# Effect of statin and other cholesterol-lowering drugs on the incidence of type-II diabetes in non-diabetic patients.

Submitted By: Bibin Joseph

1. **INTRODUCTION** –
This study investigates the impact of statin and other cholesterol-lowering drugs on the incidence of type II diabetes in non-diabetic patients.

2. METHODS
**Study Design**- A retrospective cohort study to evaluate the long-term effects of statins and other cholesterol-lowering drugs

**Study Setting** – The data is provided by the professor and is taken from the Fairview Health system.

**Study Type** – A retrospective cohort study

**Study Period**- The Patients of Interest were selected from the period 2010-2018

**Baseline Definition**: For statin users, the date of initial statin prescription; for non-users, a matched date based on age and sex.

**Incident Type-II Diabetes Definition:** Clinical recognition post-baseline, including diagnosis codes, medication initiation, or A1C increase.

**Inclusion criteria**- Patients 30 years and older, non-diabetic at baseline, with specific health measurements.

**Exclusion criteria** - Participants were excluded if they had pre-existing diabetes, incomplete medical records, no blood pressure measurements around the index date, a history of diabetes medication or diagnosis before the index date, were under 30 years old, or were in the control group and taking cholesterol medication.

**Exposure-** Statin use vs. No Statin drugs for at least 3 years.

**Outcome**- Type-2 diabetes diagnosis within 0-5 years post-index date, indicated by A1C > 6.5% or diagnosis code

**Handling Confounding**: Using propensity score matching to balance characteristics between exposed and unexposed groups.

**Statistical Analysis:** Cox proportional hazards model to estimate hazard ratios, checking for proportional hazards and linearity assumptions.

2.2**. Data**

The data elements collected in the study include patient demographics (age, sex, race), health status indicators (blood pressure, A1C levels, cholesterol levels, smoking status, medication history), and outcomes (diagnosis of type 2 diabetes). Missing values are addressed through the exclusion criteria in the study design, removing patients with incomplete data or those not meeting specific criteria (e.g., missing blood pressure measurements or A1C levels). The final combined data table integrates these elements to facilitate analysis.

Table 1. Summary of Data Summary

| **Variable** | **Count** | **Mean** | **SD** | **Missing percentage** |
|---|---|---|---|---|
| Age at lndex date | 45063 | 55.52 | 14.52 | 0.00 |
| Maximum Idl | 20141 | 112.7 | 36.35 | 55.30 |
| Maximum trigliceride | 20910 | 147.9 | 167.38 | 53.60 |
| Maximum hdl | 21162 | 51.22 | 16.47 | 53.04 |
| Maximum sbp | 31871 | 137 | 17.37 | 29.27 |
| Maximum dbp | 31871 | 84.16 | 11.28 | 29.27 |
| Maximum pulse | 30583 | 87.46 | 16.58 | 32.13 |
| Maximum bmi | 29958 | 38.22 | 425.71 | 33.52 |
| | | | | |
| | | | | |
| **Variable** | **Category** | **Count** | **Percentage** | |
| Sex | F | 26495 | 58.80 | |
| | M | 18568 | 41.20 | |
| Race | American Indian or Alaska Native | 322 | 0.71 | |
| | American Indian or Alaska Native, Asian | 1 | 0.00 | |
| | American Indian or Alaska Native, Black or African American | 5 | 0.01 | |

