Research Article
Open Access

Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia

Abstract

Aims: To investigate all-cause mortality (ACM) attributable to insulin-treated diabetes mellitus through propensity score (PS)-weighting with and without novel confounders identified by Random Survival Forest (a machine learning approach).

Methods: Prospective clinic encounter data was obtained from 1517 females with Type 2 diabetes (mean age 63±12 years) from Barranquilla, Colombia (2003 – 2016, censored August 2017) for a median 10-year mortality follow-up. Risk variables of importance for ACM were identified on RSF screening. Survival was compared in retrospective cohorts, identified by baseline treatment with glucose-lowering therapy, and balanced for confounders through PS-weighting with and without RSF variables using multivariable Cox regression.

Results: RSF screening identified new risk variables (e.g., recruitment year, parity, reproductive lifespan) for ACM in women receiving insulin. The unweighted risk estimate showed a nonsignificant increased risk for ACM [HR 1.32 (.9, 2), p=0.2] compared to noninsulin treated women. After balancing for risk covariates in the compared cohorts, PS showed no significant effect of insulin on all-cause mortality [HR 95% CI 0.83 (0.5, 1.4) p=0.5] whereas PS-weighted analyses incorporating RSF novel variables approached conservative ACM estimates [HR 95% CI 0.56 (0.3, 1.0) p=0.07)]. The estimated ACM risk from active smoking was also more conservative with RSF weighting.

Conclusion: In this observational study, insulin treatment appeared to be a surrogate for higher-risk women with diabetes mellitus. RSF-augmented PS analysis showed that insulin treatment may potentially be associated with a survival advantage compared to non-insulin treatment in older female diabetics.

Keywords: machine learning, random survival forest, propensity score weighting, all-cause mortality, females, diabetes, insulin, confounder

Article highlights

Observational studies suggest increased mortality from insulin treatment in diabetic patients. Propensity score methods enable inference from observational studies but cannot account for unknown confounders that were not included in developing the propensity model. To determine if insulin treatment increases mortality in adult females with diabetes mellitus in this observational study, we applied propensity score weighting, balancing known risk factors for mortality, enriched with new confounders identified by machine learning (RSF). Compared to unweight estimates of all-cause mortality risk, this approach resulted in a 40% lower mortality risk (p=0.07) estimate observed with insulin treatment. Thus, machine learning approaches (such as RSF) can identify novel predictors that can minimize residual confounding bias when incorporated in PS analyses.

Volume 10 Issue 2 - 2023

Carlos Cure Cure,[1] Eileen E Navarro Almario,[9] Yuan Gu,[2,4] John D Eustaquio,[5] Pablo Cure,[3,8] Anwar Husain,[7] Colin O Wu,[4,6] Xin Tian,[4,6] Ramiro Galindo,[1] Victor Crentsil,[7] George Sopko,[3] Gyorgy Csako,[3] Ahmed A Hasan,[3,4]

[1]Biomelab SAS, Barranquilla, Colombia
[2]Department of Statistics, the George Washington University, Washington, DC, USA
[3]Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA
[4]Data Science Initiative, Cardiovascular Research Laboratories, University of Maryland School of Medicine, Baltimore, MD, USA
[5]University of Maryland in Baltimore County, MD, USA
[6]Division of Intramural Research, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD.
[7]Food Drug Administration, MD, USA
[8]National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA
[9]Paraclete Professionals, LLC, Clarksville, MD, USA

Correspondence: Eileen Navarro Almario, Paraclete Professionals, LLC, Clarksville, MD, USA, Tel 4105912826, Email eileen.almari@gmail.com

Received: November 29, 2023 | **Published:** December 15, 2023

Introduction

Studies with newer glucose-lowering drugs (GLDs) suggesting mortality and morbidity advantage over insulin have prompted a reexamination of treatment approaches in patients with diabetes mellitus in the real-world setting.[1] Increased mortality attributed to insulin comes from observational studies which may be limited by treatment assignment bias. A meta-analysis of randomized trials found no such increased risk, albeit 70% of the effect was driven by a single trial[2] where women comprised 35% of the population.[25] We sought to identify risk variables for ACM in this underrepresented population[3] in a community care setting of Barranquilla, Colombia and estimate the risk attributable to insulin treatment when these variables are taken into consideration. Propensity score (PS)-based analytic tools adjust for confounding in observational studies and have enabled estimation of drug effect using data collected from clinical settings. To optimize comparability of the non-randomized treatment groups, a parsimonious model that prioritizes covariates identified by experts to affect outcome (e.g., ACM in diabetes) was used to estimate the effect attributed to insulin in a PS regression models.[4] Reliance on previously established risk variables can result in residual confounding from unmeasured factors when data is derived from populations not previously studied. The availability of high dimensional data in real world clinical settings can overcome some of this limitation but

balancing the compared populations using regression analysis may prove computationally challenging.

Machine learning methods such as Random survival forest (RSF) that can process highly dimensional data may be useful to employ in PS models. More recent PS models (high dimensional propensity score models or HDPS)[5] have integrated machine learning for variable discovery. We chose to perform our analyses sequentially, first applying RSF to identify variables important for ACM, then using PS weighting to compare the effect attributable to insulin with and without the RSF identified variables not previously identified by domain experts. Using RSF, we previously reported diabetes mellitus as a top predictor of all-cause mortality (ACM) in male and female participants in our study (**PR**evalence **o**f metab**O**lic disease and its influence on macrovasc**U**lar **D**isease and Fractures; PROUD) in Barranquilla, Colombia.[6] Moreover, insulin treatment was identified as a risk variable for ACM in subjects with diabetes mellitus, (n=1517, ~10% of the PROUD cohort). Age at menarche and number of pregnancies were additional ACM risk factors identified in females, who comprised 82.6% (9869 of 11952) of PROUD subjects. These findings motivated the investigation of the contribution of RSF to PS-weighted analysis of ACM risk in insulin and non-insulin treated female diabetics. The goal of this sub-study was to assess whether RSF can minimize residual confounding in pharmacoepidemiologic studies by identifying confounders that may have otherwise been overlooked.

Methods

Study population and design

Barranquilla is a large industrial city on the Caribbean coast of Colombia, with a population (1.2 million in 2003) consisting mainly of mixed-race inhabitants. Universal health care is provided through community health centers complemented by tertiary care through a network of university-based and independent private hospitals. Between 2003 and 2016, 12000 adult participants (82% women) recruited from a community health program[7–9] consented to provide demographic, anthropometric, medical history, reproductive, diet, and activity data in this study (PROUD) in Barranquilla, Colombia. Diabetic females with type 2 diabetes ≥18 years of age who received health care in the community center and agreed to participate were included in the current sub-study (Supplemental Figure 1). Subjects <18 years of age, with missing anthropomorphic and behavioral data were excluded from the study. The BIOMELAB Institutional Review Board, Barranquilla approved the study.

