Use of an electronic health record to identify prevalent and incident cardiovascular disease in type 2 diabetes according to treatment strategy

Mary T Korytkowski,1 Esra Karslioglu French,2 Maria Brooks,3 Dilhari DeAlmeida,4 Justin Kanter,5 Manuel Lombardero,6 Vasudev Magaji,7 Trevor Orchard,3 Linda Siminerio8

ABSTRACT

Background: The increasing use of electronic health records (EHRs) in clinical practice offers the potential to investigate cardiovascular outcomes over time in patients with type 2 diabetes (T2D).

Objective: To develop a methodology for identifying prevalent and incident cardiovascular disease (CVD) in patients with T2D who are candidates for therapeutic intensification of glucose-lowering therapy.

Methods: Patients with glycated hemoglobin (HbA1c) ≥7% (53 mmol/mol) while receiving 1–2 oral diabetes medications (ODMs) were identified from an EHR (2005–2011) and grouped according to intensification with insulin (INS) (n=372), a different class of ODM (n=833), a glucagon-like peptide 1 agonist (GLP-1RA) (n=59), or no additional therapy (NAT) (n=2017). Baseline prevalence of CVD was defined by documented International Classification of Diseases Ninth Edition (ICD-9) codes for coronary artery disease, cerebrovascular disease, or other CVD with first HbA1c ≥7% (53 mmol/mol). Incident CVD was defined as a new ICD-9 code different from existing codes over 4 years of follow-up. ICD-9 codes were validated by a chart review in a subset of patients.

Results: Sensitivity of ICD-9 codes for CVD ranged from 0.83 to 0.89 and specificity from 0.90 to 0.96. Baseline prevalence (INS vs ODM vs GLP-1RA vs NAT: 65% vs 39% vs 54% vs 59%, p<0.001) and incident CVD (Kaplan-Meier estimates: 58%, 31%, 52%, and 54%, p=0.002) were greater in INS group after controlling for differences in baseline HbA1c (3.2 ±2.0% vs 8.3 ±1.2% vs 8.2 ±1.3% vs 7.7 ±1.1% (77 vs 67 vs 66 vs 61 mmol/mol), p<0.001) and creatinine (1.15 ±0.96 vs 1.10 ±0.36 vs 1.01 ±0.35 vs 1.07 ±0.45 mg/dL, p=0.001).

Conclusions: An EHR can be an effective method for identifying prevalent and incident CVD in patients with T2D.

INTRODUCTION

Recommendations for intensification of glucose-lowering therapies for people with type 2 diabetes (T2D) have evolved over recent years from more of a prescriptive approach to one that is individualized and patient oriented.1–5 The American Diabetes Association recommends metformin as the drug of choice for initial pharmacological therapy in patients not meeting glycemic goals with lifestyle interventions alone with T2D.5 However, the majority of patients with T2D will not achieve and maintain glycemic goals with any single drug therapy. There are multiple studies demonstrating the effect of different combinations of pharmacological agents on glycemic control, but there is little available data demonstrating superiority of one drug combination over another on long-term outcomes.

The majority of clinical trials investigating pharmacological agents for achievement of glycemic control in T2D-specified changes in risk for cardiovascular disease (CVD) as the primary measure of clinical outcomes.6–12 In 2008, the Food and Drug Administration (FDA) issued an updated Guidance for Industry requiring that all new antidiabetic drugs rule out excess cardiovascular risk
prior to seeking approval for clinical use.13 CVD end points and mortality are also used as a critical factor in ongoing surveillance of existing glucose-lowering therapies.12–15

Results from large studies prespecified glycemic goals demonstrated no reductions in CVD events.5–9 16–17 Until recently, similar negative findings have been demonstrated in studies investigating CVD outcomes according to use of specific glucose lowering medications, including glitazones, or the dipeptidyl peptidase IV inhibitors (DPP-IV-I) (sitagliptin, saxagliptin, or alogliptin).8 10–12 17 18 In fact, the DPP-IV-I, saxagliptin, was associated with an increase in risk of hospitalization for congestive heart failure (CHF) and hypoglycemia.11 Whether this was due to the study medication or an epiphenomenon of the study design is not clear. The sodium-glucose cotransporter 2 inhibitor, empagliflozin, was recently reported to reduce the number of hospitalizations for CHF and all cause and cardiovascular mortality.17 There were no significant reductions in risk for myocardial infarction or stroke. To date, empagliflozin is the only agent in this class that has demonstrated reductions in CVD. However, more data regarding the long-term safety and efficacy of this group of agents are needed.

