PharmaSUG 2014 - Paper HA04

Linking Healthcare Claims and Electronic Health Records (EHR) for Patient Management – Diabetes Case Study

Paul A. LaBrec, Treo Solutions - 3M Health Information Solutions, San Diego, CA

ABSTRACT

Treo Solutions—now part of 3M Health Information Systems—conducted a pilot project to assess the feasibility of linking healthcare administrative claims data to an electronic health record (EHR) data extract to enhance patient case management activities. We linked one year of healthcare claims data (2012) to the equivalent year of medical record data abstracted from the EHR system of a large Midwest commercial insurer. The claims database identified 328,897 adult patients receiving services during 2012. Over 35,000 of these patients (10%) had a diabetes diagnosis. The clinical data set included 272,193 records on 61,532 patients in 2012 and included over 50 data elements. Measures identified in the EHR database included physical measures (the most common records), health history, health behaviors, radiologic and endoscopic tests, select prescription data and laboratory values. We abstracted a subset of EHR records for adults (ages 18-75) who had at least one diabetes-related test recommended by the National Quality Forum for use in this analysis. These tests include blood pressure, hemoglobin A1c, low-density lipoprotein, and retinal exams. From this combined database we calculated that the majority of patients with a diabetes diagnosis on claims had no diabetes test results for the study year. Furthermore, a small number of patients without a known diabetes diagnosis had at least one out-of-range diabetes test. We summarize the strengths and weaknesses of administrative claims versus EHR data for patient classification and compliance analyses, as well as methodological issues in combining claims and clinical databases. Planned follow-up analyses include medication fill rate calculations; cost of care predictions for various patient groups; and health outcomes analyses.

INTRODUCTION

The increasing collection, aggregation, standardization, and availability of electronic healthcare data from a variety of sources in various formats and delivery systems has resulted over recent years in an exponential increase of information available for the understanding and management of population health and healthcare. The variety of these databases and systems includes:

- commercial healthcare claims
- electronic medical record systems
- public insurance databases from the Centers for Medicare and Medicaid Services (CMS)
- state hospital discharge databases
- All Payer Claims Databases (APCD)
- vital statistics
- chronic and acute disease registries
- public health surveillance systems
- public and private survey research databases
- US Census surveys
- consumer attitudes and behavior databases
- clinical trial registries

Technological developments in information management systems and analytic methods have led to the explosion of 'big data' and 'data sciences' across numerous industries including health sciences and healthcare. Furthermore, the collection, movement, analysis, and delivery of healthcare claims is occurring in the regulatory environment of the Health Insurance Portability and Accountability Act (HIPAA) and Health Information Technology for Economic and Clinical Health Act (HITECH) enacted to protect Personal Health Information (PHI) (USDHHS, 2014a).

These recent technological and analytic developments have taken place in the context of the move of the US healthcare system from a predominantly fee-for-service delivery and payment model to a more population health and value-based model. Many of these developments are taking place in the regulatory environment of the Patient Protection and Accountable Care Act (PPACA) signed into law in 2010 (HHS) (USDHHS, 2014b). Features of the US healthcare system developing from this legislation include Accountable Care Organizations (ACOs) and Patient Centered Medical Homes (PCMH), and the new Health Insurance Exchanges. In each of those environments, a good understanding of population health status, health service utilization, and health provider quality and value is critical for program success (Hammond, 2011; Fillmore et al, 2013).

Prior to the advent of the PPACA, Don Berwick of the Institute for Healthcare Improvement (IHI) summarized this new population health, value-based emphasis in what he calls the "Triple Aim" (Berwick, 2008).

- Improving the patient experience of care (including quality and satisfaction)
- Improving the health of populations
- Reducing the per capita cost of healthcare

One of the ways in which the Triple Aim can be advanced in the context of health services research and healthcare delivery is comparative effectiveness research (CER) and disease management programs. This paper discusses a pilot research project to aid disease management programs in a commercial insurance healthcare analytics environment.

DESCRIPTION OF ADMINISTRATIVE CLAIMS DATA

As the name implies administrative claims data are managed primarily for the administration of payment for health services delivered by healthcare providers and facilities. Most administrative claims are based on the format of the CMS 1500 form for outpatient and provider services and the Universal Billing (UB-04) form for inpatient services. These forms collect patient information such as patient demographics (name, address, birthdate, gender, and marital status), employment and insurance status, occupational limitations, dates of service, diagnoses and procedures, service provider information, and charges for services. Due to the nature of the adjudication process for administrative claims, there is generally a 90-day claims 'run-out' period during which payer 'allowed' amounts are finalized before claims are added to a research database.

DESCRIPTION OF ELECTRONIC HEALTH RECORD DATA

An Electronic Health Record (EHR) is a digital version of a patient's paper medical chart. EHR's are real-time, patient-centered records that can make information available instantly and securely to authorized users. In addition to containing medical and treatment histories of patients, an EHR system can contain more than the standard clinical data collected in a provider's office. The Medicare and Medicaid EHR Incentive Programs provide financial incentives for the "Meaningful Use" of certified EHR technology to improve patient care (HealthIT.gov, 2014).

Meaningful Use, as defined by CMS, involves guidelines for capturing and using data elements such as patient demographics, medication use, potential medication interactions, clinical quality measures, and the protection of electronic health information.

There is debate in health services research regarding the degree to which administrative claims data continue to be valuable for health services research. A recent panel sponsored by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR, 2013) summarized two key points in this debate:

- Administrative claims data are a useful lens through which patterns of care, treatment outcomes, and health-care costs can be viewed. However, a paucity of clinical detail on study patients has historically constituted an important limitation of these data.
- Recent trends in the health sector—including the rise of patient-centered outcomes research (PCOR) and the proliferation of electronic medical records—combine to bring a heightened focus on patient characteristics and greater ability to incorporate clinical detail into retrospective research.

THE USE OF CLAIMS AND CLINICAL DATA FOR HEALTH SERVICES RESEARCH

As discussed above, administrative claims are produced primarily for the purpose of billing and paying for health services, while EHRs are produced primarily for recording and managing patient care. A recent study sponsored by the Agency for Healthcare Research and Quality AHRQ (West SL et al, 2010) evaluated the process and analytic yield of linking administrative claims data and electronic medical records using state Medicaid population claims and an academic medical center's EHRs. The study group concluded that, although many challenges exist in combining and analyzing claims and clinical data, the combination of these two sources of healthcare data creates an analytic resource stronger than either source individually, and a process worth continued evaluation and improvement.

Another study within the Kaiser diabetes patient population combined clinical databases, including diagnoses, laboratory results, prescription records, and patient reported information from laboratory records, pharmacy records, utilization records, and survey data to predict patients at high risk for short-term complications. They found a history of prior complications or related outpatient diagnoses was the strongest predictor of risk. For patients without prior history, various combinations of medication use (particularly insulin) and laboratory results (HbA1c, serum creatinine and albuminuria/microalbuminuria) were also predictive of future complications (Selby et al, 2001).

