

[9] Castelli G, Barlett SJ. Should metformin continue as first-line pharmacotherapy for patients with type 2 diabetes? Yes: Metformin is still the best choice. *Am Fam Physician.* 2024;109(3):200-201.

[10] Centers for Disease Control and Prevention. Breast cancer statistics [Internet]. Atlanta, GA: CDC; 2024 [cited 2025 Oct 13]. Available from: https://www.cdc.gov/cancer/breast/statistics/

[11] National Cancer Institute. SEER cancer stat facts: Female breast cancer [Internet]. Bethesda, MD: National Cancer Institute; 2024 [cited 2025 Oct 13]. Available from: https://seer.cancer.gov/statfacts/html/breast.html

[12] American Cancer Society. Breast cancer facts & figures 2024–2025 [Internet]. Atlanta, GA: American Cancer Society; 2024 [cited 2025 Oct 13]. Available from: https://www.cancer.org/research/cancer-facts-statistics.html

[13] American Cancer Society. Immunotherapy for breast cancer [Internet]. Atlanta, GA: American Cancer Society; 2024 [cited 2025 Oct 13]. Available from: https://www.cancer.org/cancer/breast-cancer/treatment/immunotherapy.html

[14] Merck & Co., Inc. Dosing schedule for KEYTRUDA® (pembrolizumab) [Internet]. Rahway, NJ: Merck; 2024 [cited 2025 Oct 13]. Available from: https://www.keytrudahcp.com/dosing/

[15] Gregg EW, Cheng YJ, Srinivasan M, et al. Trends in cancer mortality among people with vs without diabetes in the United States, 1988–2015. *Ann Intern Med.* 2018;167(12):913-920. doi:10.7326/M17-2697

[16] Tsilidis KK, et al. UK study shows increased cancer mortality in people with type 2 diabetes. *Diabetologia.* 2022;65(4):712-719. doi:10.1007/s00125-022-05694-9

[17] Kowall B, Rathmann W. Relationship between diabetes and cancer mortality: A pooled

analysis of prospective cohort studies. *PLoS One.* 2017;12(1):e0170084. doi:10.1371/journal.pone.0170084
[18] Bhattarai M, et al. Association between diabetes and cancer mortality in the United States. *Am J Prev Med.* 2022;62(3):412-420. doi:10.1016/j.amepre.2021.09.013
[19] Wang J, Lin J, Guo H, et al. Effect of metformin on outcomes of patients treated with immune checkpoint inhibitors. *Front Pharmacol.* 2024;15:10991879. doi:10.3389/fphar.2024.10991879
[20] Afzal MZ, Mercado RR, Shirai K. Efficacy of metformin in combination with immune checkpoint inhibitors. *J Immunother Cancer.* 2018;6(1):109. doi:10.1186/s40425-018-0428-3
[21] Li X, Zhang Y, Zhao R, et al. Clinical evidence for the prognostic impact of metformin in cancer therapy: A systematic review and meta-analysis. Cancer Treat Rev. 2024;128:103998. doi:10.1016/j.ctrv.2024.103998