Randomized controlled clinical trials (RCTs), long considered the gold standard for evaluating the safety and efficacy of new therapies, require considerable resources to prospectively study the long-term effects of any diabetes medication on CVD or other outcomes.5 13 15 19 The electronic health record (EHR) offers an alternative or supplemental approach to the RCTs for detecting signals for CVD risk and other adverse events in large patient populations that may be representative of patient populations with T2D who will be taking a medication on a chronic basis.20–24

The primary aim of this study was to examine the ability of an EHR to accurately identify pre-existing and incident CVD events using the ICD-9 codes in a cohort of patients with T2D previously identified in another study investigating the ability of an EHR to identify patterns of therapeutic intensification of glucose-lowering therapies in patients with T2D.25

### METHODS

This study was approved by the Institutional Review Board at the University of Pittsburgh. The University of Pittsburgh Medical Center (UPMC) EHR data repository includes administrative and clinical data forwarded from the health system’s clinical, administrative, and financial databases.26 27 There are currently more than 20 academic, community, and specialty hospitals; more than 500 physician offices; and over 3600 physicians in the UPMC Health System. The EMR includes patient demographics, office visits, medication lists, laboratory results, and charges from inpatient and outpatient settings throughout the health system (EpicCare; Epic Systems Corp., Verona, Wisconsin, USA; PowerChart; Cerner Corporation, Kansas City, Missouri, USA). An interface (dbMotion, Pittsburgh, Pennsylvania, USA) connects and shares key information between systems.

The criteria used to identify patients as having T2D in this study was previously validated, with a positive predictive value (PPV) of 96% and a sensitivity of 96%.26 Data extraction methods for this study were previously described.25 Briefly, ambulatory de-identified EHR information from a large institutional data repository was searched to identify any patient visit that listed an oral diabetes medication (ODM) as an active therapy between June 2005 and November 2011. Patients with glycated hemoglobin (HbA1c) ≥7% (53 mmol/mol) (index HbA1c value) who had existing prescriptions for one or two ODM, documentation of at least one follow-up visit after the index HbA1c, and available ICD-9 codes (n=3331) were included in this study (see online supplementary appendix A). ICD-9 codes were used to identify the presence of CVD according to one of three distinct categories: coronary artery disease (CAD) (ICD-9 410.xx, 411.xx, 412.00, 413.x, 414.x, and 414.xx); cerebrovascular disease (CBVD) (431, 433.xx, 434.xx, 435.x, 434.xx, 437.x, and 438.xx); and other CVD (402.xx, 416.x, 424.x, 425.x, 427.x and 427.xx, 428.x and 428.xx, 429.x, 440.x and 440.xx, 443.9 and 443.81, 707.xx, 785.4, and V49.6 and V49.7) (table 1). Other CVD referred to CVD outcomes that were not captured as CAD or CBVD.

Prevalent CVD was defined by the presence of an ICD-9 code at time of the index HbA1c value. Incident CVD was defined as a new ICD-9 code that appeared following the index date and that differed from any existing codes in patients with baseline evidence of CVD. For patients without baseline evidence of CVD, incident events were described by time of appearance in the EHR following the index date. Prevalent and incident CVD events were examined separately for all patients with or without baseline evidence of a CVD event, and according to therapeutic intensification with any insulin (INS) therapy, a glucagon-like peptide-1 receptor agent

### Table 1

<table>
<thead>
<tr>
<th>Complication</th>
<th>ICD-9 code(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary artery</td>
<td>410.xx, 411.xx, 412, 413.xx, and disease 414.xx</td>
</tr>
<tr>
<td>Cerebrovascular</td>
<td>431, 433.xx, 434.xx, 435.xx, 342. disease xx, 437.xx, and 438.xx</td>
</tr>
<tr>
<td>Other cardiovascular</td>
<td>402.xx, 416.xx, 424.xx, 425.xx, 427.xx, 428.xx, 429.x, 440.x, 443.9x, 443.81, 707.xx, 785.4x, V49.6, and V49.7</td>
</tr>
</tbody>
</table>

GLP-1RA (exenatide or liraglutide), a new class of ODM or no additional therapy (NAT).\(^\text{25}\)

Accuracy of ICD-9 coding for determination of CVD events was investigated by conducting a chart review on all participants receiving intensification with GLP-1RA therapies (n=59) and a subset of patients selected randomly from each of the other groups (n=205) (total 264 participants, 8% of included population). The chart review was performed by two endocrinologists (EKF and MTK) with extensive experience in diabetes care and use of an EHR, and a project coordinator (JK). The chart review process was initiated by a review of the office note or diagnostic study correlating with the date an ICD-9 code appeared in the EHR.