DIABETES

ETIOLOGY

Diabetes is a complex group of diseases with a variety of causes; a disorder of metabolism leading to high blood glucose or hyperglycemia. Diabetes develops when the body does not make enough insulin or is not able to use insulin effectively, or both. When pancreatic beta cells do not produce enough insulin or the body does not respond to the insulin that is present, glucose builds up in the blood instead of being absorbed by cells in the body, leading to prediabetes or diabetes. In diabetes, the body's cells are starved of energy despite high blood glucose levels. Sustained high blood glucose levels cause damage to nerves and blood vessels, leading to complications such as heart disease, stroke, kidney disease, blindness, dental disease, and amputation. Other complications of diabetes may include increased susceptibility to other diseases, loss of mobility with aging, depression, and pregnancy problems. Adult onset (Type 2) diabetes is the most common form, and develops most often in middle-aged and older people who are also overweight or obese. Other forms of the disease include juvenile diabetes (Type 1) and gestational diabetes. Prediabetes is a condition in which hemoglobin A1C levels—which reflect average blood glucose levels—are higher than normal but not high enough to be diagnosed as diabetes. People with prediabetes have an increased risk of developing Type 2 diabetes, heart disease, and stroke. Studies have demonstrated, however, that people with prediabetes who control their weight and increase their physical activity can prevent or delay Type 2 diabetes and in some cases return their blood glucose levels to normal (CDC, 2011).

EPIDEMIOLOGY

In the US, diabetes affects 25.8 million people of all ages (8.3 % of the US population). Of this number, approximately nineteen million have been diagnosed with diabetes, while an estimated seven million people have undiagnosed diabetes (NIDDK, 2011). A summary of diabetes and prediabetes incidence and prevalence is presented below.

- Among US residents ages 65 years and older, 10.9 million, or 26.9 %, had diabetes in 2010.
- About 215,000 people younger than 20 years had diabetes—Type 1 or Type 2—in the United States in 2010.
- About 1.9 million people ages 20 years or older were newly diagnosed with diabetes in 2010 in the United States.
- Diabetes is the leading cause of kidney failure, nontraumatic lower-limb amputations, and new cases of blindness among adults in the United States.
- Diabetes is the seventh leading cause of death in the United States.
- Based on fasting glucose or A1C levels, an estimated 35 % of US adults ages 20 years or older had prediabetes in 2005-2008—50 % of those ages 65 years or older. Applying those percentages yields an estimated 80 million Americans ages 20 years or older with prediabetes.
- The percentage of US adults ages 20 years or older with prediabetes was similar for non-Hispanic whites, 35 %; non-Hispanic blacks, 35 %; and Mexican Americans, 36 %.

Racial and Ethnic Differences in Diagnosed Diabetes

The National Institutes of Health and the Indian Health Service (IHS) provide national information for minority group diabetes diagnoses. After adjusting for age, survey data for people 20 years or older found the following race/ethnicity diagnosed diabetes rates: 7.1% for non-Hispanic whites, 8.4% for Asian Americans, 11.8% for Hispanics/Latinos, and 12.6% for non-Hispanic blacks. When comparing risk of diagnosed diabetes with non-Hispanic whites, Asian American adults had an 18% higher risk, 66% higher among Hispanic/Latino adults, and 77% higher among non-Hispanic black adults. Age-adjusted diabetes prevalence rates reported by the IHS vary by region and range from 5.5% among Alaska Native adults to 33.5% among American Indian adults in southern Arizona (NIDDK, 2011).

THE ECONOMIC COST OF DIABETES IN THE US

The direct and indirect costs of diabetes prevalence in the US is high, an estimated $174 billion dollars annually. Approximately $116 billion in direct medical costs are expended for diagnosis and treatment of the primary disease and its secondary consequences. In addition, there is an estimated $58 billion cost for disability, work loss, and premature mortality. Age-sex adjusted medical expenses for people with diabetes are more than 2 times higher than for people without diabetes (CDC, 2011).

DIABETES CARE MANAGEMENT

Type 2 diabetes and prediabetes are conditions in which early detection is appropriate. Both conditions are 1) common, 2) increasing in prevalence, 3) impose significant public health burdens; and there are 4) simple tests to detect preclinical disease readily available, and 5) effective interventions to prevent disease progression and reduce the risk of complications. Furthermore, early detection is important because Type 2 diabetes often has a long presymptomatic phase before diagnosis, and glycemic burden is a strong predictor of adverse outcomes. Unfortunately, Type 2 diabetes is frequently not diagnosed until complications appear. Some estimates suggest that approximately one-fourth of the US population may have undiagnosed diabetes (ADA, 2014).

Diabetes monitoring and treatment activities as standard of care, suggested by the American Diabetes Association, (*ADA, 2014*) include blood glucose and LDL cholesterol monitoring, screening and treatment for modifiable risk factors for cardiovascular disease; behavior modification including body weight management, increasing physical activity, and smoking cessation; and insulin therapy and other pharmaceutical treatment where appropriate.

METHODS

Administrative claims data used for this study were managed using a Microsoft® SQL Server 2005 and extracted into SAS® datasets for analysis using the SQL Procedure. Prior to populating the SQL claims data warehouse, claims data received from the insurer were processed using a rigorous data intake, formatting, enrichment, and validation process. The process includes tagging data with age group, claim category, care management program participation, and other flags; enriching data using various claims groupers with risk adjustment (e.g., 3M™ Clinical Risk Groups (CRG), All-Payer Refined Diagnosis Related Groups (APR-DRG), Medicare Severity Diagnosis Related Groups (MS-DRG), 3M™ Potentially Preventable Events, etc.); applying various business rules for accepting inpatient, outpatient, professional, and pharmacy claims; verifying member eligibility for services; and adding appropriate keys for linking claims data to various other data sources as required for analytics.

We used PROC SQL with Open Database Connectivity (ODBC) to extract data from our SQL Server warehouse using the following code convention, of which a generic example is reproduced below. User-defined elements in the CONNECT statement specific to server names and input and output tables are indicated in brackets with bold text and italics.

```
PROC SQL;
CONNECT to oledb(init_string="Provider=SQLOLEDB.1;Integrated Security=SSPI;Persist Security
Info=True;Initial Catalog=[SQL DATABASE NAME];Data Source=[SERVER NAME]");
CREATE TABLE [SAS LIBNAME.SAS TABLE NAME] AS SELECT * FROM CONNECTION TO OLEDB

(SELECT
        a.Person_ID,
        a.Person_DOB,
        a.Person_Gender,
        b.Pers_ID,
        b.EDC,
        c.*,
        d.*
FROM   [SQL TABLE NAME]      a
INNER JOIN [SQL TABLE NAME]  b      ON (b.Pers_ID = a.Person_ID)
LEFT JOIN [SQL TABLE NAME]   c      ON (c.Person_ID = b.Pers_ID)
LEFT JOIN [SQL TABLE NAME]   d      ON (d.Patient_Key = c.Pat_Key)
);
DISCONNECT FROM OLEDB;
QUIT;
```

Clinical data used for this investigation were provided by to us by an existing customer for whom we perform various data management and analytic services. Data files were received using a secure FTP upload and added to SQL tables on our research server using a process similar to that described above for administrative claims data. The customer also provided an identification number crosswalk file to enable us to match the patients reflected in the EHR extract with other administrative claims and eligibility databases.