8. Appendix

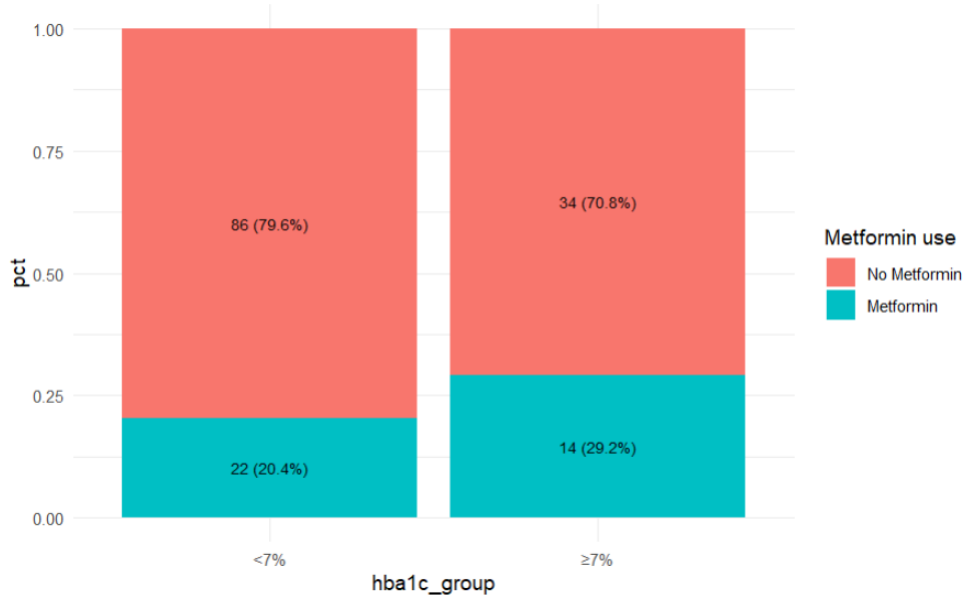

Figure 7: Distribution of metformin use by baseline HbA1c category

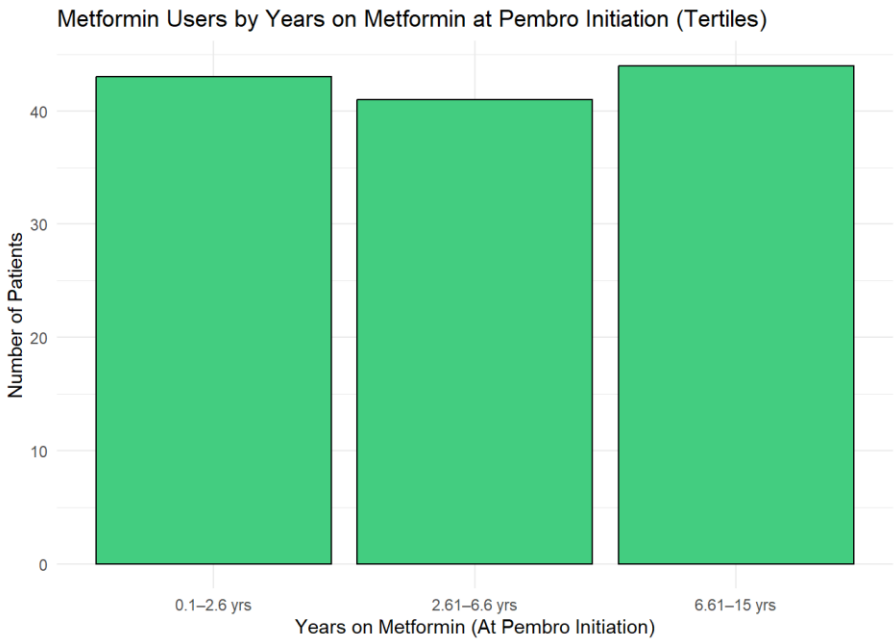

Figure 8: Classification of Metformin Duration Variable into Tertiles.

Table 5. Variable Definitions of the Patient Cohort Flagged Dataset. This table provides detailed definitions for all variables generated and incorporated as columns within the Patient Cohort Flagged Dataset. Each row corresponds to an individual patient in the cohort. For each variable, the table specifies the derivation rules employed and the temporal windows applied during variable construction.

Variable Name	Rules Used	Time Period
patient_id	Patient ID from the patient table.	1/1/2019 – 4/25/2025
BC_diagnosis_date	1. Include codes with "breast" **and** any of: "malignant", "carcinoma", or "Anaplastic large cell lymphoma". 2. Exclude terms: "secondary", "history", "screening", "susceptibility", or "lobular". 3. Group by patient_id; select **earliest** qualifying diagnosis date.	1/1/2019 – 4/25/2025
pembro_initiation	1. Records with "pembrolizumab" in description. 2. Earliest record **after** BC diagnosis date. 3. Exclude patients with no pembrolizumab post-diagnosis.	After BC_diagnosis_date
pembro_final	1. Records with "pembrolizumab" in description. 2. Select **last** pembrolizumab record.	After pembro_initiation
Metformin_start_date	1. Earliest metformin record per patient. 2. Must occur **before** pembro initiation. 3. Include only if **Metformin = 1**; else	Metformin start < pembro initiation < metformin end