Source of data and definitions

Demographic characteristics, socioeconomic parameters, anthropometric measures, gestational (reproductive and lactation) history, baseline medical and family history, diet, physical activity, and recreational habits were collected onto a standard data collection instrument as part of routine medical care. Data preprocessing reduced over 300 variables into clinically relevant concepts; definitions based on current medical standards and their linkage to the tabulated data is shown in the final list of covariates (Table 2). Derived variables, including socioeconomic status obtained by geolocation[10] are summarized in Supplemental Tables 1 and 2.

Baseline diagnosis of diabetes, hyperlipidemia, metabolic syndrome, and hypertension was based on self-reported diagnosis, confirmed by review of diagnostic test results at recruitment or treatment history from outpatient medical records. Waist circumference and obesity definitions thresholds were based on established Colombian reference standards.[11] Age at menarche and menopause, menstrual cycle history, self-reported "infertility", gravidity, parity, and lactation history were collected; reproductive lifespan at baseline (age in years at menopause minus age at menarche or interval in years between the first and last child for which maternal age is available) was derived from the collected data. Osteoporosis was based on self-reported diagnosis of osteoporosis or osteopenia and/or antiresorptive treatment. Cardiovascular disease was identified by self-report and treatment history. Healthy behavior, defined as a minimum of 3 of the 4 following factors:[12] absence of smoking, no obesity, weekly physical exercise, and self-reported healthy diet was assessed from baseline food and activity questionnaires. Baseline medication history included the name, dose, route, and year of GLD initiation. Research participants were retrospectively categorized into those receiving insulin treatment (alone or in combination with other GLD) and compared to those that did not receive treatment with insulin in addition to diet, exercise, and/or nutritional supplements. Baseline characteristics were compared to illustrate between-group differences (Table 1). For all analyses, a 2-tailed $p<0.05$ denoted statistical significance.

ACM for participants in PROUD (median follow-up of 10 years) was confirmed by linkage with the national mortality surveillance system. Mortality data consisted of the date and location of death, the physician recorded and final cause of death as reported to the national statistics and social benefit systems.[13] All data was translated from Spanish; mortality data was further recategorized into ICD codes. Survival was prospectively followed from 2003 and administratively censored on August 31, 2017. Overall survival was defined as the time (in days) from study baseline to death or censoring.

Random survival forest (RSF) analysis

RSF analysis[14] was performed on the data of female participants of the PROUD,[6] and in the sub study of females with diabetes mellitus. Variables with ≥10% missing data were excluded from all analyses (Supplemental Figure 2), leaving 85 of 88 variables for consideration in PS development. Of these, the top 50 variables, based on the estimated VIMP score[15] were built into a model on the original data without missing imputation. Reproductive variables (including reproductive lifespan, a known risk for cardiovascular mortality)[16–19] were included in RSF model development (Table 2). The final RSF model's robustness was assessed through tuning different model hyperparameters, encompassing key variables such as the number of trees, ranging from 500, 800, to 1000, determined based on the out-of-bag error rate. To enhance robustness and address missing data effectively, we conducted a sensitivity analysis employing multiple imputation by chained equations (MICE). We finalized the model with specific hyperparameters for the RSF, consisting of 1000 trees, utilizing the log rank test as the splitting rule, and opting to omit missing data in the analysis. Model development and hyperparameter tuning (to discriminate the robustness of the final RSF model) are shown in Table 2 and Supplemental Table 3, respectively.

Propensity score (PS)-weighting

The conditional probability (i.e., propensity) of receiving insulin (alone or in combination) *vs.* no insulin was estimated in a logistic regression model with insulin treatment as outcome variable conditioned on risk variables with or without the new variables identified by RSF analysis (year of recruitment, parity in pregnancy, reproductive lifespan). The final model prioritized variables identified by subject matter experts as ACM predictors in diabetes (Supplemental Table 4). We performed PS weighting rather than matching to account

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

for all subjects and covariates in this study subset of PROUD. The PS weights (calculated as an inverse of the PS) were then applied to the individuals in the insulin and non-insulin treatment cohorts[20] and the balance of confounding variables across the compared populations assessed (Supplemental Figure 3).

Survival analysis with and without PS-weighting

Survival analyses were performed using R Statistical Software (version 4.1.2, R Core Team 2021 and replicated in Stata (Stata Corp 2021 "Stata Statistical Software: Release 17". College Station, TX: StataCorp LLC). The Kaplan-Meier survival curves were separately estimated for subjects receiving insulin (combination and monotherapy) *vs*. the non-insulin treated cohort and the incidence of ACM compared using the log-rank test for the unweighted and PS-weighted (with or without new RSF variables) populations. Significance of risk predictors for ACM was confirmed by traditional Cox proportional hazard (HR) regression. As monotherapy with insulin could have represented a phenotype that is fully insulin dependent (such as Type 1 diabetes mellitus), we compared the insulin monotherapy group *vs.* all other patients as a sensitivity analysis.

Results

Baseline characteristics

Baseline characteristics of adult female diabetics receiving insulin *vs*. the non-insulin-treated cohort are summarized in Table 1. The mean age of study participants was 63 years (±11.8, range 19-91), with a majority considered as indigenous but multiracial and married or cohabiting (68%). Although the demographic, anthropometric, and gestational characteristics of the two groups were similar, insulin treated subjects differed in important diabetic, metabolic and co-morbid characteristics from their counterparts. Approximately one-third of the population belonged to the highest socioeconomic strata; subjects receiving insulin had more missing socioeconomic information and somewhat lower socioeconomic ranking.

Table 1 Baseline characteristics of 1517 female subjects with diabetes mellitus by treatment cohort

Baseline characteristics*	Insulin treated (N=176)	No insulin (N=1341)	P value+
Demographics			
Age(y) (n=1517)	63 ± 12	63 ± 12	0.87
Race (n (%)) (n=1517)			0.78
Black	2 (1)	14 (1)	
White	3 (2)	38 (3)	
Multiracial	171 (97)	1289 (96)	
Socioeconomic Parameters			
Marital status (n (%)) (n=1436)			0.29
Alone	61 (35)	395 (30)	
Single/Separated/Divorced	37 (21)	243 (18)	
Widow/widower	24 (14)	152 (11)	
With Partner	108 (61)	872 (65)	
Married	90 (51)	785 (59)	
Living together	18 (10)	87 (7)	
Socioeconomic strata (n (%)) (n=1217)			0.06
1(low)	13 (7)	130 (10)	
2	28 (16)	222 (17)	
3	30 (17)	225 (17)	
4	10 (6)	33 (3)	
5 (high)	49 (28)	477 (36)	
Anthropometrics			
SBP (mmHg) (n=1471)	140 ± 20	139 ±22	0.41
DBP (mmHg) (n=1471)	77 ± 11	79 ±11	0.02
Heart Rate (beats/min) (n=1494)	78 ± 11	78 ±10	0.52
Height (cm) (n=1494)	155 ± 7	154 ±7	0.24
Body Weight (kg) (n=1448)	68 ± 13	67 ± 13	0.17
Waist (cm) (n=1459)	92 ±11	90 ± 10	0.01
Hip (cm) (n=1459)	102 ± 11	102 ±.11	0.95
Waist/Hip Ratio(n=1459)	.90 ±.07	.88 ± .06	<0.01
BMI (kg/m²) (n=1494)	28 ±.4.99)	28 ±. 5	0.56
BMI Categories (n=1494)			0.03
Underweight	6 (3.5)	14 (1.1)	
Normal weight	35 (20.2)	332 (25.1)	
Pre-Obese	73 (42.2)	560 (42.4)	
Obese	59 (34.1)	415 (31.4)	
Gestational			
Age at menarche (y) (n=1478)	13 ± 2	13 ± 3	0.06
Age at menopause onset age (y) (n=350)	47 ± 6	48± 6	0.62
Reproductive lifespan† (years) (n=1479)	47 ± 14	47 ± 13	0.94
Number of pregnancies (mean ± SD) (n=1404)	5 ± 3	6 ± 3	0.16
Number of deliveries (n=1383)	4 ± 2	5 ± 3	0.09
Parity in pregnancy (n, %) (n=1517)			0.76
0	18 (10)	105 (8)	
1	15 (8)	91 (7)	
2	20 (11)	170 (13)	