Data manipulation and analysis was performed using SAS® v9.4. Data management and analysis programs were edited using SAS® Enterprise Guide v6.1. Data summaries presented include count, percent, mean, minimum,

maximum, and standard deviation (sd) calculated using PROC FREQ, PROC UNIVARIATE, or PROC MEANS. Statistical tests reported include chi square tests for categorical data using the FREQUENCY Procedure and t-tests for continuous data using the MEANS Procedure and the TTEST Procedure. When comparing the diabetes and non-diabetes cohorts, equality of variance of samples was determined using the F-Test and appropriate test options for equal or unequal variances were applied. We used the two-tail TTEST, making no a priori assumption about the direction or relative values for test measures between diabetics and non-diabetics. We considered diabetics and non-diabetics independent samples. The following options for PROC TTEST and PROC FREQ were used for between group comparisons.

```
PROC TTEST DATA=[LIBNAME.DATASET];
  CLASS diab_cohort;
        VAR lab1;
        WHERE lab_cnt=1;
        TITLE "Mean Highest BMI Values by Diabetes Cohort";
        TITLE2 "Adults Ages 18-75 yrs";
RUN;

PROC FREQ DATA=[LIBNAME.DATASET];
        TABLES diab_cohort*BMI_ADA  /chisq;
        WHERE lab_cnt=1;
        TITLE "BMI ADA Limits Lab Values by Cohort";
RUN;
```

For person-level analyses the analysis database--containing multiple visits per person and multiple tests per visit--had to be sorted and selected in order to represent person-level rather than record-level data. Furthermore, some analyses required the highest level lab test among many per person to be selected. In person-level analyses, a counter was used to identify the record of interest, an example of which follows. This record counter was invoked using a WHERE statement as in the PROC TTEST and PROC FREQ examples above.

```
PROC SORT DATA=[LIBNAME.DATASET1];
        BY person_id DESCENDING lab1;
RUN;

DATA =[LIBNAME.DATASET2];
  SET [LIBNAME.DATASET1];
        lab_cnt + 1;
        BY person_id;
        IF first.person_id THEN lab_cnt = 1;
        IF last.person_id THEN tot_lab = lab_cnt;
RUN;
```

In our examination of laboratory test results between cohorts, we did not apply any additional statistical control for age, gender or disease burden. The NQF ‘Optimal Diabetes Care’ criteria were designed to be applied to a diabetic population, while the ADA criteria consider the total at risk population. NQF defines a higher HbA1c threshold for the purpose of measuring ‘diabetes control’ than does ADA for the purpose of identifying potential cases of diabetes. In other words, the threshold levels themselves consider disease burden. Although an effect of age on HbA1c levels in non-diabetics has been suggested (Pani et al, 2008), no clear age-specific HbA1c thresholds have been defined. Neither NQF or ADA suggest additional age/gender-specific thresholds for the measures used in this study (NQF, 2012; ADA, 2014).

In our examination of TMA and health services utilization, however, we stratified the analysis by age, gender, and CRG weight, as these factors impact disease prevalence, total annual medical charges, and health services utilization (DeCola, 2012; Hall et al, 2010, Avrill, 1999).

COMMERCIALLY INSURED STUDY POPULATION

We began the analysis by identifying a cohort of patients covered by a mid-west US based commercial insurer. Over 80% of subjects resided in two mid-west US states; the majority (80%) were covered under a PPO plan type, while the remainder were covered by an HMO plan. We selected persons who had any utilization of inpatient, outpatient, or professional services during the 2012 calendar year. In order to investigate this population with respect to guidelines for Type 2 diabetes care suggested by the National Quality Forum (discussed below), the analysis was limited to the adult population ages 18 to 75 years. Age on December 31, 2012 determined inclusion in the study cohort. 3M™ Clinical Risk Groups (CRG) were used to define physical and behavioral health disease burden among the study population. CRGs use claims-based diagnoses to assign subjects to mutually exclusive, hierarchically ranked risk groups (Hughes, et al, 2004).

DIABETES COHORT DEFINITION

From the group of insured patients described above, we used the Episode Diagnosis Category (EDC) codes listed in Table 1 to identify our 'diabetes cohort'. EDC codes are diagnosis codes, which group similar diseases and then categorize the disease as either acute or chronic (Averill et al, 1999). We excluded Juvenile Onset (Type 1) diabetes (EDC 427) from the cohort definition. The distribution of our final list of EDC codes is presented in Table 1 below. Nearly all diabetics (93.6%) were classified with Type 2 diabetes as their dominant EDC code. A small proportion (2.6%) had the primary EDC of diabetic neuropathy.

EDC	EDC Description	% Patients
424	Diabetes	93.6%
428	Diabetes with Circulatory Complication	0.2%
429	Diabetic Coma	<0.1%
430	Diabetic Ketoacidosis	0.2%
431	Diabetic Nephropathy	0.3%
432	Diabetic Neuropathy	2.6%
433	Diabetic Retinopathy	1.8%
434	Other Diabetic Complications	1.2%

Table 1. EDC Code Distribution – Diabetes Cohort

The EHR data used in this analysis was provided by a commercial insurer serving patients in the mid-western US. The clinical dataset was extracted by the client from a MDDatacor clinical registry system and provided as multiple visit records per patient with multiple test records per visit. A total of 52 individual measures was provided. With the assistance of our Medical Director, we combined these measures into seven categories as described in Table 2. Measures categorized as laboratory measures accounted for 23.1% of unique measures, while other categories included behavioral measures (7.7%), diagnosis history (11.5%), physical measures (11.5%), radiologic tests (9.6%), scope examinations (5.8%), and indication of treatments, such as pharmaceutical and OTC medications and vaccinations (30.8%). Because no free text fields were included in the data extract, no natural language processing or other coding of open-ended data was conducted.

PCMH Trait Type	Examples	Traits	%
Behavioral Measure	Tobacco Use, Nutritional Counseling	4	7.7
History	Hx of AFib, MI, Asthma	6	11.5
Laboratory	HbA1c, LDL, FOBT, Renal Function	12	23.1
Physical Measure	Height, Weight, BMI, Blood Pressure	6	11.5
Radiologic Test	Mammogram, Ejection Fraction	5	9.6
Scope Examination	Colonoscopy, Flexible Sigmoidoscopy	3	5.8
Treatment	Beta Blocker, Aspirin, Vaccinations	16	30.8
TOTAL MEASURES		**52**	**100.0**

Table 2. Number and Type of Measures from Clinical Registry

Diabetes Care Management Assessment

In this study's EHR dataset, diabetes care management assessment was made using several variables available in the record and identified by the National Quality Forum (NQF) as elements of their "Optimal Diabetes Care" composite measure, part of the NQF Quality Positioning System. This NQF measure is used within the Medicare Shared Savings Program and Physician Quality Reporting System (PQRS) (NQF, 2012). In addition to the NQF criteria, we evaluated body mass index, as it was available and included in the American Diabetes Association (ADA) diabetes care measures (ADA, 2014). The thresholds for diabetes care criteria we used to compare diabetic and non-diabetic patients are listed in Table 3. The NQF threshold of < 8% HbA1c was used to indicate 'controlled' diabetes, as was a total blood pressure threshold of <140/90 mm Hg and a LDL limit of < 100 mg/dL. Other indicators available in the EHR data and included in this diabetes analyses were history of tobacco use, retinal exam and aspirin use.