	NA.	
Metformin_end_date	1. Latest metformin record per patient. 2. Must occur **after** pembro initiation. 3. Include only if **Metformin = 1**; else **NA**.	Metformin start < pembro initiation < metformin end
metformin	1. = 1 if patient has metformin record. 2. Must have pembro initiation between metformin start and end dates. 3. Else = 0.	Metformin start < pembro initiation < metformin end
Death_date	1. Use last day of month from month_year_death in demographics (e.g., "202404" → "2024-04-30"). 2. Must be **after** metformin end and pembrolizumab final.	Death_date > metformin_end_date, pembro_final
Year_of_birth	Remove patient from cohort if missing.	—
age_at_pembro_initiation	Year of pembro initiation − Year of birth.	—
age_at_diagnosis	Year of BC diagnosis − Year of birth.	—
age_at_met_initiation	Year of metformin start − Year of birth.	—
Year_of_pembro_initiation	Year of pembrolizumab initiation.	—
years_between_met_and_pembro	Duration (in years) of metformin use at pembrolizumab initiation. Set to 0 if **Metformin = 0**.	—
metformin_years_at_pembro_tertile	1. = 0 when Metformin = 0. 2. Otherwise, discretize years_between_met_and_pembro into tertiles: • 0.1–2.6 years • 2.61–6.6 years • 6.61–15 years	—
surgery	Set surgery = "yes" if any qualifying record exists; otherwise "no".	surgery_date ≤ pembro_initiation + 6 months
radiation	Set radiation = "yes" if qualifying record exists; else "no".	radiation_date < pembro_initiation
chemo	From chemo_lines table. Set chemo = "yes" if any record exists; else "no".	chemo_start < pembro_initiation
tamoxifen_ai_status	If: • only tamoxifen → "tamoxifen_only" • only AI → "ai_only" • both → "both" • neither → "no"	med_start < pembro_initiation
BMI	Latest BMI record **before** pembrolizumab initiation.	date < pembro_initiation
HbA1c	Latest HbA1c record **before** pembrolizumab initiation.	date < pembro_initiation
insulin_at_pembro	1. Identify insulin or GLP-1 receptor	insulin start < pembro

	agonist use from standardized terminology list: glargine, degludec. 2. From medication_ingredient records, find patients whose pembrolizumab initiation date falls between earliest and latest insulin use dates. 3. Flag as 1 if insulin was used at pembro initiation, else 0.	initiation < insulin end
num_other_med_classes	1. Identify all non-insulin and non-metformin diabetes drug classes (e.g., sulfonylureas, SGLT-2 inhibitors, DPP-4 inhibitors, etc.) from standardized terminology list. 2. Count unique drug classes with active use at pembrolizumab initiation. 3. Categorize counts as 0, 1, or 2+.	medication start < pembro initiation < medication end
num_other_meds	1. Identify all non-metformin diabetes medications (including insulin and other oral/GLP-1 agents). 2. Count number of individual medications with active use at pembrolizumab initiation. 3. Categorize counts as 0, 1, 2, or 3+.	medication start < pembro initiation < medication end
days_on_pembro	1. Calculate time difference between pembro_final and pembro_initiation in days using difftime.	From pembro_initiation to pembro_final
obesity	1. Identify diagnosis codes containing "obesity" in standardized terminology. 2. Include diagnoses recorded within 1 year before pembro initiation. 3. Flag 1 if present, else 0.	1 year prior to pembro initiation
hypertension	1. Identify diagnosis codes containing "hypertension". 2. Include diagnoses recorded within 1 year before pembro initiation. 3. Flag 1 if present, else 0.	1 year prior to pembro initiation
hyperlipidemia	1. Identify diagnosis codes containing "hyperlipidemia". 2. Include diagnoses recorded within 1 year before pembro initiation. 3. Flag 1 if present, else 0.	1 year prior to pembro initiation
liver_disease	1. Identify diagnosis codes containing "liver disease".	1 year prior to pembro initiation