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

Table 1 Continued...

Baseline characteristics*	Insulin treated (N=176)	No insulin (N=1341)	P value+
3	30 (17)	232 (17)	
4	24 (14)	191 (14)	
5-10	65 (37)	479 (36)	
11-15	4 (2)	69 (5)	
16-20	4 (2)	4 (0.1)	
Total	176 (100)	1341 (100)	
Number of abortions (n=688)	1.8 ± 1	1.7 ± 1	0.49
Number of live births (n=1516)	3.7 ± 2.5	4 ± 2.7	0.76
Cumulative lactation (months) (n=1518)	32 ± 51	32± 47	0.85
Baseline diabetes data			
Age at diabetes diagnosis (y)	48.3 ± .13	56.6 ± .12	<0.01
Diabetes duration at baseline (y)	14.7 ± 9.0	6.6 ± 7.3	<0.01
Family history of diabetes	110 (62.5)	684 (51.0)	<0.05
Glucose Lowering Interventions (n, row %)	176 (11.6)	1341 (88.4)	
Insulin treated diabetics	176 (100)		
Insulin monotherapy	96 (54.5)		
Insulin combination therapy	80 (45.5)		
with sulfonylurea	19		
with metformin	44		
with sulfonylurea and metformin	14		
with other GLD‡	3		
Non-insulin treated diabetic		1341 (100)	
Metformin		201 (14.9)	
Sulfonylurea		472 (35.2)	
Metformin and Sulfonylurea		335 (24.9)	
Other GLD		73 (5.4)	
Diet and/or nutritional supplements, and exercise		264 (19.7)	
Other baseline diseases			
Metabolic disease			
Dyslipidemia	110 (62.5)	859 (64.1)	0.68
Osteoporosis*(includes Osteopenia)	6 (3.4)	99 (7.3)	0.05
Metabolic Syndrome	117 (66.5)	757 (56.5)	0.01
Hyperuricemia	5 (2.8)	71 (5.3)	0.2
Other noncommunicable disease			
Hypertension	164 (93.2)	1182 (88.1)	0.047
Stroke and/or MI	18 (10.2)	95 (7.1)	0.9
Stroke	6 (3.1)	50 (3.7)	0.83
MI	12 (6.8)	52 (3.8)	0.07
Cancer	6 (3.4)	36 (2.7)	0.62
Lung Disease	3 (1.7)	8 (0.6)	0.13
Liver Disease	8 (4.6)	15 (1.2)	<0.01
Kidney disease	16 (9.1)	28 (2.1)	<0.01
Dementia	0	2 (0.1)	0.61
Communicable disease			
Tuberculosis	3 (1.7)	13 (0.97)	0.42
HIV	1 (0.6)	0	0.12
Diet & habits §			
Active Smoking	12 (7)	49 (3)	0.045
Smoking History	49 (28)	305 (23)	0.13
Alcohol Use	8 (4%)	81 (6)	0.43
Exercise	52 (30)	415 (31)	0.7
Healthy diet||	75 (46)	615 (46)	0.45
Healthy lifestyle@	17 (10)	166 (12)	0.3

*Categorical variables presented as frequencies and column percentages (n, %), numerical variables as mean and standard deviation (SD) unless stated otherwise.

†duration of fertility at recruitment = age at menopause or age at baseline – age at menarche of if either is missing, calculate reproductive lifespan based on interval between the first and the 4th child (the latter being the last pregnancy for which age maternal age is recorded in the database.) On this basis, there were 39 subjects with missing information.

‡ includes unrecalled medication name, or DPP4,

§ active smoking- smoking of any tobacco product in any amount, weekly alcohol - intake of at least a single drink of alcoholic beverage once a week, regular exercise - at least 30 minutes of moderate activity at least weekly

|| Khera et al (12)

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

Metabolic syndrome was more frequent in the insulin- than non-insulin treated participants (67% *vs.* 56%). Dyslipidemia occurred in ~ two-thirds of both groups, whereas slightly more insulin-treated subjects had both dyslipidemia and osteoporosis. Seven percent of the population had osteoporosis, of whom 70% received pharmacologic treatment, which was twice as common in the noninsulin- compared to the insulin-treated group. Hypertension, cardiovascular, liver, lung, or kidney disease were more frequently reported in the insulin- than non-insulin treated group.

Mean onset of menarche was at 13 (± 1.8) years of age. Less than a quarter of the population (23%) reported onset of menopause (mean 48 ± 6 years). Parity in pregnancy was high in both study groups: 41% had > 4 and 5% had > 10 pregnancies. Ninety two percent of women had become pregnant (mean 6 ± 3 pregnancies) and delivered (mean 5 ± 2 deliveries) over their reproductive lifespan at baseline (mean 47 ± 13 years). Cumulative duration of lactation over successive pregnancies was a mean of 32 months. However, lactation duration was not reported in 39% of fertile women, presumed to have bottle fed their infants. Less than half of women (n=688, 45%) reported spontaneous abortions; contraceptive history is not available. Albeit rare in both study groups, active smoking was more commonly reported in the insulin-treated group. Alcohol use was uncommon. Daily exercise (30%) and a healthy diet (45%) were similar in the two treatment groups.

Whereas historical data on fasting and postprandial hyperglycemia was obtained for all participants, confirmatory biochemical test values were available in only 8-23% of the population at baseline and, therefore, not included in risk analysis. Similarly, bone mineral density data, available in <50% of subjects, was not used in risk analysis.