| | | | | |
|---|---|---|---|---|
| | American Indian or Alaska Native, Black or African American, White | 1 | 0.00 | |
| | American Indian or Alaska Native, White | 32 | 0.07 | |
| | Asian | 1078 | 2.39 | |
| | Asian, Black or African American | 1 | 0.00 | |
| | Asian, Native Hawaiian or Other Pacific Islander, White | 1 | 0.00 | |
| | Asian, White | 13 | 0.03 | |
| | Black or African American | 2658 | 5.90 | |
| | Black or African American, Black or African American | 7 | 0.02 | |
| | Black or African American, Native Hawaiian or Other Pacific Islander | 1 | 0.00 | |
| | Black or African American, White | 19 | 0.04 | |
| | Native Hawaiian or Other Pacific Islander | 37 | 0.08 | |
| | Native Hawaiian or Other Pacific Islander, White | 8 | 0.02 | |
| | White | 40879 | 90.72 | |
| Dead | Dead | 154 | 0.34 | |
| | Alive | 44909 | 99.66 | |
| Smoker status | Never | 18802 | 41.72 | |
| | Not Asked | 1751 | 3.89 | |
| | Passive | 235 | 0.52 | |
| | Quit | 10918 | 24.23 | |
| | Unknown | 4565 | 10.13 | |
| | Yes | 8792 | 19.51 | |
| Med | No Medication | 1548 | 3.44 | |
| | Non-htn medication | 38715 | 85.91 | |
| | Non-Statin | 915 | 2.03 | |
| | Statin | 3885 | 8.62 | |
| Statin use | No | 41178 | 91.38 | |
| | Yes | 3885 | 8.62 | |
| Outcome | No Diabetes | 31437 | 69.76 | |
| | Diabetes | 13472 | 29.90 | |
| | Dead | 154 | 0.34 | |

**Missing values**

Table 1. It shows the summary table, and as you can see, many of the variables are missing, and the missing percentage is from 0-50%. So, imputing the data with mean or median will create bias in the modeling. The missing values are removed from the data, and that data is used for the final model.

**Statistical Analysis**

Propensity model: "Propensity score matching (PSM) is a statistical technique used in observational studies to estimate the causal effect" of a treatment or intervention on an outcome variable. The variables Age at Index date, sex of the patient, race of the patient, maximum BMI, smoking status, maximum LDL, maximum HDL, and maximum triglycerides are used for Propensity score matching.

The dataset with missing variables contains 13591 variables. After propensity matching, the number of variables in the dataset became 3232, so we can conclude that a total of 3232 patients are matched using Propensity score matching.

Common support of exposed and unexposed patients. Since I have done propensity score matching (PSM) for the data. It ensures that for every treated subject, there is a comparable control subject with a similar propensity score. This ensures that there is overlap, which is crucial for valid comparison and to avoid extrapolation beyond the range of observable data.

The balance for confounders after PSM is done by visualization and T-tests.

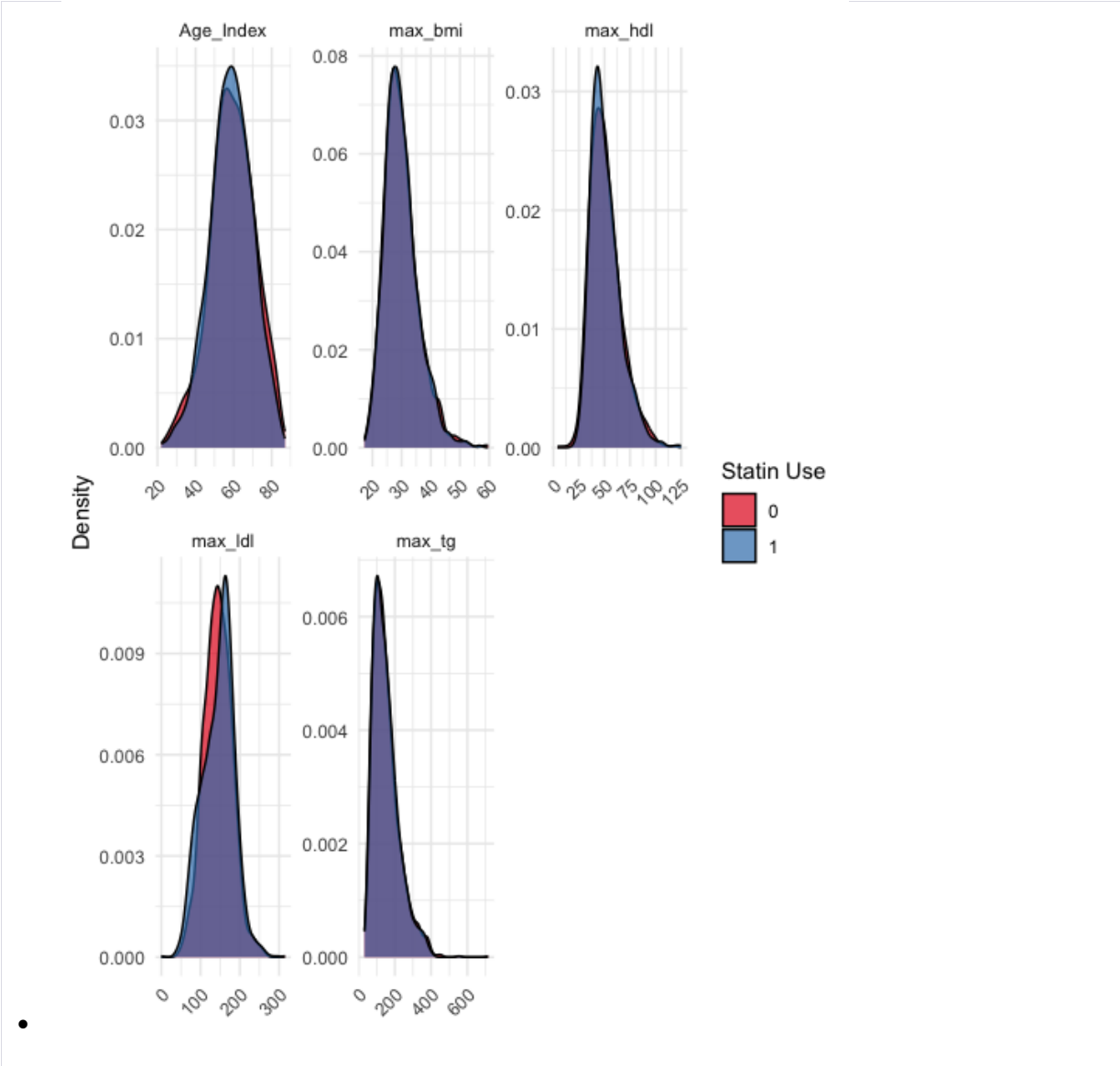
Table 2 shows the T-test values for the variables.

