Statins and Risk of Cancer
A Retrospective Cohort Analysis of 45,857 Matched Pairs From an Electronic Medical Records Database of 11 Million Adult Americans

Claudio Marelli, MD,* Candace Gunnarsson, EdD,* Susan Ross, MD,* Sara Haas, MS,* Donna F. Stroup, PhD, MSc,* Paul Clopton, PhD,† Paul Clopton, MS,‡ Anthony N. DeMaria, MD§

Cincinnati, Ohio; Chalfont St Giles, Bucks, United Kingdom; and San Diego, California

Objectives
The purpose of this study was to determine whether cancer can be attributed to statin use among a general population of older adults in the United States with at least 3 years of follow-up.

Background
Statins are widely prescribed drugs in the United States for the management of dyslipidemia, atherosclerosis, and cardiovascular event risk reduction. Unsettled scientific debate about the association of statins with cancer continues, with high-profile studies showing conflicting results.

Methods
A retrospective cohort analysis of the incidence of cancer in older adults who have and who have not used statins was performed. More than 11 million analyzable patient records from January 1990 through February 2009 were drawn from the General Electric Centricity electronic medical records database. Propensity matching found pairs of patients receiving and not receiving statin therapy who shared similar propensities for statin use.

Results
Propensity score methods matched 45,857 comparison pairs of patients taking a statin and patients not taking a statin. The average time in the database was 8 years, with pairs being followed for an average of 4.6 and 4.7 years. After matching, the incidence of cancer in patients taking a statin was 11.37% compared with 11.11% in matched patients not taking a statin. Multivariate-matched Cox regression analysis showed a nonsignificant hazard ratio of 1.04 (95% confidence interval: 0.99 to 1.09). Kaplan-Meier curves for diagnosis of any cancer up to 10 years also showed no difference for patients taking a statin and those not taking a statin.

Conclusions
This retrospective analysis of nearly 46,000 propensity-matched pairs demonstrated no statistically significant increased risk of cancer associated with statins. (J Am Coll Cardiol 2011;58:530–7) © 2011 by the American College of Cardiology Foundation

Statins (or 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors) are among the most prescribed drugs in the United States for the management of dyslipidemia, atherosclerosis, and cardiovascular event risk reduction. Recent published results of the JUPITER (Justification for the Use of Statins in Primary Prevention: An Intervention Trial Evaluating Rosuvastatin) (1) has further fueled the debate about the value of expanding indications for statin therapy to lower risk individuals to reduce their relative risk of cardiovascular disease (2). Available analyses (3,4) assume that statin therapy is well tolerated and safe, even with long-term use, but the evidence acquired so far is limited by several factors: the number and characteristics of the subjects studied, relatively short follow-up period, and constraints of patient selection and rigid treatment protocols typical of randomized, controlled trials.

For more than a decade, the scientific debate about the association of statins with cancer has been unsettled (5–18). Recently, this debate was rekindled with the high-profile publication of 2 studies of a new lipid-lowering agent, ezetimibe, in combination with a statin. The first study, the SEAS (Simvastatin and Ezetimibe in Aortic Stenosis) trial (19), a randomized, placebo-controlled study of safety in 1,873 patients with a median follow-up of 52.2 months, reported that cancer occurred significantly more frequently in the treatment group (relative risk 95% confidence interval [CI]: 1.13 to 2.12). The other study (20), a hypothesis-testing meta-analysis of cancer data from more than 20,617 patients in 2 large ongoing trials of the same regimen: SHARP
Statins and Cancer Risk: A Retrospective Analysis

JACC Vol. 58, No. 5, 2011
July 26, 2011:530–7
Marelli et al.

(Study of Heart and Renal Protection) and IMPROVE-IT (Improved Reduction of Outcomes: Vytorin Efficacy International Trial), came to the opposite conclusion, namely, that these trials do not provide evidence of an association between lipid-lowering drugs and cancer. Meta-analyses of the 2 ongoing trials yielded a risk ratio of 0.96 (95% CI: 0.82 to 1.12), and of all 3 of the aforementioned trials, a risk ratio of 1.06 (95% CI: 0.92 to 1.22). It is extremely important to answer this question unambiguously because statins (with or without ezetimibe) are widely prescribed throughout the world. More definitive studies, particularly real-world observational studies involving more patient data with longer follow-up, are clearly needed.