	2. Include diagnoses recorded within 1 year before pembro initiation. 3. Flag 1 if present, else 0.	
kidney_disease	1. Identify diagnosis codes containing "kidney disease" or "renal disease". 2. Include diagnoses recorded within 1 year before pembro initiation. 3. Flag 1 if present, else 0.	1 year prior to pembro initiation
num_comorbidities	1. Sum of binary indicators for obesity, hypertension, hyperlipidemia, liver disease, and kidney disease (rowSums).	1 year prior to pembro initiation
metastatic	1. Identify diagnosis codes containing "secondary malignant" or "secondary and unspecified malignant" neoplasms, excluding certain codes (198.2, C79.2, C79.81, 198.81). 2. Include diagnoses **on or after** BC_diagnosis_date and **on or before** pembrolizumab initiation date. 3. Flag 1 if present, else 0.	BC diagnosis < secondary diagnosis < pembrolizumab initiation
death_event	1. Flag 1 if death_date is not missing, otherwise 0.	Through censoring date
censor_date	1. For each patient, select the **latest start_date** in medication_ingredient. 2. Use this as censoring date if death date is not available.	1/1/2019 – 4/25/2025
os_days	1. Calculate time from pembro_initiation to either death_date or censor_date (whichever comes first). 2. Measured in days using difftime.	From pembro initiation to death or censoring

Table 6. Associated Codes used in Variable Definitions. This table shows the codes and code systems isolated from the Standardized Terminology Table when defining each variable.

Variable Name	Associated Codes (ICD / HCPCS / RxNorm / CPT / LOINC)
BC_diagnosis_date	*ICD-9/10*: Include codes with "breast" and "malignant", "carcinoma", or "Anaplastic large cell lymphoma"; exclude "secondary", "history", "screening", "susceptibility", "lobular". **ICD-10-CM:** C43.52, C44.501, C44.511, C44.521, C44.591, C4A.52, C50, C50-C50, C50.011, C50.012, C50.019, C50.021, C50.022, C50.029, C50.1, C50.11, C50.111, C50.112, C50.119,

	C50.12, C50.121, C50.122, C50.129, C50.2, C50.21, C50.211, C50.212, C50.219, C50.22, C50.221, C50.222, C50.229, C50.3, C50.31, C50.311, C50.312, C50.319, C50.32, C50.321, C50.322, C50.329, C50.4, C50.41, C50.411, C50.412, C50.419, C50.42, C50.421, C50.422, C50.429, C50.5, C50.51, C50.511, C50.512, C50.519, C50.52, C50.521, C50.522, C50.529, C50.6, C50.61, C50.611, C50.612, C50.619, C50.62, C50.621, C50.622, C50.629, C50.8, C50.81, C50.811, C50.812, C50.819, C50.82, C50.821, C50.822, C50.829, C50.9, C50.91, C50.911, C50.912, C50.919, C50.92, C50.921, C50.922, C50.929, C84.7A, D05, D05.1, D05.10, D05.11, D05.12, D05.8, D05.80, D05.81, D05.82, D05.9, D05.90, D05.91, D05.92. **ICD-9-CM:** 174, 174.0, 174.1, 174.2, 174.3, 174.4, 174.5, 174.6, 174.8, 174.9, 175, 175.0, 175.9, 233, 233.0
pembro_initiation, pembro_final	**HCPCS:** J9271 **RxNorm:** 1547545 *(pembrolizumab)*
Metformin_start_date, Metformin_end_date, metformin	**RxNorm:** 6809 *(metformin)*
surgery	**CPT:** 19499, 19301–19307, 19328, 19340–19342, 19357, 19361–19369, 38500–38525, 38740–38745 **ICD-10-PCS:** Codes starting with ^0H[BTRUXHWP] or ^07[BTC] **ICD-9-CM:** 85.20–85.23, 85.33–85.36, 85.41–85.48, 85.50–85.54, 85.60–85.67, 40.23–40.24, 40.50–40.54, 40.19
radiation	**CPT:** 77261–77295, 77300, 77331, 77336, 77338, 77370, 77306–77307, 77371–77373, 77385–77386, 77401–77412, 77424, 77425, 77427, 77431, 77432, 77435, 77470, 77750–77799 **HCPCS:** G6001, G6002, G6017, G0339, G0340, G6003–G6016 **ICD-10-PCS:** Codes starting with ^DM **ICD-9-CM:** 92.21–92.29, 92.30–92.33
tamoxifen_ai_status	**Tamoxifen:** 10324 **Aromatase inhibitors (AI):** 84857 (*Anastrozole*), 72965 (*Letrozole*), 258494 (*Exemestane*)
BMI	**LOINC:** 39156-5
HbA1c	**LOINC:** 4548-4
chemo	Uses **chemo_lines table** (no specific code).
Insulin_at_pembro	**RxNorm**: 1670007 (*Degludec*), 274783 (*Glargine*), 139825 (*Detemir*)
num_other_meds, num_other_med_classes	*Sulfonylureas:* **LOINC:** 73360-0, 73361-8, LP15165-1, LP164791-8, LP391453-0; **RxNorm:** 1670007, 33738 *(Glimepiride)*. **LOINC:** 10539-5, 27025-6, 47204-3, 48319-8, 48326-3, 6944-3, LP14719-6, LP164792-6, LP390543-9; **RxNorm:** 16681 *(Glipizide)*.