| Variable | t_value | p_value |
|---|---|---|
| Age Index | 1.56 | 0.12 |
| Maximum bmi | 0.11 | 0.92 |
| Maximum ldl | -0.61 | 0.54 |
| Maximum hdl | 0.05 | 0.96 |
| Maximum triglyceride | 0.24 | 0.81 |

The p-values for all the covariates are well above the common alpha level of 0.05, indicating no statistically significant differences in the means of these covariates between the treatment and control groups after matching. This lack of significant differences suggests that the propensity score matching process has effectively balanced these covariates between the groups, reducing

the likelihood that observed differences in outcomes are due to these confounders rather than the treatment effect.

Figure 1 shows the graphical representation of the PSM matching for continuous variables.




Measuring the Effect of Exposure

The effect of statin use is measured using an outcome model. Statin use's impact on the outcome of interest was investigated using a Cox proportional hazards model. Recognizing the potential confounding influence of demographics like Age at Index date, sex, and race, these factors were included as key predictors. Maximum BMI, reflecting overall health and obesity, along with smoker status, were incorporated due to their known associations with health outcomes.

Importantly, essential lipid profile elements like maximum LDL, maximum HDL, and maximum triglyceride were included. This comprehensive model structure aimed to adjust for potential confounding effects, isolating the distinct impact of statin use from other influencing factors. By effectively controlling for these covariates, the study's robustness was strengthened, leading to more reliable and valid conclusions about the true effect of statin use within the limitations of observational data.