Toward this end, we performed a retrospective cohort analysis of the incidence of cancer in older adults according to their use of statins. Analyzable patient records were drawn from the General Electric (GE) Centricity database, a large commercial electronic medical records (EMR) system, used by more than 20,000 clinicians for the management of medical records of more than 30 million patients throughout the United States (21). This nationally representative database of anonymized longitudinal medical records contains many clinical and demographic variables and allows patients’ clinical conditions to be tracked over time.

Propensity score methods were used to develop matched comparison groups of patients taking a statin and those not taking a statin. This report presents our findings regarding the incidence of cancer in these matched groups.

Methods

The primary objective of the study was to determine whether cancer can be attributed to the use of statins among a general population of older U.S. adults with at least 3 years of follow-up. Study design and patients. We analyzed records from the GE Centricity EMR database. The a priori protocol and statistical analysis plan were reviewed by a central institutional review board, and an informed consent exemption was obtained. Patient records had to contain a valid patient age and documented activity dates between January 1, 1990, and February 28, 2009, to be included in the analyzable dataset.

For patients in the database up to February 2009, it was determined whether they had ever received statin therapy. Time zero was defined as the point at which patients began to take statins or, if never on statins, the date of the first recorded low-density lipoprotein (LDL) or total cholesterol level in the database.

For those with sufficient medication history, men 45 years of age and older and women 55 years of age and older at time zero were selected, with the age cutoff based on the SHAPE (Screening for Heart Attack Prevention and Education) Task Force report (22). Patients were excluded if they had insufficient history in the database, defined for statin users as <1 year before the start of medication or <2 years after the start date and for those not taking a statin, <3 years of history in the database, younger than the age cutoffs at the start of medication or at time zero, and any patients with a diagnosis of cancer before medication or time zero. The remaining patients were treated as a cohort for analysis.

Outcomes. The primary outcome was diagnosis of any type of cancer recorded in the medical record after time zero.

Calculation of propensity scores. Propensity scores for receipt of statins were calculated for each of the patients included in the analysis based on a nonparsimonious multivariable logistic regression model (23). Due to the poor reporting of body mass index (BMI) and very low density lipoprotein (VLDL), we imputed values for these measurements before time zero based on age- and sex-specific means. There were matches on 35 characteristics (see Online Table 1 for complete listings of all variables). Characteristics hypothesized to be associated with statin use included smoking status, age, time in the database, calendar time, sex, race, concomitant diagnoses, medications taken before time zero or statin initiation, total cholesterol, VLDL, low-density lipoprotein (LDL), triglycerides, the number of office visits before time zero or statin initiation, record of Pap or prostate-specific antigen test, and BMI.

Propensity score matching. The goal of the propensity matching was to find pairs of patients taking and not taking a statin who shared similar propensities for statin use based on the matching variables. A SAS version 9.1 (SAS Institute, Cary, North Carolina) macro was created to use nearest-neighbor matching on the estimated propensity scores to choose matches for the statin patients (24). We randomly ordered patients taking and those not taking a statin and then chose the patients not taking a statin with the propensity score closest to the first patient taking a statin. We matched the propensity scores to 3 decimal places. Both patients were then removed from the pool of patients available for matching. This procedure was repeated for each patient taking a statin (25). We chose a stringent (conservative) stopping rule to control as much as possible for confounding in these observational data.

Assessment of residual bias. Differences in the distribution of patient characteristics were assessed before and after matching. We used independent-sample tests for the descriptive statistics before matching and paired-sample tests after matching.