	Meglitinides: **LOINC:** 49487-2, 49702-4, 73690-0, LP164294-3, LP390550-4; **RxNorm:** 25789 *(Nateglinide)*. **LOINC:** 38542-7, 48328-9, 73550-6, 73551-4, 73689-2, LP164293-5, LP390542-1; **RxNorm:** 274332 *(Repaglinide)*. *TZDs:* **LOINC:** 100351-6, 73204-0, 73205-7, LP164295-0, LP165192-8, LP171629-1, LP390551-2; **RxNorm:** 1991302 *(Pioglitazone)*. *Alpha-glucosidase Inhibitors:* **LOINC:** LP64443-2; **RxNorm:** 139953, 4821 *(Acarbose)*. **RxNorm:** 141626 *(Miglitol)*. *DPP-4 Inhibitors:* **LOINC:** LP403053-4 *(Alogliptin)*. **LOINC:** LP403052-6; **RxNorm:** 1100699 *(Linagliptin)*. **LOINC:** LP402865-2, LP403055-9; **RxNorm:** 1373458 *(Saxagliptin)*. **LOINC:** LP402864-5, LP403054-2; **RxNorm:** 1368001 *(Sitagliptin)*. Bile Acid Sequestrants: **RxNorm:** 475968 *(Colesevelam)*. Dopamine-2 Agonists: **LOINC:** 11232-6; **RxNorm:** 593411 *(Bromocriptine)*. SGLT-2 Inhibitors: **LOINC:** LP403210-0; **RxNorm:** 1440051 *(Canagliflozin)*. **LOINC:** LP403211-8; **RxNorm:** 1488564 *(Dapagliflozin)*. **LOINC:** LP432640-3; **RxNorm:** 1545653 *(Empagliflozin)*. **RxNorm:** 1551291 *(Etrugliflozin)*. Amylin: **RxNorm:** 60548 *(Pramlintide)*. GLP-1 Receptor Agonists: **RxNorm:** 73044 *(Dulaglutide)*. **RxNorm:** 857974 (E*xenatide*). **SNOMED:** 105200005 *(Liraglutide)*. **SNOMED:** 121465001 *(Lixisenatide)*. **SNOMED:** 5178781 *(Semaglutide)*.
obesity	**ICD-9-CM:** 278.00, 278.01**. ICD-10-CM:** E66.01, E66.09, E66.1, E66.2, E66.8, E66.813, E66.9, O99.211, O99.212, O99.213
hypertension	**ICD-9-CM:** 401.0, 401.1, 401.9. **ICD-10-CM:** I10, I15.0, I15.1, I15.2, I15.8, I15.9, I27.0, I27.20, I27.21, I87.312, K76.6, O10.913, O10.919
hyperlipidemia	**ICD-9-CM:** 272.4, 272.2. **ICD-10-CM:** E78.49, E78.5, E78.2
liver_disease	**ICD-10-CM:** K75.6, K75.9
kidney_disease	**ICD-9-CM:** 585.3, V18.69. **ICD-10-CM:** D63.1, E11.22, I12.0, I12.9, I13.0, I13.10, I13.11, I13.2, N18.1, N18.2, N18.3, N18.30, N18.31, N18.32, N18.4, N18.5, N18.6, N18.9
metatstatic	**ICD-9-CM:** 196.3, 196.9. **ICD-10-CM:** C77.0, C77.1, C77.2, C77.3, C77.4, C77.5, C77.8, C77.9, C78.00, C78.01, C78.02, C78.1, C78.2, C78.4, C78.5, C78.6, C78.7, C78.80, C78.89, C79.00, C79.01, C79.02, C79.31, C79.49, C79.51, C79.52, C79.71, C79.72, C79.82, C79.89, C79.9.