Treatment for diabetes

Insulin use either in mono- or combination therapy constituted 12% of the subjects (Table 1). Of the insulin-treated subjects, slightly more subjects were receiving insulin alone than in combination with oral GLDs. Glucose-lowering interventions in most non-insulin subjects consisted of sulfonylurea, metformin, or their combination, while 20% were treated with behavioral modification such as diet, exercise, nutritional supplements (e.g., dietary fiber, psyllium, chromium picolinate) or adjunctive therapy (e.g., lipid lowering with fibric acid). The proportion of GLD treated subjects increased over the period of recruitment (Figure 1). Not surprisingly, insulin treated subjects had more serious diabetes than their non-insulin treated counterparts as evidenced by a mean onset of diabetes of 48 years in insulin treated and 57 years in non-insulin treated subjects. The mean duration of diabetes was 15 years for insulin treated diabetics; approximately twice that observed for the non-insulin treated subjects (7 years). In the insulin-treated group, subjects recruited later in the study tended to be younger than those recruited earlier, compared to those that received no insulin (Figure 1, p=0.01). In the non-insulin treated group, subjects recruited later tended to have more long-standing diabetes (p <0.01). No such trend was evident in the insulin-treated group. Duration of diabetes and BMI remained essentially the same over the recruitment period in both groups (Figure 1).

All-cause mortality (ACM)

A total of 202 deaths occurred among 1517 diabetic females over 12,620 person years of observation. Of these, 24 deaths over 1193 person years occurred in the insulin treated (incidence rate =2.0 per 100 person-years) *vs.* 178 deaths (incidence rate=1.6 per 100 person-years) over 11, 427 person years in the non-insulin treated group. The median age of the time at censoring was similar between insulin- *vs.* non-insulin-treated subjects (70 *vs.* 73 years, respectively), whereas the median attained age at death was younger in the insulin *vs.* the non-insulin treated subjects (67 *vs.* 76 years) (Supplemental figure 4, Panel A). Median time to censoring was 6.3 years compared to 9.3 years in the insulin *vs.* the non-insulin treated subjects, respectively. Median time to death, was 4.8 years in the insulin treated *vs.* 5 years in the non-insulin treated cohorts (Supplemental figure 4, Panel B). Three subjects in the insulin treated and three in the non-insulin treated groups died at or below age 50: a 25-, 34- and 50-year-old subject died at 4.1, 3.1 and 2.9 years of observation in the insulin treated group and a 29-, 41- and 37-year-old with a time to death of 3.2, 1.4 and 3.7 years of observation, respectively in the non-insulin treated group. Among the insulin treated subjects, death rate was higher but time to death was longer (16 of 96 (16.7%) subjects died with a mean time to death of 2131 days) in the insulin monotherapy treated group, compared to subjects treated with insulin combined with an oral GLD, (8 of 80 (10%) subjects died with a mean time to death of 980 days).

Variable	N	coeff	SE	p	95% CI	Variable	N	coefficient	SE	p	95 % CI
Age	176	-.73	.28	0.01	-1.29, -.17	Age	1341	.26	.10	0.01	.56,.46
Duration	176	-.04	.22	0.85	-.47, .39	Duration	1341	.30	.06	0.00	.18, .42
BMI	173	-.02	.12	0.86	-.26, .22	BMI	1321	-.02	.04	0.62	-.10, .06

Figure 1 Proportion of insulin and oral GLD use (top) and trends in mean age, duration of diabetes and BMI (bottom) by year of recruitment for insulin and non-insulin treated cohorts.

Screening for ACM risk predictors by RSF and incorporation into PS-weighting

In our earlier study, RSF screening verified by Cox regression found age, diabetes, marital status, BMI and baseline comorbid conditions as significant ACM risks in all PROUD participants whereas reproductive and comorbid variables of fracture and cancer

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

were significant in female participants.[6] Hyperparameter tuning found a stable model with the top 50 VIMP, with parity, anthropometric or vital signs variable categories ranked high, followed by diet / behavior, metabolic /co-morbid disorders. GLD treatment, also significantly predicted ACM, defined treatment cohort assignment for this exploratory sub-study.

Table 2 lists variables in descending order by their VIMP score, significance in univariate regression, integration in the PS weights and inclusion in the final multivariable regression. RSF identified the known ACM risk factors in diabetes mellitus (Supplemental Table 4) and new variables not previously reported as ACM risks in diabetes mellitus (recruitment year, parity, and reproductive lifespan, the latter derived from baseline age or age at menopause minus age at menarche). Recruitment year [HR 3.8 95% CI (1.9, 7.6)] was a study-specific category. Of the reproductive and gestational variables, parity, and reproductive lifespan were retained in the logistic regression model to estimate the probability of treatment with insulin, selected based on their predictiveness for ACM. Thus, the weighted comparison of the treatment cohorts was based on variables known to predict ACM in diabetes (PS weighting) or known variables plus the new RSF identified variables (RSF complemented PS weighting) (Supplemental Table 1).

Table 2 Risk variables identified by RSF, their VIMP score, HR for ACM, consideration, and selection for PS weighing), and used in multivariable regression for all insulin comparisons

Hierarchy of 82 variables in 1517 females with diabetes mellitus	VIMP† Score	Univariate Cox HR (95%CI) for ACM in females with diabetes mellitus	Considered in PS analysis		Final variables included in	
			weighting	outcome	PS Weighting with Novel RSF variables bolded	Multivariable regression
Parity	0.084522	1.09 (1.0, 1.1)	x		**Parity**	Osteoporosis
BMI 1,3,*	0.069393	.93 (.90, .96)	x		**Reproductive lifespan**	Active smoking
Weight 2,3	0.063659	.97 (.95, .96)	x		**Year of recruitment‡**	Race
Waist circumference 3	0.054312	.98 (.97, .99)	x		**Age‡§**	Kidney disease
Maximum age mother 2	0.053557	1.02 (0.01, 1.02)	x		**WHO BMI categories**	
Duration of diabetes 2,*	0.053311	1.04 (1.0, 1.06)	x		Waist Hip ratio	
Age of first child	0.047738	1.02 (1.0, 1.0)	x		Duration of diabetes	
Reproductive lifespan**	0.045612	1.01 (1.0, 1.03)		x	Family history of diabetes	
WHO BMI categories	0. 040774	1.16 (1.1, 1.27)	x		Hypertension	
Height 3	0.040476	.96 (.94, .97)	x		Hypercholesterolemia	
Age 1,2*	0.040147	1.05 (1.0, 1.07)	x		Cancer	
Age at menarche 3	0.032738	1.11 (1.03, 1.19)	x		Cardiovascular disease	
Number of pregnancies 3	0.031677	1.08 (1.0, 1.1)	x		History of cigarette smoking	
Hip circumference 3	0.030277	.97 (.96, .99)	x		History of alcohol use	
Systolic BP 3	0.027123	1.01 (1.0, 1.01)	x			
Weekly alcohol use	0.023001	1.02 (1.01, 1.03)		x		
Duration of hypertension	0.021203	1.02 (1.00, 1.03)	x			
Diastolic blood pressure *	0.019218	1.00 (.99, 1.01)	x			
Lactation in months	0.018834	1.0 (1.00 ,1.00)	x			
Metabolic overlap	0. 017533	1.00 (.90, 1.12)		x		
Waist hip ratio	0.017378	2.03 (.21, 19.3)	x			
Heart rate 1	0.015727	1.00 (.99, 1.02)	x			
Smoking duration 2	0.011598	1.01 (1.0, 1.02)	x			
Cancer History 1	0.011259	1.97 (1.01, 3.85)	x			
Number of abortions	0.01084	1.09 (.98, 1.22)	x			
Active smoking 3*	0.010804	1.98 (1.2, 3.3)		x		
CVD *	0.010561	1.95 (1.34, 2.84)	x			

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

Table 2 Continued...