An EHR was included in this study if it included any of the following test records: Hemoglobin A1c (% HbA1c), total blood pressure (mm Hg), low-density lipoprotein (mg/dL), aspirin use history, body mass index (BMI kg/m^2), tobacco use history, or retinal exam. In the EHR dataset, 'history' variables were indicated only with a "Y" if present at a visit. In addition to using NQF study criteria for diabetes, we examined HbA1c levels recommended by the ADA for defining both diabetic and prediabetic patients (Table 4).

Medical Charges and Utilization

We compared aggregated 2012 charges for the diabetes cohorts using the measure "total medical allowed (TMA) dollars." TMA is the sum of inpatient, outpatient (incl. hospital emergency department), and professional charges, and does not include pharmacy charges. For the financial analyses presented TMA was trimmed by excluding $0 dollar amounts, capped at $100,000 to limit the impact of outliers on average charges, and rounded to whole dollars. After trimming, 317,057 (96.4%) records remained for the financial analysis; 1,417 (0.5%) of those records had values for TMA above $100,000 which were capped. In order to conduct stratified analyses, we created categories for the continuous variables age and CRG weight by splitting the study population at the median (46 years for age and 1.8 for CRG weight).

Health services utilization was measured using total 2012 inpatient, outpatient, and professional claims separately. Inpatient admits included total 2012 hospital admissions for each person. Outpatient visits included total 2012 outpatient facility (e.g., hospital outpatient, emergency department) visits. Professional visits included total 2012 provider visits (e.g., physician office visits). Since 93.5% of persons had no 2012 inpatient admits, 52.7% had no 2012 outpatient visits, and both distributions were positively skewed, we categorized the distribution of these visit variables as "0", "1", and "2+". Although fewer patients (3.7%) had no professional visits, 65.4% had between 1 and 12 visits. and the range also was positively skewed. Furthermore, 12 visits per year is an average of 1 visit per month. For those reasons we categorized professional visits as "0", "1-12", and "13+".

Diabetes Management Criterion	NQF Optimal Care	Study Measures
Hemoglobin A1c (HbA1c)	< 8%	< 8%
Total Blood Pressure (BP)	< 140/90 mm Hg	< 140/90 mm Hg
Low-Density Lipoprotein (LDL)	< 100 mg/dL	< 100 mg/dL
Aspirin Use (History of ischemic vascular disease only)	Daily	Any in 2012
Body Mass Index	n/a	< 25 kg/m^2*
Tobacco Use	None	None
Retinal Exam	≥ Every 2 yrs	Any in 2012

Table 3. Diabetes Care Criteria Measured *American Diabetes Association

Category	ADA	NQF
Normal	< 5.7%	< 8%
Prediabetic	5.7 – 6.4%	---
Diabetic	≥ 6.5%	≥ 8%

Table 4. Study Criteria for Hemoglobin A1C

RESULTS

The full clinical record extract received included nearly 3 million records on over 126,000 unique individuals (Table 5). Although the date range of records provided spanned over five years (2008 through part of 2013), the year with the single largest number of records was 2012. The present analysis was limited to calendar year 2012. After applying inclusion criteria for age (ages 18 – 75 years) and diabetes care measures, and after linking the EHR records to available claims records through the process described above, 272,193 EHR records for 61,532 unique patients was available for analysis.

After linking the claims and clinical records, 328,897 patients were included in the study, 61,532 (18.7%) of whom had an electronic health record meeting study criteria (Table 6). While most study patients (89.2%) were non-diabetic, a greater proportion of diabetic (24.9%) than non-diabetic patients (18.0%) had at least one clinical (EHR) record. Furthermore, diabetics with EHR data had an average of 9.4 records per person versus 3.6 for non-diabetics.

The mean number of member months of claims data for all study subjects was 10.2 and varied little between diabetics (10.3) and non-diabetics (10.2) or between patients with an EHR record (10.7) or without (10.1).

Clinical Data Extract	Records (%)	Persons (%)
Full extract received	2,710,432 (100.0)	125,865 (100.0)
2012 records	764,859 (28.2)	92,309 (73.3)
2012 adult (ages 18-75) diabetes-related test records	272,193 (10.0)	61,532 (48.9)

Table 5. Study Patients and Test Records – EHR Data Only

Patients with Claims Records	Patients with EHR Records Available						Records per Person
	No	%	Yes	%	TOTAL	%	
Diabetic	26,625	75.1%	8,825	24.9%	35,450	10.8%	9.4
Non-Diabetic	240,740	82.0%	52,707	18.0%	293,447	89.2%	3.6
TOTAL	267,365	81.3%	61,532	18.7%	328,897	100.0%	4.4

Table 6. Study Patients by Cohort – EHR and Claims Data

Summary demographics and risk status for each cohort is described in Table 7. Of the 328,897 adults in the diabetes cohort analysis, the majority (89.2%) were considered 'non-diabetic' based on EDC classification in claims as described above. Slightly less than half (49.3%) of the diabetic cohort was female, while 59.5% of non-diabetics were female. The diabetic population was also an average of ten years older (54.2 yrs vs 43.7 yrs). As expected, the disease burden of the diabetic population, as measured by average CRG weight, was significantly higher at 3.9 versus 1.6. CRG weight can be interpreted as a measure of the expected resource consumption of a patient compared to the patient population average. The population CRG weight average is 1.0. A patient with a CRG weight of 2.0 is expected to use 2x the average resources as measured by Total Cost of Care (TCC) per member per month (PMPM). Total Cost of Care for the CRG weights used in this study is defined as the sum of inpatient, outpatient, professional, and pharmacy costs. All differences between cohorts displayed in Table 6 were statistically significant at the p<.0001 level.

Among all clinical measures received for our study patients, the overwhelming majority of records were the physical measures: BMI, weight, height, and total blood pressure. These test results accounted for 75.3% of all test results among the study cohort who had EHR data (Table 8). The volume of other measures represented in the data extract fell precipitously after those physical measures with tobacco use, mammogram, Influenza A vaccine, PAP test, Hemoglobin A1c, and Low-Density Lipoprotein (LDL), each having at least one record for between approximately 2% to 4% of patients in the database. Five of the seven diabetes study criteria we used are represented in this list of the most common measures in the EHR extract. We expected to find a much greater prevalence of HbA1c testing among diabetics, and the EHR data supported that expectation. Among the 8,825 diabetics identified in the EHR data (Table 6) 5,465 (61.9%) had at least one HbA1c test in 2012, while only 248 (0.5%) of non-diabetics had that test.