An ICD-9 code was defined as being accurate when there was chart documentation of the complication (true positive). An ICD-9 code was defined as being inaccurate in the absence of chart validation of an ICD-9 code (false positive). A false negative was defined as evidence of a complication by the chart review without a recorded ICD-9 code and a true negative as the absence of an ICD-9 code when a complication was not documented in the patient record.

**Statistics**

The baseline characteristics of patient population are presented as means and SDs for continuous variables and as percentages for categorical variables. Time-to-event analyses were undertaken using the index date (ie, intensification or the first recorded HbA1c value \(\geq 7\%\) (53 mmol/mol)) as time zero, and time was measured from the index date. Kaplan-Meier estimates and log rank statistics were used to compare the risk of cardiovascular events over time. Hazard ratios (HRs) and 95% CIs are reported. For the validation analyses, sensitivity, specificity and positive predicted values were computed and reported for each of the events. Following confirmation of the approximate normality of the continuous variables, analysis of variance statistics were used to compare the mean values among the three intensification groups and among the four groups including the control group. Categorical variables were compared using \(\chi^2\) statistics. Unadjusted and multivariable-adjusted (age, sex, race, body mass index (BMI), HbA1c, creatinine, and baseline CVD) Cox proportional hazard regression models were used to estimate the association between the four groups and risk of cardiovascular events.

**RESULTS**

The baseline clinical characteristics of the 3331 patients meeting inclusion criteria grouped according to intensification strategy are presented in table 2. The majority (58\%, n=1937) of patients had baseline evidence of CVD using ICD-9 criteria.

The chart review revealed that ICD-9 codes were highly sensitive and specific for each CVD complication (table 3). The positive predictive value of an ICD-9 code ranged from 0.77 for CBVD to 0.89 for other CVD. The diagnoses most frequently represented in the other CVD category were hypertensive heart disease (ICD-9 402 group) (28\%), cardiac dysrhythmias (ICD-9 427 group) (18\%), heart failure (ICD-9 428 group) (17\%), and other peripheral vascular disease (ICD-9 443 group) (12\%). No group differences were observed in the incidence of CAD (0.22 vs 0.10 vs 0.24 vs 0.19, p=0.36).

### Table 2  Baseline clinical characteristics and prevalence CVD according to intensification strategy

<table>
<thead>
<tr>
<th>Groups with DM therapy intensified</th>
<th>Insulin</th>
<th>GLP-1RA</th>
<th>+1 ODM</th>
<th>p value</th>
<th>Control</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>372</td>
<td>59</td>
<td>883</td>
<td>0.0006</td>
<td>1937</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (years)</td>
<td>60.8±14.2</td>
<td>56.9±11.7</td>
<td>62.8±12.8</td>
<td>0.0147</td>
<td>22.6</td>
<td>0.03</td>
</tr>
<tr>
<td>Percentage of African-American</td>
<td>24.7</td>
<td>8.6</td>
<td>20.5</td>
<td>0.35</td>
<td>55.4</td>
<td>0.33</td>
</tr>
<tr>
<td>Percentage of female</td>
<td>54.1</td>
<td>62.1</td>
<td>52.5</td>
<td>0.58</td>
<td>32.5±7.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>33.0±7.2</td>
<td>36.8±6.6</td>
<td>33.6±7.1</td>
<td>0.0011</td>
<td>7.6±1.1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>HbA1c (%)</td>
<td>9.2±1.9</td>
<td>8.3±1.2</td>
<td>8.2±1.3</td>
<td>&lt;0.0001</td>
<td>61</td>
<td>0.0008</td>
</tr>
<tr>
<td>Creatinine (mg/dL)</td>
<td>1.15±0.96</td>
<td>1.01±0.36</td>
<td>1.01±0.36</td>
<td>0.0008</td>
<td>1.07±0.45</td>
<td>0.0001</td>
</tr>
<tr>
<td>Prevalence of CVD at baseline</td>
<td>48.9</td>
<td>30.5</td>
<td>38.4</td>
<td>0.0006</td>
<td>42.0</td>
<td>0.002</td>
</tr>
<tr>
<td>Related to CAD*</td>
<td>16.9</td>
<td>3.4</td>
<td>12.1</td>
<td>0.0054</td>
<td>15.7</td>
<td>0.003</td>
</tr>
<tr>
<td>Related to CBVD*</td>
<td>54.3</td>
<td>28.8</td>
<td>38.3</td>
<td>&lt;0.0001</td>
<td>45.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Related to any CVD*</td>
<td>65.3</td>
<td>39.0</td>
<td>54.1</td>
<td>&lt;0.0001</td>
<td>59.1</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Cases refers to number of participants in each group with available data for each specified characteristic. Data are shown as mean±STD unless otherwise indicated.