Hierarchy of 82 variables in 1517 females with diabetes mellitus	VIMP† Score	Univariate Cox HR (95%CI) for ACM in females with diabetes mellitus	Considered in PS analysis		Final variables included in	
			weighting	outcome	PS Weighting with Novel RSF variables bolded	Multivariable regression
Duration of dyslipidemia	0.008423	1.00 (.95, 1.05)		x		
Ischemic CVD 3	0.008343	2.06 (1.31, 3.23)		x		
Coffee cups / day	0.006964	1.02 (.98. 1.17)	x			
Kidney disease*	0.00635	1.00 (.99, .41)		x		
Liver disease	0.005848	.52 (.73, 3.71)	x			
Family history of diabetes 3	0.005457	.62 (.47, .82)	x			
Metabolic syndrome	0.005452	.54 (.38, .76)		x		
Duration of osteoporosis	0.005439	.97 (.89, 1.08)		x		
Duration of hypercholesterolemia	0.00454	1.0 (.96, 1.04)	x			
Daily fruit intake 1	0.004497	.75 (.54, 1.05)		x		
Stroke 1,3	0.001738	2.1 (.89, 4.99)	x			
Respiratory disease	0.003859	.75 (.31, 1.83)	x			
Osteoporosis	0.003768	.88 (.51, 1.51)		x		
History of MACE	0.003663	1.65 (1.2. 2.2)	x			
Hypertensive CVD	0.003423	.99 (.14, 7.14)	x			
Recruitment year 2,3	0.003402	3.8 (1.9, 7.6) ‡	x			
CVD history & treatment	0.003299	1.6 (1.1, 2.45)	x			
Dyslipidemia	0.002844	.68 (.51, .91)		x		
Insulin monotherapy *	0.002793	1.51 (.90, 2.51)				
Family history of hypertriglyceridemia	0.002755	.93 (.57, 1.5)	x			
Osteoporosis treatment 3	0.002573	.78 (.41, 1.5)	x			
Marital status (binary)	0.000368	1.0 (.80, 1.3)		x		
History of Alcohol Use	0.001727	.63 (.32, 1.22)	x			
Number of lipid disorders	0.001505	.85 (.78, .94)	x			
Dyslipidemia with Lab tests	0.001369	.43 (.29, 63)		x		
Healthy behavior	0.001196	.78 (.49, 1.2)		x		
Family History of Cancer	0.001174	.67 (.39, 1.1)	x			
Triglyceride treatment	0.000911	1.2 (.87, 1.78)		x		
Fish at least 2X /week	0.00088	.80 (.61, 1.06)	x			
Family history of Cholesterol	0.000679	.79 (.52, 1.2)	x			
All metformin therapy	0.000569	.97 (.72, 1.3)				
History of Hypertriglyceridemia	0.000514	1.1 (.82, 1.5)		x		
Race	0.000494	.90 (.73, 1.11)		x		

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

Table 2 Continued...

Hierarchy of 82 variables in 1517 females with diabetes mellitus	VIMP† Score	Univariate Cox HR (95%CI) for ACM in females with diabetes mellitus	Considered in PS analysis		Final variables included in	
			weighting	outcome	PS Weighting with Novel RSF variables bolded	Multivariable regression
Combination GLD treatment	0.000417	1.2 (.85, 1.6)				
Marital Status	0.000368	1.0 (.95, 1.1)		x		
Baseline Lung Disease	0.000303	1.6 (.38, 6.25)	x			
Metformin monotherapy	0.000299	.66 (.40, 1.1)				
Cholesterol treatment*	0.000213	1.1 (.81, 1.44)		x		
Family history of osteoporosis	0.000154	.38 ((.16, .93)	x			
Sulfonylurea monotherapy	0.000134	1.4 (1.1, 1.9)				
Hypertension 1	0.00005	1.6 (1.1, 2.9)	x			
History of Alzheimer's disease	0	0		x		
HIV	0	0	x			
Daily alcohol intake	0	0	x			
History of Infertility	0	.81 (.20, 3.3)	x			
Hypercholesterolemia on baseline tests	-0.00006	.33 (.12, .90)		x		
Treatment for hypertension*	-1.6E-05	1.4 (1.0, 1.8)		x		
History of hypercholesterolemia	-0.00021	1.1 (.82, 1.4)		x		
Exercise 1	-0.00024	.62 (.45, .87)		x		
History of smoking*	-0.00028	1.65 (1.23, 2.23)	x			
Healthy diet	-0.00034	.86 (.65, 1.13)		x		
Raised non HDLC Lab tests	-0.00045	.67 (.43, 1.0)		x		
Family history of obesity	-0.00077	.78 (.46, 1.3)	x			
Tuberculosis	-0.00087	1.9 (.72, 5.2)	x			
Any Treatment for diabetes	-0.00099	1.5 (1.2, 2.0)				
Socioeconomic stratum	-0.00138	.89 (.81, .995)	x			

For select definitions, see footnotes to Table 1. Bolded items in column 1 are variables included in the final multivariable regression (column 7). Bolded items in column 6 are novel ACM risk variables identified through RSF screening.

* identified in published ACM risk scores for diabetes or integrated into treatment guidelines for management of diabetes. CVD=cardiovascular disease

** Reproductive lifespan = age at baseline or menopause minus age at menarche+1 (if missing, maternal age at fourth child minus maternal age at first child or menarche+1)

1,2,3 significant in multivariable Cox regression in participants of both sexes, females, females with diabetes mellitus, respectively.

† from Lu, Ishwaran et al(15)

‡ analyzed as a categorical value by year from 2003-2016

§ analyzed as a categorical value => 50 years based on the Cigolle et al (38)

Start of diabetes used to define duration of diabetes and not independently assessed.

Treatment variables used to develop cohorts and not used for PS- weighting

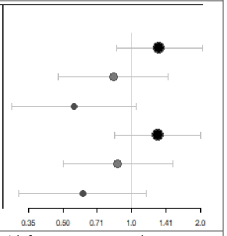
Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

Survival analysis in insulin treatment cohorts and four major risk variables

The univariate unweighted comparison of insulin *vs.* non-insulin treated cohorts suggested an increased risk for ACM, whereas the PS-weighted risk estimate moved in the opposite direction (Figure 2A, Supplemental Table 6). However, neither were statistically significant on the ACM risk. Incorporation of the novel RSF risk variables resulted in 40% reduction in ACM risk estimate that approached significance (p=.07). The same trend was observed in the multivariable analyses for unweighted, PS-weighted and RSF complemented PS-weighted analyses. Of further note is that, albeit statistically insignificant, the estimate of ACM risk from insulin monotherapy (Supplemental Table 7), was found protective in the RSF complemented PS-weighted univariate and multivariable sensitivity analyses, whereas the unweighted and PS-weighted estimates indicated increased risk in both univariate and multivariable analyses. The Kaplan Meier analyses also suggest a protective insulin effect on the cumulative probability of survival in the RSF complemented PS-weighted analysis (Supplemental Figure 5). Of the 4 variables included in the multivariable Cox regression, only smoking was significantly associated with mortality in the univariate analysis [HR (95% CI) of 1.98 (1.2, 3.3), p= .01)] (Table 2, column 3). Although statistically significant in all multivariable analyses, the two-fold increased ACM risk estimate observed with the unweighted analysis for active smoking was magnified in the PS-weighted estimates whereas a more modulated risk estimate was obtained in the RSF complemented PS-balanced analysis (Figure 2B and Supplemental Figure 6). Neither osteoporosis, race, nor kidney disease was associated with an increased ACM risk in the unweighted or PS-weighted comparisons (with or without RSF variables). In fact, the pattern of HR estimates for osteoporosis resembled those of insulin treatment: a nonsignificant trend toward a protective effect in both univariate and multivariable analyses.