Comparative study measures between the diabetes cohorts are presented for continuous variables in Tables 9a and 9b. The NQF and ADA diabetes control targets are included. When all test results for each individual are considered (Table 9a), the cohorts differed to a statistically significantly degree on all measures except LDL, where average values were nearly identical. In addition, both cohorts met on average each of the NQF/ADA targets except BMI, where both cohorts exceeded the target of < 25 kg/m^2 (diabetics at 35.2 and non-diabetics at 29.1). As demonstrated by the standard deviations and maximum values of the measures, there are patients in each cohort who missed targets. When only the highest individual test result for each patient is considered (Table 9b) the cohorts also differed significantly on all measures except LDL. As with the view of all tests, the average highest individual test for BMI also was above the threshold for both cohorts. Unlike the average of all tests, however, the average highest LDL for both diabetics (102.5) and non-diabetics (103.2) was above the threshold of < 100.

Cohort	N (%)	Females (%)*	Mean Age (sd)*	Mean CRG Weight (sd)*
Diabetic	35,450 (10.8)	14,786 (49.3)	54.2 (11.0)	3.9 (9.3)
Non-Diabetic	293,447 (89.2)	174,572 (59.5)	43.7 (13.7)	1.6 (3.8)
TOTAL	328,897 (100.0)	192,058 (58.4)	44.8 (13.8)	1.8 (4.7)

Table 7. Study Cohort Demographic Profile and Health Status

*Differences between groups for each measure were significant at *p < .0001

In addition to examining test measures as continuous variables, we categorized those measures using the NQF and ADA criteria described in the Methods section in order to be able to identify patients with respect to thresholds for laboratory measures. Table 10 describes the highest individual HbA1c test levels by cohort using the NQF

categories: controlled (<8.0%) and uncontrolled (≥8.0%); and ADA categories: normal (<5.7%), prediabetic (5.7-6.4%), and diabetic (≥6.5%). A total of 5,713 persons had at least one HbA1c value for analysis; 5,465 (96.7%) of those patients were known diabetics. Among those diabetic patients the majority (72.4%) are considered 'controlled' using the NQF criterion. Applying the stricter thresholds in the ADA criteria to examine the diabetic population, we found that 4.3% had an HbA1c level considered in the 'normal' range, while 26.9% would be considered 'prediabetic', and 68.8% would be considered 'diabetic' based on this criterion alone. When these sets of criteria were applied to the highest HbA1c result (n=248) for the non-diabetic cohort, we found that 6 of those patients (2.4%) had levels considered 'uncontrolled' using the NQF criterion. When using the ADA criterion 20.6% of non-diabetics with an HbA1c result would be considered 'diabetic' by this criterion alone. An additional 56.1% with test results would be considered 'prediabetic' and only 23.4% would be considered 'normal.'

We examined the claims records of those 6 non-diabetics with elevated HbA1c in greater detail, including pharmacy claims (Table 11). Two (33.3%) of those patients had a somewhat elevated CRG weight, although both were still below the non-diabetic cohort average CRG weight of 1.6 (Table 7). One of those patients had significant pharmacy costs ($4,252).

Test Name	Number	%
BMI	543,100	26.4
Weight	540,226	26.3
Height	233,395	11.4
Total Blood Pressure	232,263	11.3
Tobacco Free	76,176	3.7
Mammogram	50,654	2.5
Tobacco Use Indicated	48,725	2.4
Influenza A Vaccine	41,675	2.0
Pap Test Cervical Cancer Screening	40,733	2.0
Hemoglobin A1c	39,156	1.9
Low-density lipoprotein	37,587	1.8
All Other Traits	173,290	8.4
TOTAL	**2,056,980**	**100.0**

Table 8. Most Common Clinical Measures Represented in Study Population

Diabetes Cohort	Measure (Units)	Tests (n)	Mean Value	SD	Min	Max	p > |t|	Control Target
Diabetic	HbA1c (%)	10,856	7.2	1.5	3.3	18.0	<.0001	< 8
Non-Diabetic		365	6.1	0.9	4.9	13.9		
Diabetic	BMI (kg/m^2)	27,945	35.2	8.0	5.1	95.4	<.0001	< 25
Non-Diabetic		122,400	29.1	6.8	5.0	91.2		
Diabetic	Systolic BP (mm Hg)	25,071	128.9	15.7	57.0	230.0	0.02	< 140
Non-Diabetic		43,296	129.2	15.6	53.0	240.0		
Diabetic	Diastolic BP (mm Hg)	25,071	77.3	10.2	16.0	130.0	<.0001	< 90
Non-Diabetic		43,296	80.1	10.5	11.0	160.0		
Diabetic	LDL (mg/dL)	7,386	97.1	35.1	5.7	299.0	0.09	< 100
Non-Diabetic		1,811	98.7	34.2	24.0	255.0		

Table 9a. Clinical Test Measure Study Cohort Comparison – All Tests

Diabetes Cohort	Measure (Units)	Persons (n)	Mean Value	SD	Min	Max	p > |t|	Control Target
Diabetic	HbA1c (%)	5,465	7.5	1.7	4.1	18.0	<.0001	< 8
Non-Diabetic		248	6.1	0.9	5.0	13.9		
Diabetic	BMI (kg/m^2)	8,262	35.4	8.2	10.9	95.4	<.0001	< 25
Non-Diabetic		51,358	29.1	6.7	5.2	91.2		
Diabetic	Systolic BP (mm Hg)	7,256	137.3	16.4	90.0	230.0	<.0001	< 140
Non-Diabetic		15,587	136.4	16.3	80.0	240.0		
Diabetic	Diastolic BP (mm Hg)	7256	78.3	10.0	36.0	130.0	<.0001	< 90
Non-Diabetic		15,587	81.0	10.5	32.0	140.0		
Diabetic	LDL (mg/dL)	4,554	102.5	36.2	15.0	299.0	0.57	< 100
Non-Diabetic		1,245	103.2	35.3	30.0	255.0		

Table 9b. Clinical Test Measure Study Cohort Comparison – Highest Test per Person

		NQF		ADA		
	Total Persons	Controlled (<8%)	Uncontrolled (≥8%)	Normal (<5.7%)	Prediabetic (5.7-6.4%)	Diabetic (≥6.5%)
Diabetes Cohort	(n) (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Diabetic	5,465 (96.7)	3,958 (72.4)	1,507 (27.6)	232 (4.3)	1,472 (26.9)	3,761 (68.8)
Non-Diabetic	248 (4.3)	242 (97.6)	6 (2.4)	58 (23.4)	139 (56.1)	51 (20.6)
Total	5,713 (100.0)	4,200 (73.5)	1,513 (26.5)	290 (5.1)	1,611 (28.2)	3,812 (66.7)

Table 10. HbA1c Thresholds Using NQF and ADA Criteria – Highest test per Person

Patient	Gender	Age	EDC	EDC Description	CRG Weight	Highest HbA1c (%)	Total Medical Allowed ($)	Rx Allowed ($)
1	M	40	851	HIV Disease	1.18	8.4	622	606
2	F	61	740	Cluster - Minor Infections	0.21	11.3	122	35
3	F	61	844	Signs, Symptoms, and Findings	0.27	10.1	285	0
4	M	56	844	Signs, Symptoms, and Findings	0.45	13.9	516	30
5	M	61	842	History of Significant Prescription Medication NEC	1.06	8.8	173	4,252
6	M	58	844	Signs, Symptoms, and Findings	0.66	9.4	1,918	45

Table 11. Select Characteristics of Non-Diabetics with Elevated HbA1c Values

As described in Table 3 there were three indicator variables included among the NQF diabetes-related test measures and available in the clinical data feed: tobacco use, aspirin use, and retinal exam. These variables were coded with

a “Y” when present. There were no “N” values in the data, so it is not clear whether or not any missing values can be inferred as “N”. For this analysis we calculated the prevalence of “Y” indicators among total EHR records for diabetics and non-diabetics (Table 12). The prevalence of aspirin use among diabetics was higher than non-diabetics (4.5% to 1.1%). We did not confirm through additional claims data investigation, however, whether or not the aspirin use among diabetics recorded in the EHR data was associated with patients with a history of ischemic vascular disease (IVD) per the NQF monitoring recommendation. Retinal exams were far more prevalent among diabetic (10.7%) than non-diabetics (<0.1%). Recorded rates of tobacco use in the EHR were similar between diabetics (19.6%) and non-diabetics (21.1%).

	Diabetic	Diabetics w/ EHR Data (n=8,825)	Non-Diabetic	Non-Diabetics w/ EHR Data (n=52,707)	Total	Total Patients w/ EHR Data (n=61,532)
Indicator	n	%	n	%	n	%
Aspirin Use	397	4.5%	581	1.1%	978	1.6%
Retinal Exam	947	10.7%	25	0.0%	972	1.6%
Tobacco Use	1,726	19.6%	11,096	21.1%	12,822	20.8%

Table 12. Prevalence of Diabetes Care Indicators by Diabetes Cohort

TOTAL MEDICAL ALLOWED AND UTILIZATION

We compared the charges for medical care incurred among the study population using the measure Total Medical Allowed (TMA). Mean TMA for the study population with trimmed records available for analysis (n=317,057) was $4,662 (sd=$11,025). We compared TMA between diabetes cohorts stratifying the analysis on gender, age group, and disease burden as measured by CRG weight. These variables were stratified as described above. Results of the stratified analyses for TMA are presented in Table 13. In each risk factor category, the mean TMA for diabetics was greater than that for non-diabetics. As expected the group with the largest mean TMA ($10,778; sd=$19,126) was diabetics with a CRG weight greater than the study population median. The group with lowest mean TMA ($1,383; sd=$2,564) was non-diabetics with a CRG weight below the median.

We also compared utilization of health services by these same risk factor/cohort groups. Results of the stratified analyses for utilization are presented in Table 14. A greater proportion of the diabetic cohort had at least 1 inpatient admit (11.8% versus 5.9%), while a slightly greater proportion of non-diabetics had at least 1 outpatient visit (59.1% versus 56.9%). A greater proportion of non-diabetics had between 1 and 12 professional visits (67.4% versus 49. 6%); while more diabetics had 13 or more professional visits (46.3% versus 29.0%). In each age group (above and below the median age) diabetics had a greater proportion of patients in the highest frequency group of admits and visits. When comparing diabetes cohort-specific utilization rates between patients above or below the median CRG weight, differences were less pronounced than for other risk stratification groups. In fact, the inpatient utilization rates were not statistically significant between diabetes cohorts for those patients below the median CRG weight (those using fewer resources than others). Other chi square results in Table 14 were significant at the p < .0001 level.

Risk Factor	Cohort	n	Total Medical Allowed $ (sd)	**p > |t|**
Females	Diabetic	16,946	8,368 (16,175)	< .0001
	Non-diabetic	169,520	4,324 (9,689)	
Males	Diabetic	17,065	8,161 (7,898)	< .0001
	Non-diabetic	113,526	4,090 (4,029)	
Age <= median	Diabetic	7,208	6,674 (14,453)	< .0001
	Non-diabetic	153,515	3,458 (8,044)	

Table 13. Total Medical Allowed by Risk Factors and Diabetes Cohort

Risk Factor	Cohort	n	Total Medical Allowed $ (sd)	p > |t|
Age > median	Diabetic	26,803	8,692 (17,450)	< .0001
	Non-diabetic	129,531	5,144 (11,861)	
CRG Wt <= median	Diabetic	9,324	1,608 (3,104)	< .0001
	Non-diabetic	148,188	1,383 (2,564)	
CRG Wt > median	Diabetic	24,687	10,778 (19,126)	< .0001
	Non-diabetic	134,858	7,357 (13,578)	

Table 13 (cont). Total Medical Allowed by Risk Factors and Diabetes Cohort

		Inpatient Admits			Outpatient Visits			Professional Visits		
Risk Factor	Cohort	0 (%)	1-2 (%)	3+ (%)	0 (%)	1-2 (%)	3+ (%)	0 (%)	1-12 (%)	13+ (%)
Females	Diabetic	15,366 (87.9)	1,494 (8.5)	626 (3.6)	6,299 (36.0)	3,685 (21.1)	7,502 (42.9)	554 (3.2)	7,790 (44.6)	9,142 (52.3)
	Non-diabetic	162,776 (93.2)	10,236 (5.9)	1,560 (0.9)	86,656 (49.6)	39,208 (22.5)	48,708 (27.9)	5,212 (3.0)	112,076 (64.2)	57,284 (32.8)
Males	Diabetic	15,917 (88.6)	1,461 (8.1)	586 (0.4)	8,984 (50.0)	3,068 (17.1)	5,912 (32.9)	920 (5.1)	9,777 (54.4)	7,267 (40.5)
	Non-diabetic	113,345 (95.4)	4,428 (3.7)	1,102 (0.9)	7,236 (59.9)	22,899 (19.3)	24,740 (20.8)	5,530 (4.7)	85,587 (72.0)	27,758 (23.4)
Age <= median	Diabetic	6,795 (88.9)	633 (8.3)	215 (2.8)	3,863 (50.5)	1,399 (18.3)	2,381 (31.2)	440 (5.8)	4,220 (55.2)	2,983 (39.0)
	Non-diabetic	149,105 (93.8)	8,784 (5.5)	1,131 (0.7)	94,932 (59.7)	31,278 (19.7)	32,810 (20.6)	5,714 (3.6)	112,742 (70.9)	40,564 (25.5)
Age > median	Diabetic	24,488 (88.0)	2,322 (8.4)	997 (3.6)	11,420 (41.0)	5,354 (19.3)	11,033 (39.7)	1,034 (3.7)	13,347 (48.0)	13,426 (48.3)
	Non-diabetic	127,016 (94.5)	5,880 (4.4)	1,531 (1.1)	62,960 (46.8)	30,829 (22.9)	40,638 (30.2)	5,028 (3.7)	84,921 (63.2)	44,478 (33.1)
CRG Wt <= median	Diabetic	10,138 (99.6)	38 (0.4)	1 (<0.1)	6,479 (63.7)	1,978 (19.4)	1,720 (16.9)	874 (8.6)	8,008 (78.7)	1,295 (12.7)
	Non-diabetic	155,894 (99.7)	528 (0.3)	13 (<0.1)	103,192 (66.0)	32,695 (20.9)	20,548 (13.1)	8,475 (5.4)	133,069 (85.0)	14,891 (9.5)
CRG Wt > median	Diabetic	21,145 (83.7)	2,917 (11.5)	1,211 (4.8)	8,804 (34.8)	4,775 (18.9)	11,694 (46.3)	600 (2.4)	9,559 (37.8)	15,114 (59.8)
	Non-diabetic	120,227 (87.8)	14,136 (10.3)	2,649 (1.9)	54,700 (39.9)	29,412 (21.5)	52,900 (38.6)	2,267 (1.7)	64,594 (47.1)	70,151 (51.2)

Table 14. Health Services Utilization by Risk Factors and Diabetes Cohort

DISCUSSION

Our data provide evidence of additional clinical testing among the diabetic versus non-diabetic patient populations. The overall number of EHR records per patient was higher among diabetics—more than 2.5 times greater—than non-diabetics. The most common test measures found in both cohorts were physical measures—height, weight, BMI, and blood pressure. This finding was expected as these measures are routinely collected for all patient healthcare encounters. The volume of some other measures reflected in the data, however, was less than anticipated. We expected, for example, with respect to diabetics that there would be records of foot exams, as this is a common exam to assess peripheral neuropathy (Boulton et al, 2008). Although 'foot exam' was an available test field in the clinical database, there were no records for that test. We were, on the other hand, able to document a much higher rate of HbA1c testing among diabetics (61.9%) than non-diabetics (0.5%). We observed that nearly three-quarters of known diabetics had HbA1c levels considered "controlled" by NQF standards. Given the prevalence of prediabetes in the non-diabetes cohort (56.1%) (Table 10), along with a mean BMI of 29.1 kg/m^2 (Table 9b), we might have expected a somewhat higher prevalence of testing in that cohort.

The ability to identify patients who have diabetes test measures outside the normal range, yet have no known diagnosis of diabetes in claims data is an important finding of the study. This linked analysis can produce a list of those patients for follow-up by healthcare providers and care managers to assess whether or not they are prediabetic or have undiagnosed diabetes. Early detection of predisease can greatly improve patient quality of life and reduce healthcare costs (ADA, 2014). We were able to identify 6 such patients for recommended additional follow-up.

The generally increased costs and utilization among diabetic patients observed in this study was an expected outcome. When the analysis was limited to patients with CRG weight above the median, the difference persisted. A cost and utilization comparison group for the diabetic cohort in future analyses will be the subset of non-diabetics with other chronic conditions. Although BMI (from a risk perspective) and CRG weight (from an expected cost perspective) are to some degree measures of comorbidity, we did not explicitly investigate the effects of comorbidities in the diabetic population on test measures, utilization and costs. Physical comorbid conditions can exacerbate a diabetic patient's general health risk, and mental health comorbidities such as depression can severely impact a patient's ability to care for their chronic conditions (Piette and Kerr, 2006).

Based on our identification of non-diabetics with HbA1c levels in the ADA 'prediabetic' range, additional investigations into the health of this population are warranted. We plan to examine the additional years of EHR data available to identify the presence of other out-of-range test values in previous and subsequent years. Given the increasing prevalence of diabetes in recent years, early diagnosis and treatment has become increasingly important.

Although statistical hypothesis tests were performed to assess differences between the diabetes and non-diabetes cohorts, the large number of patients in our study resulted in some statistically significant differences that may not be clinically significant. For example, the difference of 0.9 mm Hg in systolic blood pressure between cohorts, while statistically significant holds no clinical relevance. This caution in interpreting statistical test results is widely applicable when analyzing the very large number of patients usually found in healthcare claims databases.

Limitations of the study include the potential for misclassification of patients if any errors exist in the patient identification crosswalk file provided by the insurer to connect EHR IDs to claims IDs. Although some behavioral data exist in the EHR records (e.g., smoking and aspirin use), important dietary risk factor information does not. Furthermore, no diabetes treatment data (e.g., insulin or metformin prescriptions) were assessed as part of this study.

Since the audience of the conference presentation of this paper primarily works with clinical trial data, we present a summary (Table 15) comparing and contrasting various dimensions of clinical trials, EHRs, and healthcare claims. From a SAS® programming perspective one of the key differences between clinical trial data and healthcare claims data is the size of datasets. While clinical trial database are relatively small, healthcare claims databases can be quite large, making efficient data processing and careful attention to code accuracy more important in program development, as risk of reaching disk space limitations or wasting processing time are greater with large datasets.

With respect to analytic techniques, while the randomized clinical trial (RCT) is the gold standard in pharmaceutical research, this design is rare in disease management programs. Generally, all patients with a given disease/severity history or risk are encouraged to participate in disease management programs. Their participation, however, cannot be randomized. In this case, other natural experimental designs such as difference-in-differences (D-in-D) analysis are required. In a D-in-D analysis, patients who chose to participate in a care management program such as a Patient Centered Medical Home (PCMH) are compared to similar patients (via matching techniques) who chose not to participate. PCMH patients are followed retrospectively and prospectively around a program entry time point to assess utilization and cost outcomes pre- and post- program. Their results are compared to comparable non-PCMH (control) patients over similar time periods. Post-program differences in outcomes are compared to assess program impact (Faries et al, 2010).

Feature	Clinical Trials	Electronic Health Records	Healthcare Claims
Description	Mainly small datasets (dozens to thousands of records) based on Case Report Forms (CRF) designed for research to support drug development.	Datasets of varied sizes created from domains related to patient care designed for the managing the health history, diagnosis, and treatment of patients in a clinical setting.	Mainly very large datasets (thousand to millions of records) based on standardized healthcare insurance eligibility and service billing records designed primarily for the purpose of paying for health services.
Financial data	No	No	Yes
Patient Populations	Clinical Trial participants.	Patients of individual or group practice.	Covered lives in commercial or public insurance programs.
Data structures	Various CRF designs, with some standards such as CDISC and ADaM. Structures vary, particularly in data such as laboratory data.	Numerous user interface and data management products producing many data elements in both structured and unstructured data.	Fairly standardized claim data formats, although data warehouse structures can vary by payer.
Completeness	Fairly complete. Data submission requirements for clinical trials and direct follow-up with data submitters.	Completeness can vary by record system and healthcare provider diligence.	Varies by submitter and diligence in including data not directly necessary for payment.
Research Methods	Structured (RCT) analyzing clinical outcomes (efficacy). --blinded and unblinded.	Various descriptive and analytic methods, including textual analysis on free-text data.	Descriptive and analytic methods using natural experimental designs for non-random patient/provider behavior.
Analytics	Hypothesis testing to support safety and efficacy assessment. Primarily summary analysis with limited and predefined interim analyses. Static datasets and TLF output with interpretation in CSR.	Analysis of various patient physical, clinical, and behavioral measures to classify and monitor patient health. Often real-time or near real-time dashboards for providers to employ in daily patient management.	Effectiveness of interventions (therapies and clinical practice programs like patient-centered medical homes) in clinical practice environment. Support of payment transformation and quality/value improvement programs. Analytics to compare and predict health status, risk, and cost. Flexible population and provider analytics on regularly refreshed databases of adjudicated claims.
Data Consumers	Pharma/biotech companies, FDA, general public.	Healthcare providers, researchers.	Health insurers, providers, researchers, sometimes public (de-identified claims in APCD).

Table 15. Comparative Features of Healthcare Analytic Data Sources

CONCLUSION

Through this pilot project we demonstrated our ability to link administrative claims and clinical data in order to examine the relationship between diagnostic, demographic, clinical, and patient financial data. We demonstrated the ability to receive clinical data outside our usual administrative claims data feed, but with the same data transfer security and similar data integrity testing. Finally, we demonstrated that we can conduct useful analyses that provide insight into patient care and health status that is not available using either claims data or clinical data alone.

Next steps in this analysis include more detailed investigation into the health services utilization and charge profiles of the diabetes and non-diabetes cohorts is planned. Of particular interest to our research group and our customers is the development of predictive models for health status, health behaviors, health costs and outcomes. We have developed such models for various measures of costs and utilization using traditional regression-based techniques and more contemporary machine-learning algorithms. We will apply those techniques to additional analyses of claims/clinical aggregated databases. The EHR data feed analyzed includes measures relevant to other diseases in addition to diabetes. We plan to conduct similar analyses using measures relevant to heart disease and respiratory disease. We plan to conduct a search for ischemic heart disease history among diabetic patients in the claims data in order to assess aspirin use in that diabetic subpopulation and fully assess that NQF measure. In addition, we plan to conduct a more detailed analysis of pharmacy claims data in conjunction with the EHR data. The EHR data include indicators for whether or not a prescription was written for certain classes of medications (e.g., beta blockers and ACE Inhibitors). Linking this information with pharmacy claims we can calculate prescription fill rates and medication adherence and persistence rates. Furthermore, we can assess utilization of diabetes treatments such insulin and metformin. Using disease screening test indicators we can calculate utilization rates for mammography, colonoscopy, flexible sigmoidoscopy, and other tests. Most of the analyses conducted or discussed herein can be conducted using multiple years of data to assess trends and validate findings from single-year studies.

Finally, we plan to review these and subsequent findings with our insurance partner in order to determine whether or not these linked analyses are identifying any additional patients who require additional disease assessment or treatment.

REFERENCES

American Diabetes Association (ADA). January 2014. “Standards of Medical Care in Diabetes – 2014.” Diabetes Care.

Averill RF, Goldfield NI, Eisenhandler J, et al. 1999. “Development and Evaluation of Clinical Risk Groups (CRGs).” Wallingford, CT: 3M HIS Research Report.

Berwick DM, Nolan TW, Whittington J. May 2008. “The Triple Aim: Care, Health, and Cost.” Health Affairs.

Boulton AJ, Armstrong DG, Albert SF, et al. August 2008. “Comprehensive Foot Examination and Risk Assessment: A Report of the Task Force of the Foot Care Interest Group of the American Diabetes Association, with Endorsement by the American Association of Clinical Endocrinologists.” Diabetes Care.

Centers for Disease Control and Prevention (CDC). 2011. “National Diabetes Fact Sheet: National Estimates and General Information on Diabetes and Prediabetes in the United States, 2011.” Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

DeCola PR. 2012. “Gender Effects on Health and Healthcare,” in Schenck-Gustafsson K, DeCola PR, Pfaff DW, Pisetsky DS (eds): Handbook of Clinical Gender Medicine. Basel, Karger.

Faries, Douglas; Haro, Josep Maria; Obenchain, Robert L.; Leon, Andrew C, (2010-01-25). Analysis of Observational Health Care Data Using SAS (Kindle Locations 210-213). SAS Institute/SAS Publishing. Kindle Edition.

Fillmore H, Dubard CA, Ritter GA, Jackson CT. September 21, 2013. “Health Care Savings with the Patient-Centered Medical Home: Community Care of North Carolina’s Experience.” Popul Health Manag.

Hall MJ, DeFrances CJ, Williams SN, et al. 2010. “National Hospital Discharge Survey: 2007 Summary. National Health Statistics Reports; no 29.” Hyattsville, MD: National Center for Health Statistics.

Hammond E. “Success Factors for Creating Accountable Care Organizations.” http://healthaffairs.org/blog/2011/06/03/success-factors-for-creating-accountable-care-organizations/. Health Affairs Blog. 4/20/14.

HealthIT.gov. “What is an Electronic Health Record (EHR)?” http://www.healthit.gov/providers-professionals/faqs/what-electronic-health-record-ehr. 4/19/14.

Hughes JS, Averill RF, Eisenhandler J, et al. January 2004. “Clinical Risk Groups (CRGs): A Classification System for Risk-Adjusted Capitation-Based Payment and Health Care Management.” Med Care.

ISPOR Issues Panel. May 20, 2013. “Are Administrative Claims Data Becoming Obsolete?” New Orleans, LA: ISPOR 18th Annual International Meeting, Session 1.

National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). September 9, 2013. “National Diabetes Statistics, 2011.” Bethesda, MD: National Diabetes Information Clearinghouse.

National Quality Forum (NQF). 2012. “Optimal Diabetes Care (Composite Measure).” Quality Positioning System (QPS) Measure Description Display Information. Minneapolis, MN: MN Community Measurement.

Pani LN, Korenda L, Meigs JB, et al. October 2008. “Effect of Aging on A1C Levels in Individuals without Diabetes: Evidence from the Framingham Offspring Study and the National Health and Nutrition Examination Survey 2001–2004.” Diabetes Care.

Piette JD and EA Kerr. “The Impact of Comorbid Chronic Conditions on Diabetes Care.” March 2006. Diabetes Care.

Selby JV, Karter AJ, Ackerson LM et al. August 2001. “Developing a Prediction Rule From Automated Clinical Databases to Identify High-Risk Patients in a Large Population With Diabetes“. Diabetes Care.

U.S. Department of Health and Human Services (HHS). “HIPAA Administrative Simplification Statute and Rules.” http://www.hhs.gov/ocr/privacy/hipaa/administrative/index.html. 4/19/14, (2014a).

U.S. Department of Health and Human Services (HHS). “Read the Law.” http://www.hhs.gov/healthcare/rights/law/. 4/19/14, (2014b).

ACKNOWLEDGEMENTS

The author acknowledges the assistance of Melissa Gottschalk in validation of the SAS® code used in this analysis, in preparing references, and in manuscript editing. The author wishes to thank Herb Fillmore for advice on the statistical analyses and Gordon Moore for advice on clinical concepts.

CONTACT INFORMATION

Name: Paul A. LaBrec
Enterprise: Treo Solutions-3M Health Information Systems
Address: 125 Defreest Drive
City, State ZIP: Troy, NY 12180
Work Phone: 518.426.4315
E-mail: plabrec@treosolutions.com
Web: www.treosolutions.com