Statistical analysis. To assess the extent to which propensity matching reduced confounders, the distribution of several variables were compared before and after matching, taking into consideration sex, age, race, smoking status,
time in the database, concomitant diagnoses, preventive screenings (Pap or prostate-specific antigen test), concomitant medications, and calendar time. Unmatched group comparisons were made using t tests for independent samples and chi-square tests. Matched group comparisons were made using paired t tests and McNemar tests. Conditional Cox regression was used to account for the matched pairs to estimate the association of statin use with cancer. In the multivariable Cox regression, there was adjustment for propensity score, as well as other covariates, to control for any residual confounding; however, the propensity score was nonsignificant in the final model (Online Table 2). Covariates included in the Cox regression model were all variables included in the propensity model (including flags for imputed BMI and VLDL); concomitant diagnoses (e.g., chronic obstructive pulmonary disease) before time zero; use of various medications (tumor necrosis factor alpha inhibitors, immunosuppressants, glucocorticoids, omega-3, acetylsalicylic acid, hormones, or nonsteroidal anti-inflammatory drugs) before time zero; and metabolic measurements (LDL, VLDL, total cholesterol, triglycerides, BMI). In addition, the definition of a diagnosis of a condition before time zero was adjusted to include persons on condition-specific medication for diabetes, hypertension, peripheral vascular disease, and ischemic heart disease. For example, the use of sulfonlureas or insulin was used as a surrogate for a diagnosis of diabetes mellitus. Interactions were formally tested by entering interaction terms between statin use and the main predictor variables, adjusting for propensity to use statins. Only 2 interaction terms were statistically significant: acetylsalicylic acid–statins and hormones–statins. However, these 2 interaction terms did not substantially alter the findings; therefore, no interaction terms were included in the final model. All data analyses were performed using SAS version 9.1 (proc logistic for generating propensity scores, proc lifetest for generating Kaplan-Meier curves, proc SQL for generating plots, and proc freq, proc means, proc ttest, and proc sort for descriptive statistics and significance testing).

Results

Patient attrition and characteristics. As of February 2009, the database contained information on 11,196,881 patients, of whom 1,191,822 (10.6%) had ever used a statin. After exclusion of patients with insufficient history, age outside the target range, or diagnosis of cancer before time zero, 203,763 patients on medication and 159,004 patients never on medication remained for analysis (Fig. 1). The use of these variables for exclusion was required to eliminate possible confounding influences and adhere to age guidelines. Patients taking a statin differed from those not taking a statin for most characteristics before matching (Table 1). Specifically, before matching, statin users were older with higher comorbidity burden and concomitant medication use compared with those not taking a statin. After matching, all baseline covariates were well balanced between the 2 groups, as shown by nearly identical frequencies in those patients taking a statin and those not taking a statin (Table 1). Matched patients had an average time in the database of approximately 8 years, with 85% >5 years. The follow-up interval after time zero for patients taking a statin was 4.7 ± 2.2 years and 4.6 ± 2.2 years for matched patients not taking a statin.

Statin use and cancer. Before matching, cancer occurred in 23,906 of the 203,763 patients taking a statin (11.7%) and in 17,457 of the 159,004 patients (11.0%) not taking a statin (Table 2). After the 1:1 matching, however, there were 45,857 patients in each group, and the incidence of cancer in patients taking a statin decreased to 11.37% compared with 11.11% in matched patients not taking a statin. No particular cancer type predominated (Online Table 3).

The multivariate–matched Cox regression analysis showed a nonsignificant hazard ratio of 1.04 (95% CI: 0.99 to 1.09) (Online Table 2). Similarly, Kaplan-Meier curves for diagnosis of any cancer up to 10 years showed no difference for patients taking a statin and those not taking a statin (data not shown). Interactions of statins with clinically relevant covariates were significant only for acetylsalicylic acid and hormones. As expected, collinearity was apparent for LDL and total cholesterol. This did not invalidate any results because the objective was to assess the overall association between statin use and cancer and not to interpret the coefficients of the covariates.

Discussion

This retrospective database analysis of a general population of 45,857 matched pairs of patients with an average of 4.6 years of follow-up after time zero revealed no statistically significant association of statins with cancer. Results are consistent with the fact that a plausible biological mechanism to implicate statins as causative agents in cancer has never been demonstrated. In fact, numerous preclinical studies have supported the potential anticancer activity of these compounds, providing evidence of their antiproliferative, proapoptotic, anti-invasive, and radiosensitizing properties (26–28). In this analysis, the distributions of cancer types seen in both groups were comparable. Cancer is a multifactorial disease, and if statins were truly associated with an increase in cancer, we would have expected to see an increase in at least 1 type of cancer, as is typically seen with toxin-related neoplasms.

Most importantly, there are ample clinical trial data in the literature to refute a statin association with cancer. In 2005, the Cholesterol Treatment Trialists published a meta-analysis of 14 randomized trials of statins involving more than 90,000 patients, with a pooled risk ratio for cancer after randomization of 1.0 (95% CI: 0.95 to 1.04) (29). More recently, Alsheikh-Ali et al. (8) reported a Bayesian meta-analysis of 15 randomized trials of statins with more than
437,000 patients. Adjusting for LDL levels, they reported an incidence rate ratio of 1.02 (95% CI: 0.95 to 1.10), and concluded that statins have no effect on cancer risk at all LDL levels. Furthermore, in a recent synthesis of multiple studies using real-world data, Taylor et al. (30) reported a meta-analysis of 20 case-control trials of more than 100,000 cancer cases and more than 3.4 million controls. They computed an odds ratio of 0.71 (95% CI: 0.56 to 0.89) for any cancer, suggesting a protective effect of statins. Clearly, the weight of clinical evidence to date supports these findings showing no significant association between statin use and incidence of cancer.

Historically, the use of observational data to help answer clinical questions has been limited by the available datasets, chiefly administrative claims data, which are better suited for use and costs studies. Yet with the increasing uptake of EMRs by healthcare providers, and their resulting databases of large numbers of patients with clinically rich, longitudinal data, interest is greater than ever in using these datasets to help answer clinical questions. Witness the U.S. Food and

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>All 'Cleaned' patients in February 2009 database (DB)</td>
<td>N=11,196,881</td>
</tr>
<tr>
<td>All patients NEVER on a Statin</td>
<td>N=10,005,059</td>
</tr>
<tr>
<td>Patients on medication with sufficient history in DB*</td>
<td>N=278,984</td>
</tr>
<tr>
<td>Patients on medication started too young</td>
<td>N=59,706</td>
</tr>
<tr>
<td>Patients on medication, males over 45, females over 55 when started med</td>
<td>N=219,188</td>
</tr>
<tr>
<td>Patients with Diagnosis (Dx) of cancer prior to medication start date</td>
<td>N=15,425</td>
</tr>
<tr>
<td>Medication sample for use in study</td>
<td>N=203,763</td>
</tr>
<tr>
<td>Patients NOT Matched</td>
<td>N=157,906</td>
</tr>
<tr>
<td>Final Sample Size - Medication Users - AFTER MATCHING</td>
<td>N=45,857</td>
</tr>
<tr>
<td>Final sample size - NEVER on medication - AFTER MATCHING</td>
<td>N=45,857</td>
</tr>
<tr>
<td>Patients NEVER on medication with insufficient history in DB†</td>
<td>N=9,609,910</td>
</tr>
<tr>
<td>Patients NEVER on medication with sufficient history in DB†</td>
<td>N=395,149</td>
</tr>
<tr>
<td>Patients too young as of TIMEZERO (date of first recorded total cholesterol)</td>
<td>N=224,183</td>
</tr>
<tr>
<td>Patients with Dx of cancer prior to TIMEZERO</td>
<td>N=11,962</td>
</tr>
<tr>
<td>Patients NEVER on medication, males over 45, females over 55 at TIMEZERO</td>
<td>N=170,966</td>
</tr>
<tr>
<td>Patients NEVER on medication with insufficient history in DB†</td>
<td>N=9,609,910</td>
</tr>
<tr>
<td>Patients NEVER on medication with sufficient history in DB†</td>
<td>N=395,149</td>
</tr>
</tbody>
</table>

*Patients on medications (med) must have 1 year prior to medication start date AND 2 years after start date. †Patients NEVER on medication must have at LEAST 3 years of history in the database to be included in the study. ‡Males must be over 45 years of age at start of medication, Females must be over 55 years of age at start of medication.
Drug Administration’s new Sentinel Initiative whereby it hopes to actively query diverse automated healthcare data holders, including EMRs, to evaluate possible medical product safety issues quickly and securely (31). The GE Centricity database used in this study is perfectly suited for safety analyses such as these because it provides a very large, current, and representative sample of ambulatory patients in the United States.

Statistical methodologies for analysis of these large EMR databases are still evolving (32–34). Propensity scoring

Continued on next page
methods are increasingly used to provide comparison of groups that are balanced with regard to important determinants of exposure to the intervention of interest. Our matching process created pairs with propensity scores within 4 decimal places and has thus provided extremely well-matched comparison groups. Although residual confounding cannot be excluded, we are fairly confident that our cancer results are not being obscured by unmeasured confounders.

New EMR technologies are enabling powerful statistical evaluations of what would be considered huge databases from the perspective of the clinical trialist. Indeed, it is from the impressively large numbers of the GE Centricity EMR database that the strength of the current study’s findings emerges. Nevertheless, we recognize 3 potential limitations of this approach.

**Study limitations.** First, propensity scoring methods may not completely eliminate bias that is due to unmeasured or hidden covariates (35). There is the possibility that the comparison groups may differ as an artifact of the method used to assign patients to treatment groups. Specifically, control group patients could never be prescribed statins for the duration of their observation interval, but statin group patients could be assigned to the statin group on the basis of a single statin prescription decision, regardless of their subsequent medication adherence or duration of use. Further, the propensity match itself results in further winnowing of the data from 159,004 potential pairs to 45,857 analyzable pairs. This stringent restriction was necessary because any result based on the full dataset of 159,004 pairs could easily be attributed to confounding. We imputed values for 2 variables (BMI and VLDL) before time zero to increase the sample size available for the match; however, we adopted a conservative approach by including flags for these in our Cox regression.

Second, in fitting the model for propensity matching, we were forced to omit data for records missing those variables. We used variables for elimination that were clearly associated with confounding (age outside the target range, insufficient history, or pre-existing diagnosis of cancer). Although this requirement resulted in substantial winnowing of the data (retention of 17.1% of patients on medication and 1.6% of patients never on medication), this restriction of data is necessary to control for confounding in observational data. Recognizing that the requirement for complete data may create some bias in the results, the distribution of known variables for the group taking a statin and the group not taking a statin were analyzed and found to be similar.

Third, these results may not be generalizable to the overall population of U.S. adults taking statins. Although the population of adults in the GE database is representative of the U.S. population, the representativeness of the matched cohorts is restricted due to the database winnowing inherent in the analysis. The results of the analysis may be

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Continued</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strata according to time zero, calendar time</td>
<td>On Medication</td>
</tr>
<tr>
<td>Before January 1, 2000</td>
<td>Before Matching</td>
</tr>
<tr>
<td>January 1, 2000 to December 31, 2001</td>
<td>17.148 (8.42)</td>
</tr>
<tr>
<td>January 1, 2002 to December 31, 2003</td>
<td>26.807 (13.16)</td>
</tr>
<tr>
<td>January 1, 2004 to December 31, 2005</td>
<td>47.629 (23.37)</td>
</tr>
<tr>
<td>January 1, 2006 to December 31, 2007</td>
<td>72.504 (35.58)</td>
</tr>
</tbody>
</table>

No. of imputations per measure

- Imputed BMI: 20.914 (10.26)
- Imputed VLDL: 173.191 (85.0)
- BMI: 30.04 ± 5.88
- LDL: 132.87 ± 36.59
- VLDL: 180.37 ± 78.24
- Triglycerides: 179.74 ± 78.24
- Total cholesterol: 215.49 ± 42.39

No. of office visits before time zero

- Before Matching: 8.59 (13.03)
- After Matching: 6.18 (10.71)

Values shown are n, n (%), or mean ± SD. *p values for the UNMATCHED data were done using t-tests for independent samples or chi-square, but for the MATCHED data paired t-tests and McNemar tests were used. Unmatched tests were used for comparisons before matching and matched tests were used for comparisons after matching.

ASA = acetylsalicylic acid; BMI = body mass index; COPD = chronic obstructive pulmonary disease; COX = cyclooxygenase; LDL = low-density lipoprotein; NSAID = nonsteroidal anti-inflammatory drug; PSA = prostate-specific antigen; TNF = tumor necrosis factor; VLDL = very low density lipoprotein.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Cancer and Statin Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>203,763</td>
</tr>
<tr>
<td>Cancer after time zero, n (%)</td>
<td>23,906 (11.73)</td>
</tr>
</tbody>
</table>
REFERENCES


Conclusions

This EMR database analysis of more than 91,000 U.S. adults with complete clinical datasets who were propensity matched with an average of 4.6 years of follow-up, demonstrated no statistically significant increased risk of cancer associated with statins.

Acknowledgments

The authors are indebted to the dozens of physicians and thousands of patients who contributed EMR data to the analyzable database for research purposes. The authors also acknowledge the members of the New England Institutional Review Board (NEIRB), Newton, Massachusetts, for reviewing our protocol and granting exemption based on their determination that the research involving human subjects is exempt per the federal government guidelines for research using existing data. And finally, the authors are grateful for the statistical review and advice of David Kleinbaum, Ph.D, at Emory University, Atlanta, Georgia.


Key Words: cancer • drug • risk • safety • statins.

APPENDIX

For supplemental tables, please see the online version of the article.