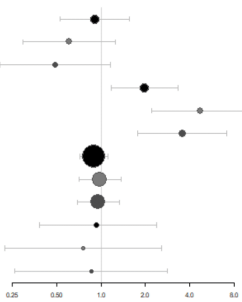
A. Insulin *vs.* Non-Insulin Treatment

	Insulin Use	HR	Conf. Int	P-value
Univariate	Unweighted	1.3	(0.9, 2.0)	0.2
	PS Weighted	0.8	(0.5, 1.4)	0.5
	PS-RSF Weighted	0.6	(0.3, 1.0)	0.07
Multivariable	Unweighted	1.3	(0.8, 2.0)	0.2
	PS Weighted	0.9	(0.5, 1.5)	0.6
	PS-RSF Weighted	0.6	(0.3, 1.2)	0.1



B. Additional risk factors in multivariable analysis comparing subjects with factor present *vs.* absent

Factor	Method	HR	Conf. Int	P-value
Osteoporosis	Unweighted	0.9	(0.5, 1.6)	0.7
	PS Weighted	0.6	(0.3,1.3)	0.2
	PS-RSF Weighted	0.5	(0.2,1.2)	0.1
Active Smoking	Unweighted	2.0	(1.1,3.4)	0.01
	PS Weighted	4.7	(2.2,10.2)	<0.01
	PS-RSF Weighted	3.6	(1.8,7.2)	<0.01
Race	Unweighted	0.9	(0.7,1.1)	0.3
	PS Weighted	1.0	(0.7,1.4)	0.9
	PS-RSF Weighted	1.0	(0.7,1.3)	0.8
Kidney Disease	Unweighted	0.9	(0.4,2.3)	0.9
	PS Weighted	0.8†	(0.2,2.6)	0.7
	PS-RSF Weighted	0.9	(0.3,2.8)	0.8



† The value of HR for PS weighted kidney disease is 0.86.

Figure 2 Forest plot of unweighted (top), PS-weighted (middle), and PS with novel RSF variable-weighted (bottom) hazard ratio and 95% CI estimates for ACM from insulin treatment (panel A) and from osteoporosis, active smoking, race, and kidney disease (Panel B).

Discussion

Despite the recognized efficacy of glucose-lowering with insulin, reported adverse effects and increased risk for mortality raise uncertainty about the net benefit of insulin treatment.[21–24] When the insulin and non-insulin treated cohorts were well balanced on risk factors known to predict ACM, our analysis found no significant effect of insulin on ACM, with a risk estimate suggesting potential benefit. Incorporation of new variables identified by machine learning (RSF) in weighted PS analysis, resulted in a more conservative risk estimate that needs confirmation in a larger study. A meta-analysis of randomized controlled trials concluded that insulin treatment does not increase ACM, with 80% of the effect driven by a large trial of basal insulin[25] that included older adults and Hispanic women. The meta-analysis further found no increased risk for cardiovascular outcomes 3-component MACE (myocardial infarction, stroke, cardiovascular death) and heart failure. Our report from observational data obtained in the community practice from Barranquilla, Colombia came to similar conclusions, made possible by PS weighting enhanced by the incorporation of RSF identified variables. Among adult diabetic females in Barranquilla, Colombia insulin treatment serves as a proxy for a more severe or advanced disease compared to subjects receiving other glucose-lowering therapies.

In this community-based health practice setting where participants were followed for mortality for a median of 10 years, an excess of deaths per person year and a younger attained age at death was observed in the insulin-treated, compared to the non-insulin treated cohort. The unweight PS estimate implied an increased risk for ACM with insulin treatment whereas addition of RSF-identified novel variables in the PS weighted comparison revealed a 40% ACM risk reduction from insulin treatment. Although statistical significance is limited by the study size, this effect was consistently observed in the participants receiving insulin as mono- or combination therapy compared to non-insulin treatment in both univariate and multivariable analyses.

Our previous report identifying diabetes and gender specific variables as ACM risk predictors identified by RSF in over 12,000 PROUD participants[5] motivated this follow-on sub-study of diabetic females. In the present study, domain expert involvement in data collection, processing and item reduction, and employment of standard definitions for derived variables enhance understandability of machine learning methods. RSF identified variables important for mortality by multiple, independent resampling of the data (trees in the forest) without pre specification[15] and is thus not susceptible to overfitting.[26] PS variable selection prioritizes variables already known by experts to be associated with either treatment choice or outcome. That is, discovery of new variables by PS is not the analytic goal, whereas combining it with a machine learning application such as RSF in pharmaco-epidemiology has the potential to uncover new variables not previously established to affect outcome.

PS-weighting enables causal estimation from observational studies by balancing known outcome modifiers[27] derived from large population-based studies in diabetes (Supplemental Table 4). RSF identified the main risk predictors, save for renal biomarkers which were not collected in PROUD. Thus, regardless of the analytic methods employed, unmeasured confounders remain a limitation even in high density data and can be an artifact of local testing conditions in pharmacoepidemiology studies. RSF identified variables relevant to mortality in females; studies conducted for cardiovascular, and ACM confirm the importance of these variables. In the reproductive and cardiovascular literature[16–19] parity and reproductive lifespan are cited

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

as risks for ACM (Supplemental Table 4, Panel B).[28] Moreover, parity is a reported risk for diabetes in Colombian women.[29] Our literature search of ACM risk variables in diabetes found only one study that specifically assessed ACM risks in females with diabetes mellitus[30] but did not include covariates related to reproduction. Exploration of parity and duration of fertility in the full PROUD data and in the subset of women with diabetes suggest a non-linear relationship with ACM (data not shown). The impact of parity and duration of fertility on ACM will need to be assessed in a larger cohort of female diabetics to confirm our findings, given that PS weighting is susceptible to model misspecification compounded by variables with a non-linear relationship with the outcome of interest (Supplemental Appendix).[31]

The large number of variables, including reproduction, long-term follow-up for mortality and incorporation of Colombian standards[11] for the characteristics assessed (e.g. obesity classification) are strengths of the study. On the other hand, diabetes subclassification, HbA1c values and a history of gestational diabetes was not collected, limiting generalization of our study findings. GLD treatment is limited to data at baseline and treatment modification during follow-up was not collected, thus attribution of 10-year survival to treatment at baseline has limitations. Nonetheless, the Kaplan Meier survival curves in the RSF complemented PS weighted analyses show an early survival drop off in the non-insulin treated cohort while sustained benefit over follow-up time is observed in the insulin treated cohort.

Data provenance and concept definition are important when interpreting existing data in real world settings. Studies that used insulin treatment assignment as a surrogate for diabetes phenotypes in identifying participants from real-world resources[32,33] found ICD coding and laboratory testing of value, both limitations in our study. Biomarkers for incipient kidney disease were lacking and bone densitometry was missing in approximately 50% of our participants. Based on validated surveillance definition buttressed by treatment data, we found an osteoporosis prevalence lower than the 35% reported by Londono et al.,[32] and closer to the 2.4% rate of osteoporosis reported by Fernández-Ávila[34] in the general Colombian population. With our definition,[28,35] a protective trend was noted for ACM in subjects with osteoporosis.

Our ML model developed in the larger PROUD population motivated this investigation (insulin treatment as a risk variable) and illustrates the stable model that identified novel ACM covariates across the larger PROUD population. Participants in PROUD were predominantly females; there were too few male subjects for inclusion. We note also that the ranking of the top variables (performed on the entire population) did not parallel the size of the estimated HR estimates across the compared groups; initial univariate estimates found very small increments in ACM risk from parity or reproductive lifespan for example. VIMP values are indicative of their potential importance in predicting outcome, whereas PS weighted comparisons are conducted to infer an effect attributable to the intervention after eliminating confounding. For example, VIMP variables were larger for the reproductive variables compared to active smoking, whereas, despite the low prevalence of smoking in the population, active smoking emerged as the dominant risk factor for ACM; as previously reported in women at least 35 years of age.[36] Incorporation of RSF identified variables in PS analyses that routinely prioritizes subject matter defined variables, has the potential to provide additional insight in outcomes relevant to patient subgroups.

In addition to the comparatively small study population, the above study limitations need to be considered in interpreting the generalizability of our conclusions. Replication of these analyses in adequately sized populations (age, sex, and race/ethnicity) may better enable assessment of insulin effect. Nonetheless, prospective long-term studies in underrepresented populations may be infeasible or ethically challenging to conduct with therapeutic products that have become standard of care. PROUD was a large study of volunteers from a community practice, with long term follow-up of a population underrepresented in randomized controlled trials. Information on parity and reproductive lifespan on ACM has led to updated recommendations on prevention of cardiovascular disease in women.[37] As CVD is a major component of ACM in diabetic patients, we encourage evaluation of the impact of reproductive factors in larger cohorts of females with diabetes mellitus.[38]

Acknowledgments

I. **Acknowledgements:** Drs. Frank Pucino, Anna Kettermann, and Yves Rosenberg for critical thinking and insightful discussion.

II. **Funding and assistance:** The study received no external funding and is made possible by the voluntary contributions of the Barranquilla community and the MATIG authorship.

III. **Author contributions and guarantor statement**

CCC, ENA, YG, JDE, PCR, AH, CW, RG, XT, GS, GC, and AH were involved in the conception, design, and conduct of the study and/or the data processing, analysis, and interpretation of the results. EN wrote the first draft of the manuscript, and all authors edited, reviewed, and approved the final version of the manuscript. CCC is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

MATIG members

Keith Burkhart,[2] Karim Calis,[2] Iffat Chowdhury,[2] Sean Coady,[3] Gyorgy Csako,[3] Carlos Cure Cure,[1] Pablo Cure,[3,7] Victor Crentsil,[2] Gauri Dandi,[3] Amit Dey,[3] Michael Domanski,[4] John Eustaquio,[5] Jerome Fleg,[3,2] Yuan Gu,[4,6] Noah Hasan,[3] Anwar Hussain,[2] Danielle Jateng,[2] Anna Kettermann,[2] Eileen Navarro Almario,[3,8] Ramiro Orozco,[1] Frank Pucino,[2] Ahmed A Hasan,[3,4] Tejas Patel,[2] Yves Rosenberg,[3] George Sopko,[3] Bereket Tesfaldet,[2] Xin Tian,[3,4] Colin O Wu,[9,4] Victoria Xin[3]

[1]Biomelab SAS, Barranquilla, Colombia

[2]Food and Drug Administration

[3]Meta-Analytical Inter-Agency Group (MATIG), Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

[4]Data Science Initiative, Cardiovascular Research Laboratories, University of Maryland School of Medicine, Baltimore, MD, USA

[5]University of Maryland in Baltimore County, MD, USA

[6]Department of Statistics, the George Washington University, Washington, DC, USA

[7]National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA

[8]Paraclete Professionals, LLC, Clarksville, MD, USA

[9]Office of Biostatistics Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

Data availability

The datasets generated during and/or analyzed during the current study are not publicly available due to ongoing assessments for additional publication(s) but are available from the corresponding author upon reasonable request.

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

Disclaimer

The manuscript's content is solely the responsibility of the authors and does not represent the official views of the National Heart, Lung, and Blood Institute, National Institutes of Health; the National Center for Advancing Translational Sciences, National Institutes of Health; the Food and Drug Administration or the United States Department of Health and Human Services.

Conflicts of interest

All authors declare no conflicts of interests in relation to the work presented in this manuscript.