*Presented as % of participants.

BMI, body mass index; CAD, coronary artery disease; CBVD, cerebrovascular disease; CVD, cardiovascular disease; GLP-1RA, glucagon-like peptide receptor 1 agonist; HbA1c, glycated hemoglobin; ODM, oral diabetes medications.
The observed baseline prevalence of CVD was highest prevalence in the INS group (n=372) and lowest in the GLP group (table 2). These differences among treatment groups were observed for CAD, CBVD, and other CVD. Patients intensified with INS had higher HbA1c and serum creatinine and more representation of African-American patients (table 2). Patients intensified with GLP-1RA (n=59) were younger and more obese than the other groups. The majority of patients (n=2017) did not receive any additional diabetes therapy during the time period of the study (NAT group). This group was older and leaner with a lower HbA1c than any of the intensified groups.

The cumulative probability of incident CVD also varied among intensification groups, with the highest incidence again observed in the INS group event over the 4-year period of follow-up (INS vs GLP-1RA vs ODM vs NAT: 0.58 vs 0.51 vs 0.52 vs 0.54, p=0.009) (figure 1). The other CVD accounted for the major portion of these differences (0.47 vs 0.21 vs 0.38 vs 0.40, p=0.008), with a smaller contribution for incident CBVD (0.21 vs 0.03 vs 0.15 vs 0.17, p=0.09).

Incident CVD was further examined according to the presence (n=1937) or absence (n=1394) of prevalent CVD at baseline. The cumulative probability of incident CVD over years 1–4 of follow-up in patients with baseline evidence of CVD was higher (0.50, 0.45, 0.59 and 0.68, respectively) than what was observed among patients without baseline evidence of CVD (n=1394) (0.09, 0.16, 0.22, and 0.30, respectively). Among patients with prevalent CVD at time of HbA1c ≥7% (53 mmol/mol), there were no differences for any category of incident CVD over 4 years according to the treatment group. Among patients without prevalent CVD at baseline, the cumulative probability of incident CVD was highest in the INS group only for the other CVD category (0.42 vs 0.00 vs 0.28, vs 0.26, p=0.02). Of note, there were no incident other CVD events in the GLP-1RA group.

Cox regression models were used to investigate the association between diabetes intensification therapy and incident CVD in relation to the NAT group (table 4). The unadjusted HR for any incident CVD event in the entire cohort of 3331 patients was lowest among those receiving intensification with GLP-1 RA (HR=0.44 (CI 0.24 to 0.82), p=0.01). This remained significant after adjusting for baseline differences in age, sex, race, BMI, HbA1c, creatinine, and baseline CVD (HR=0.55 (CI 0.280 to 0.994), p=0.048). When incident CVD was investigated according to those with and without prevalent CVD at baseline, no differences were observed among the groups for unadjusted and adjusted HR.

**DISCUSSION**

The primary purpose of this investigation was to investigate the ability of an EHR to accurately identify CVD events in a group of patients with T2D. This study was performed using data from an earlier study investigating prescribing patterns for patients with T2D and HbA1c values ≥7.0% (53 mmol/mol) while receiving treatment with one or two ODM. The results for prevalent and incident CVD events are reported according to treatment strategy employed among patients having available follow-up data from this earlier report. The results of this study are consistent with prior studies demonstrating a high prevalence and incidence of CVD in a patient population with T2D (table 2, figure 1). As anticipated, the incidence of CVD was higher among patients with baseline evidence of CVD, which was attributed predominantly to other CVD (ie, PVD and CHF).

The results demonstrating the cumulative probability of CVD events in this study are consistent with what was reported in the CALIBER study, which included over 34 000 individuals with T2D. The cumulative incidence of CVD was 58% in women and 67% in men over the 5 years of data collection in CALIBER, with PVD and CHF identified as the major contributors.

While the current study was not designed as a comparison of CVD outcomes according to intensification strategy, the observation that the INS group had more prevalent and incident CVD than other groups is also consistent with what has been observed in previous studies. The higher numbers in the INS group are consistent with what has been observed in previous studies.
likely a reflection of higher baseline comorbidity rather than attribution to INS therapy itself (table 2).31 In subgroup analyses from earlier studies using DPP-IV inhibitors, INS-treated participants with a longer duration of T2D were at higher risk for CVD events (HR 1.23 (CI 0.95 to 1.59)).10 11 Similar to what was observed in the INS group in this study, individuals with renal impairment were also at higher risk for these events.10