**RESULTS**

Table 3. Summary Table Cox Model

| | coef | exp(coef) | se(coef) | z | Pr(>\|z\|) |
|---|---|---|---|---|---|
| statin_use | -0.311 | 0.733 | 0.099 | -3.146 | 0.001655** |
| Age_Index | 0.005 | 1.005 | 0.005 | 1.008 | 0.314 |
| sexM | -0.013 | 0.987 | 0.108 | -0.117 | 0.907 |
| raceAsian | 0.549 | 1.732 | 1.075 | 0.511 | 0.610 |
| raceBlack or African American | 0.787 | 2.196 | 1.034 | 0.761 | 0.447 |
| raceNative Hawaiian or Other PacificIslander | 0.910 | 2.485 | 1.421 | 0.641 | 0.522 |
| raceWhite | 0.537 | 1.711 | 1.005 | 0.534 | 0.593 |
| max_bmi | 0.045 | 1.046 | 0.008 | 5.585 | 0.0000000234 *** |
| smoker_statusNot Asked | -0.114 | 0.892 | 0.508 | -0.224 | 0.823 |
| smoker_statusPassive | -0.727 | 0.483 | 1.004 | -0.724 | 0.469 |
| smoker_statusQuit | 0.231 | 1.260 | 0.119 | 1.942 | 0.052 |
| smoker_statusUnknown | 1.272 | 3.569 | 0.456 | 2.791 | 0.005262** |
| smoker_statusYes | 0.310 | 1.364 | 0.125 | 2.480 | 0.01314* |
| max_ldl | -0.001 | 0.999 | 0.001 | -1.052 | 0.293 |
| max_hdl | -0.003 | 0.997 | 0.004 | -0.737 | 0.461 |
| max_tg | 0.002 | 1.002 | 0.001 | 3.404 | 0.000665*** |

Table 4.

| | exp(coef) | exp(-coef) | lower .95 u | upper .95 |
|---|---|---|---|---|
| statin_use | 0.733 | 1.365 | 0.604 | 0.889 |
| Age_Index | 1.005 | 0.995 | 0.996 | 1.014 |
| sexM | 0.988 | 1.013 | 0.800 | 1.219 |
| raceAsian | 1.732 | 0.577 | 0.210 | 14.256 |
| raceBlack or African American | 2.196 | 0.455 | 0.290 | 16.662 |
| raceNative Hawaiian or Other Pacific Islander | 2.485 | 0.402 | 0.153 | 40.287 |

| | | | | |
|---|---|---|---|---|
| raceWhite | 1.711 | 0.584 | 0.239 | 12.272 |
| max_bmi | 1.046 | 0.956 | 1.030 | 1.063 |
| smoker_statusNot Asked | 0.892 | 1.121 | 0.330 | 2.414 |
| smoker_statusPassive | 0.483 | 2.069 | 0.068 | 3.459 |
| smoker_statusQuit | 1.260 | 0.794 | 0.998 | 1.590 |
| smoker_statusUnknown | 3.569 | 0.280 | 1.460 | 8.724 |
| smoker_statusYes | 1.364 | 0.733 | 1.067 | 1.742 |
| max_ldl | 0.999 | 1.001 | 0.996 | 1.001 |
| max_hdl | 0.997 | 1.003 | 0.989 | 1.005 |
| max_tg | 1.002 | 0.998 | 1.001 | 1.004 |

Table 3 and Table 4 summarize the Cox model for the effect of statin use on diabetes, which showed a statistically significant association with the event. The hazard ratio for statin use was 0.7327314, indicating that the risk of the event occurring was about 27% lower in the statin-using group than the non-users. This finding was underscored by a p-value of 0.001655, reinforcing the robustness of this association.

Apart from statin use, the model included demographic factors like Age Index, sex, and various racial categories. However, none of these demographic variables demonstrated a statistically significant impact on the time to event, as suggested by their p-values above the threshold of 0.05. This lack of significance implies that these demographic aspects did not independently influence the risk of the event.

The model also assessed health and lifestyle variables, notably max BMI and smoker status. The max BMI variable was particularly noteworthy, with its significant effect ($p < 0.001$) and a hazard ratio of 1.0464116, subtly indicating an increased hazard with rising BMI levels. Additionally, the smoking status 'Yes' was found to significantly affect the risk, highlighting the critical role of smoking behavior in altering health outcomes. The lipid profile variables, max triglycerides (max_tg) stood out with a significant effect ($p = 0.000665$), albeit with a hazard ratio close to 1, suggesting only a slight increase in risk with higher triglyceride levels.

The overall model fit and significance, as indicated by a concordance index of 0.617 and extremely low p-values in the likelihood ratio test, Wald test, and Score (log-rank) test, affirm the model's robustness and the significance of its findings.

In summary, the results of this Cox proportional hazards model revealed statin use as a key protective factor against the risk of the event(diabetes incidence). The model's findings also emphasized the importance of considering a range of factors, such as BMI and smoking status, in comprehensively assessing health risks.

Appendix

Figure 2