References

1. Sütő G, Molnár GA, Rokszin G, et al. Risk of morbidity and mortality in patients with type 2 diabetes treated with sodium-glucose cotransporter-2 inhibitor and/or dipeptidyl peptidase-4 inhibitor: a nationwide study. *BMJ Open Diabetes Res Care*. 2021;9(1):e001765.
2. Gerstein H, Yusuf S, Riddle MC, et al. Rationale, design, and baseline characteristics for a large international trial of cardiovascular disease prevention in people with dysglycemia: the ORIGIN Trial (Outcome Reduction with an Initial Glargine Intervention). *Am Heart J*. 2008;155(1):26-32:32.e1-6.
3. Downing NS, Shah ND, Neiman JH, et al. Participation of the elderly, women, and minorities in pivotal trials supporting 2011–2013 U.S. Food and Drug Administration approvals. *Trials*. 2016;17(1):199.
4. Patrick AR, Schneeweiss S, Brookhart MA, et al. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. *Pharmacoepidemiology and Drug Safety*. 2011;20(6):551–559.
5. Tazare J, Wyss R, Franklin JM, et al. Transparency of high-dimensional propensity score analyses: Guidance for diagnostics and reporting. *Pharmacoepidemiology and Drug Safety*. 2022;31(4):411–423.
6. Cure-Cure CA, Cure P, Gu Y, et al. Abstract 16252: Predictors of all cause mortality and their gender differences in a hispanic population from barranquilla-colombia using machine learning with random survival forests. *Circulation*. 2018;138(Suppl_1):A16252.
7. Cure P, Hoffman HJ, Cure-Cure C. Parity and diabetes risk among hispanic women from Colombia: cross-sectional evidence. *Diabetology & metabolic syndrome*. 2015;7(1):1–5.
8. Camacho PA, Otero J, Pérez M, et al. The spectrum of the dyslipidemia in Colombia: The PURE study. *International Journal of Cardiology*. 2019;284:111–117.
9. Cure-Cure C, Capozza RF, Cointry GR, et al. Reference charts for the relationships between dual-energy X-ray absorptiometry-assessed bone mineral content and lean mass in 3,063 healthy men and premenopausal and postmenopausal women. *Osteoporosis International*. 2005;16(12):2095–2106.
10. McHale TC, Romero-Vivas CM, Fronterre C, et al. Spatiotemporal Heterogeneity in the Distribution of Chikungunya and Zika Virus Case Incidences during their 2014 to 2016 Epidemics in Barranquilla, Colombia. *Int J Environ Res Public Health*. 2019;16(10):1759.
11. Ramírez-Vélez R, Correa-Bautista JE, Martínez-Torres J, et al. LMS tables for waist circumference and waist–height ratio in Colombian adults: analysis of nationwide data 2010. *European Journal of Clinical Nutrition*. 2016;70(10):1189–1196.
12. Khera AV, Emdin CA, Drake I, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *New England Journal of Medicine*. 2016;375(24):2349–2358.
13. DANE. *Population-and-demography/births-and-deaths Colombia: Colombia National administrative department of statistics*. 2017.
14. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circ Res*. 2017;121(9):1092–1101.
15. Lu M, Ishwaran H. A prediction-based alternative to P values in regression models. *J Thorac Cardiovasc Surg*. 2018;155(3):1130–1136. e4.
16. Mishra SR, Chung HF, Waller M, et al. Duration of estrogen exposure during reproductive years, age at menarche and age at menopause, and risk of cardiovascular disease events, all-cause and cardiovascular mortality: a systematic review and meta-analysis. *Bjog*. 2021;128(5):809–821.
17. Carlqvist E, Johnson L, Nilsson PM. Shorter reproductive life span is associated with increased cardiovascular risk and total mortality in Swedish women from an observational, population-based study. *Maturitas*. 2022;164:69–75.
18. Li X, Wang S, Dunk M, et al. Association of life-course reproductive duration with mortality: a population-based twin cohort study. *American Journal of Obstetrics and Gynecology*. 2022;227(5):748.e1–748.e13.
19. Ley SH, Li Y, Tobias DK, et al. Duration of reproductive life span, age at menarche, and age at menopause are associated with risk of cardiovascular disease in women. *J Am Heart Assoc*. 2017;6(11):e006713.
20. Chesnaye NC, Stel VS, Tripepi G, et al. An introduction to inverse probability of treatment weighting in observational research. *Clin Kidney J*. 2022;15(1):14–20.
21. Gamble JM, Chibrikov E, Twells LK, et al. Association of insulin dosage with mortality or major adverse cardiovascular events: a retrospective cohort study. *Lancet Diabetes Endocrinol*. 2017;5(1):43–52.
22. Price HI, Agnew MD, Gamble JM. Comparative cardiovascular morbidity and mortality in patients taking different insulin regimens for type 2 diabetes: a systematic review. *BMJ Open*. 2015;5(3):e006341.
23. Nyström T, Bodegard J, Nathanson D, et al. Second line initiation of insulin compared with DPP-4 inhibitors after metformin monotherapy is associated with increased risk of all-cause mortality, cardiovascular events, and severe hypoglycemia. *Diabetes Res Clin Pract*. 2017;123:199–208.
24. Roumie CL, Greevy RA, Grijalva CG, et al. Association between intensification of metformin treatment with insulin vs sulfonylureas and cardiovascular events and all-cause mortality among patients with diabetes. *Jama*. 2014;311(22):2288–2296.
25. Mannucci E, Targher G, Nreu B, et al. Effects of insulin on cardiovascular events and all-cause mortality in patients with type 2 diabetes: A meta-analysis of randomized controlled trials. *Nutrition, Metabolism and Cardiovascular Diseases*. 2022;32(6):1353–1360.
26. Wang H, Li G. A Selective Review on Random Survival Forests for High Dimensional Data. *Quant Biosci*. 2017;36(2):85–96.
27. Thomas LE, Li F, Pencina MJ. Overlap Weighting: A Propensity Score Method That Mimics Attributes of a Randomized Clinical Trial. *JAMA*. 2020;323(23):2417–2418.
28. Kravdal Ø, Tverdal A, Grundy E. The association between parity, CVD mortality and CVD risk factors among Norwegian women and men. *Eur J Public Health*. 2020;30(6):1133–1139.
29. Cure P, Hoffman HJ, Cure Cure C. Parity and diabetes risk among hispanic women from Colombia: cross-sectional evidence. *Diabetology & Metabolic Syndrome*. 2015;7(1):7.
30. Wan EYF, Fong DYT, Fung CSC, et al. Prediction of five-year all-cause mortality in Chinese patients with type 2 diabetes mellitus - A population-based retrospective cohort study. *J Diabetes Complications*. 2017;31(6):939–944.
31. Shiba K, Kawahara T. Using Propensity Scores for Causal Inference: Pitfalls and Tips. *J Epidemiol*. 2021;31(8):457–463.

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control*. 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267

32. Londono J, Valencia P, Santos AM, et al. Risk factors and prevalence of osteoporosis in premenopausal women from poor economic backgrounds in Colombia. *Int J Womens Health*. 2013;5:425–430.

33. Richesson RL, Rusincovitch SA, Wixted D, et al. A comparison of phenotype definitions for diabetes mellitus. *J Am Med Inform Assoc*. 2013;20(e2):e319–e326.

34. Fernández Ávila DG, Bernal Macías S, Parra MJ, et al. Prevalence of osteoporosis in Colombia: Data from the National Health Registry from 2012 to 2018. *Reumatol Clin (Engl Ed).* 2021;17(10):570–574.

35. Leslie WD, Lix LM, Yogendran MS. Validation of a case definition for osteoporosis disease surveillance. *Osteoporosis International*. 2011;22(1):37–46.

36. Prevention CfDCa. *Smoking Cessation: A report of the Surgeon General*. The Benefits of Smoking Cessation on Overall Morbidity, Mortality, and Economic Costs. 2020.

37. Cho L, Davis M, Elgendy I, et al. Summary of Updated Recommendations for Primary Prevention of Cardiovascular Disease in Women: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2020;75(20):2602–2618.

38. Cigolle CT, Blaum CS, Lyu C, Ha J, Kabeto M, Zhong J. Associations of age at diagnosis and duration of diabetes with morbidity and mortality among older adults. *JAMA Network Open*. 2022;5(9):e2232766.

Citation: Cure CC, Almario EEN, Gu Y, et al. Does insulin therapy affect all-cause mortality? machine learning complements propensity score analysis in a pharmacoepidemiologic study of adult diabetic females in Barranquilla, Colombia. *J Diabetes Metab Disord Control.* 2023;10(2):144–155.
DOI: 10.15406/jdmdc.2023.10.00267