The potentially favorable CVD profile observed in the small number of participants in the GLP-1RA group needs to be interpreted cautiously owing to the very low number of participants in this group. While agents in this class have been described as having potential cardio-protective effects, there is currently no evidence from RCCTs in humans to allow conclusions to be drawn regarding their cardiovascular safety or efficacy.32 33 The small numbers of participants receiving intensification of therapy with GLP-1RA is of some interest in that this class was considered to be second or third tier therapy at the time of data collection in this study.12 These results differ from several earlier EHR investigations demonstrating intensification with GLP-1RA in approximately 7% of patients.34 35 The reasons for the infrequent use of these agents (3.6% of those receiving therapeutic intensification) are not known, but it is possible that this class of agents are being prescribed to participants who were not candidates for inclusion in this study, including those with HbA1c <7% (53 mmol/mol), who may be treated with >2 ODM, or who may already be receiving INS therapy.25

There are several limitations to this study, with the inability to obtain information from the EHR regarding duration of diabetes being one of the most important limitations.25 Prevalent and incident CVD can be related to diabetes duration and may have affected the observed differences among intensified groups. However, the risk for CVD events has been demonstrated to be present at the time T2D is diagnosed, indicating that risk factors are present well before the onset of overt hyperglycemia.35 36 Another limitation is that associated with the use of an EHR for documentation of CVD. Data obtained from an EHR have been demonstrated as being incomplete and prone to error.37 Not all study variables were available for each patient at each time point, as has been observed in other studies using EHRs to report clinical outcomes data.34 38 Another limitation of this study is that we do not have estimates of the numbers of patients in the database who sought care in an outside hospital system, which could lead to either under- or overestimates of CVD prevalence and incidence. The very small number of participants meeting criteria for this study who were prescribed GLP-1RA created disproportionate comparisons with the other groups. Finally, the use of ICD-9 codes has now been replaced by ICD-10 codes, which will require additional validation.39

In summary, we report that an EHR has the potential to provide important information regarding the prevalent and incident CVD in a described population of patients with T2D. These observations suggest that data mining with an EHR has the ability to complement, support or enhance information obtained from prospective randomized clinical trials.

| Table 4 Associations between therapeutic intensification therapy and incident cardiovascular disease in relation to the group receiving no additional therapy |
|----------------|----------------|----------------|----------------|----------------|
| Type of DM therapy | Risk for occurrence of any CVD | | | |
| | | | | |
| | Unadjusted | Adjusted | | |
| | RR 95% CI | p Value | RR 95% CI | p Value |
| Patients with and without baseline CVD | | | | |
| N=3331 | Insulin | 1.06 0.86 to 1.32 | 0.56 | 1.24 0.98 to 1.56 | 0.08 |
| | GLP-1RA | 0.44 0.24 to 0.82 | 0.01 | 0.53 0.28 to 0.99 | 0.048 |
| | ODM | 0.97 0.83 to 1.14 | 0.73 | 1.08 0.92 to 1.28 | 0.34 |
| Patients without baseline CVD | | | | |
| N=1394 | Insulin | 1.06 0.63 to 1.76 | 0.84 | 1.38 0.79 to 2.39 | 0.26 |
| | GLP-1RA | 0.25 0.06 to 1.03 | 0.055 | 0.30 0.07 to 1.24 | 0.096 |
| | ODM | 1.01 0.72 to 1.41 | 0.94 | 1.20 0.84 to 1.71 | 0.32 |
| Patient with baseline CVD | | | | |
| N=1937 | Insulin | 1.06 0.84 to 1.34 | 0.63 | 1.18 0.91 to 1.53 | 0.21 |
| | GLP-1RA | 0.53 0.26 to 1.07 | 0.08 | 0.64 0.31 to 1.30 | 0.21 |
| | ODM | 0.96 0.80 to 1.14 | 0.63 | 1.06 0.88 to 1.17 | 0.56 |

CVD, cardiovascular disease; DM, diabetes mellitus; GLP-1RA, glucagon like peptide-1 receptor agonist; ODM, oral diabetes medications.
Cardiovascular and Metabolic Risk

and Robert Kudray, Center for Assistance in Research using eRecord (CARE) at UPMC.

Contributors MTK, TO, MB, and LS had significant input toward the design and execution of the study. ML and JD contributed to data collection and medical record reviews. DD provided expertise for certain aspects of the study. MTK, TO, MB, and ML all contributed to interpretation of results. MTK prepared the manuscript draft with input and contributions from other authors.

Funding Funding for this study was provided by Sanofi-Aventis. The funder of this study had no role in the conduct of the study, data collection, analysis, or interpretation of data or in the writing of this report.

Competing interests None declared.

Ethics approval University of Pittsburgh IRB.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES


