Collaborative Analytics in Action: A Case Study Focused on Treatment Patterns

Gary Schneider, MSPH, ScD Epidemiologist
Stephanie Reisinger Vice President, Technology Solutions
Matthew Reynolds, PhD Vice President, Scientific Development

Introduction

For the past seven years, Evidera scientists have been on the forefront of research into the use of a Common Data Model (CDM) to enable standardized healthcare analytics, participating as the principal investigator on several Observational Medical Outcomes Partnership (OMOP) research initiatives[1,2] as well as a collaborator with the Observational Health Data Sciences and Informatics (OHDSI) program.[3] A companion article in this issue of *The Evidence Forum*, *"Will the Growing Reliance on Real-world Data Fuel Fundamental Changes in the Way We Approach Database Analyses?"* describes in detail how a standardized approach to database analysis can enable an environment of "collaborative analytics" where analysis programs and clinical event definitions are collaboratively developed, shared, and re-used within and across organizations. This article describes the results of a collaborative analytics research project performed by Evidera scientists in collaboration with scientists at GSK and BMS. The research has been presented at ISPOR[4] and ICPE.[5]

Background

Across the industry there has been increasing interest in the use of a Common Data Model (CDM) to facilitate systematic analyses of large administrative claims (Claims) and electronic medical records (EMR) databases for real-world evidence generation, and recent research highlights the benefits of this approach.[6]

The concept of the CDM is that data from disparate databases can be transformed into a common data format using consistent assumptions. After transformation, systematic analysis can be performed in a rapid and efficient manner. Because the data has been transformed using consistent rules and analyzed using a single, standardized analysis module written for the CDM, the results across disparate data sources can be efficiently produced and meaningfully compared.

This article presents the results of a collaborative analysis of treatment patterns in patients diagnosed with depression across five electronic healthcare databases. Prior to analysis, each of the five databases used in the analysis was transformed into the OMOP CDM format, and then analyzed with a single standardized treatment pattern modular program written to conform to the OMOP CDM. Evidera scientists performed the analysis on one of the databases; the other four were analyzed by scientists at GSK and BMS using licensed observational databases. The parameters used as input to the treatment patterns modular program were identical for each execution. Results of the analysis were compared to better understand similarities and differences across databases and patient populations.

Methods

Source data

Source data came from five distributed sources of HIPAA-compliant patient data, details of which are provided below. Each database was transformed into an OMOP-compliant CDM prior to analysis. Databases were distributed across four physical locations in the U.S. (Pennsylvania, Massachusetts, Connecticut, and North Carolina). Access to the Truven, Pharmetrics, and GE data were covered by data licenses and analyzed independently by the data licensors.

Data sources used were:

- **CCMC - Truven Marketscan**: Commercial Claims and Medicare supplemental claims data. These data are fully integrated, patient-level data containing inpatient, outpatient, drug, laboratory, health risk assessment, and benefit design information from 87 million commercial and 10 million Medicare patients in the most recent five years across the U.S.

- **Medco - Medco Pharmacy Claims**: Commercial Claims data (pharmacy and integrated medical claims) on a subset of 12.7 million patients in the most recent five years across the U.S.
- **GE - GE Centricity**: Ambulatory Electronic Medical Record (EMR) data on approximately 13.5 million patients contributed by 30,000 clinicians in 49 states within the U.S.
- **PM - IMS Pharmetrics**: Commercial Claims data (pharmacy and integrated medical claims) on a subset of approximately 35 million patients in the most recent five years across the U.S.
- **MDCD - Truven Medicaid**: Government Medicaid Claims data originating from multiple states within the U.S. on approximately 12 million patients.

Table 1: Diagnosis codes used in depression cohort definition

Major depression diagnosis codes
296.2 Major depressive disorder, single episode
296.20 Major depressive affective disorder, single episode, unspecified
296.21 Major depressive affective disorder, single episode, mild
296.22 Major depressive affective disorder, single episode, moderate
296.23 Major depressive affective disorder, single episode, severe, without mention of psychotic behavior
296.24 Major depressive affective disorder, single episode, severe, specified as with psychotic behavior
296.25 Major depressive affective disorder, single episode, in partial or unspecified remission
296.26 Major depressive affective disorder, single episode, in full remission
296.3 Major depressive disorder, recurrent episode
296.30 Major depressive affective disorder, recurrent episode, unspecified
296.31 Major depressive affective disorder, recurrent episode, mild
296.32 Major depressive affective disorder, recurrent episode, moderate
296.33 Major depressive affective disorder, recurrent episode, severe, without mention of psychotic behavior
296.34 Major depressive affective disorder, recurrent episode, severe, specified as with psychotic behavior
296.35 Major depressive affective disorder, recurrent episode, in partial or unspecified remission
296.36 Major depressive affective disorder, recurrent episode, in full remission
298.0 Depressive type psychosis
Adjustment disorder with depressed mood
309.0 Adjustment disorder with depressed mood
309.1 Prolonged depressive reaction
Adjustment disorder with mixed anxiety and depressed mood
309.28 Adjustment disorder with mixed anxiety and depressed mood
311 Depressive disorder, not elsewhere classified

Common data model[7]

The standardized format of the OMOP CDM is patient-centric, organizing de-identified patient data into a "Person Timeline" format to facilitate longitudinal analysis. Information included for each person includes a unique identifier, demographic information, and an "observation period" during which healthcare encounters (e.g., conditions, medications, procedures, and visits) are recorded. All healthcare encounters include a start date, as well as an end date where appropriate.

Standardization of the data content is accomplished via a Terminology Dictionary that includes standardized condition and drug vocabularies. ICD-9-CM codes and drug product identifiers (e.g., National Drug Code, Generic Product Identifier) from source data were mapped into the standardized vocabulary.

Cohort definition

Patients, between the ages of 18 to 65, were selected who had a diagnosis of depression between January 1, 2008, and June 30, 2009. Depression was identified using ICD-9-CM codes listed in Table 1. Patients were required to have 180 days of depression-free eligibility prior to their index depression diagnosis.

Analysis

Descriptive statistics of age, gender, and the number and proportion of patients with a qualifying first-line treatment were computed separately for each database/condition combination. "Overall means," e.g., the average across all databases for age, gender, and first-line treatment were calculated as weighted averages of the database-specific mean values.

Treatment Patterns - Patients who were newly diagnosed (i.e., no depression diagnosis during the 180-day baseline interval) and newly treated (i.e., no baseline antidepressant prescription/use) with a first-line antidepressant within 60 days following index depression diagnosis were identified. First-line medications were categorized into antidepressant drug class (i.e., Selective serotonin reuptake inhibitors (SSRI), Serotonin–norepinephrine reuptake inhibitors (SNRI), Tricyclic antidepressants (TCA), Monoamine oxidase inhibitors (MAOI), and Other). Prescriptions of the same antidepressant occurring within 30 days of each other were combined into one first-line treatment episode. Patients were followed for 365 days following the start of first-line treatment. Medication treatment patterns (definitions below) were identified by examining the data through 30 days following the end of the first-line treatment episode:

- **Continued:** First-line treatment episode continued beyond 365 days.
- **Discontinued:** First-line treatment episode discontinued, with no other antidepressant prescribed within 30 days after discontinuation.
- **Augmented:** A second antidepressant was prescribed during the first-line treatment episode, with at least one additional prescription of the first-line treatment occurring after the prescription for the second antidepressant.
- **Switched:** A second antidepressant was prescribed either during the first-line treatment episode or within 30 days after first-line treatment episode ended. No additional prescriptions for first-line treatment occurred after initiation of the second antidepressant.

Mean and median treatment days were evaluated for each treatment group. In addition, the total number of treatment days occurring during the 365 day follow-up was tabulated. Similarly, the Proportion of Days Covered (PDC) occurring during follow-up was calculated by dividing the number of first-line treatment episode days occurring during the 365-day follow-up period by 365, and multiplying the result by 100. Note that although follow-up for treatment days was limited to 365 days, overlapping prescriptions were not accounted for (i.e., if a refill occurred prior to the end of days' supply from the immediately preceding prescription, the overlap would be counted twice), meaning that treatment days greater than 365 was possible.

Results

All analyses results described below were produced in less than two days (design through analysis completion).

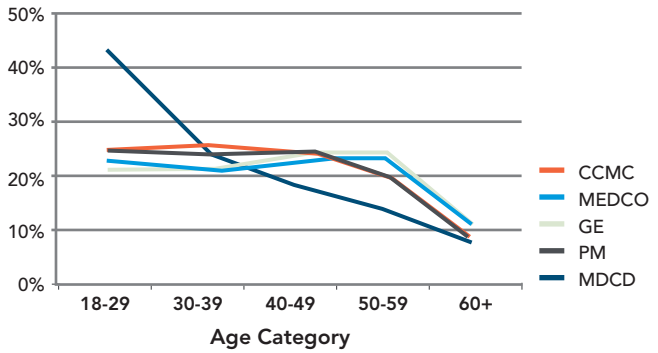
Descriptive information

Demographic characteristics were generally similar across all databases. Overall, approximately two-thirds of subjects were female; only the MDCD data varied from this substantially, having 77.7% females. The average age was 39.2 years; with the MDCD subjects being notably younger (34.8 years) than subjects originating from other data sources (Table 2). The age distributions of data used for the treatment patterns analysis, by database, are presented in Figure 1.

Table 2: Demographic characteristics by data source

	Overall	CCMC	Medco	GE	PM	MDCD
N	1,391,915	633,755	131,428	132,938	420,905	72,889
% of Total	100.0%	45.5%	9.4%	9.6%	30.2%	5.2%
Female (%)	67.7%	65.6%	63.6%	70.9%	65.0%	77.7%
Age (Mean):	39.2	39.7	41.7	42.1	40.2	34.8

Figure 1: Age category by data source, treatment pattern data extract



First-line treatments and treatment patterns

Overall, 17.4% of patients had a qualifying first-line antidepressant treatment; this ranged from 9.6% (Medco) to 29.4% (GE) (Table 3). The type (class) of first-line treatment was very similar across all databases, with SSRIs accounting for 72-75% of all first-line treatments; followed by Other antidepressants (12-17%), SNRIs (8-11%); and TCAs (1-3%) (Table 3). MAOIs represented .01% or less of first-line treatments in each database with too few first-line treatments in any database for meaningful comparison.

Discontinuation was the most common treatment pattern (62.5%), followed by Continuation (17.1%), Switched (12.3%), and Augmentation (8.1%) (Table 3). Overall patterns of discontinuation were consistent across commercial claims (i.e., CCMC, Medco and PM) and government claims (i.e., MDCD) databases (65-69%); whereas the rate of discontinuation estimated from EMR (GE) data was notably lower (45.7%) (Table 3 and Figure 2). The rate of Continuation varied by type of database: Government Claims (5%), Commercial Claims (14-15%), and EMR (32%). The Switching rate was consistent across all database types (12-14%). The Augmentation rate was also consistent across 4 of 5 databases (7-8%), with Government Claims being higher (13%).

Treatment days varied by database, with Government Claims (MDCD) exhibiting the shortest first-line treatment days (112 days [mean], 32 days [median]), and EMR (GE) the longest (414 days [mean], 205 days [median]) (Table 3). Treatment lengths among the Commercial Claims databases (i.e., CCMC, Medco and PM) were very similar with mean values between 193 and 214 days and median values between 85 and 88 days (Table 3). The similarities between commercial claims data, as well as the comparably higher treatment days when calculated from the EMR (GE) data and the lower treatment days when calculated from government claims (MDCD), are maintained when examined by antidepressant class (Figure 3).

The PDC followed a trend similar to that of treatment days with all commercial claims data having very similar PDCs (0.38-0.39), while the EMR (GE) data had the highest PDC (0.58) and the government claims (MDCD) with the lowest (0.26). SNRIs had a slightly higher overall PDC than all other classes, ranging from 0.30-0.53 across all databases (Figure 4). For all individual databases other than GE, the SNRI PDC was the highest of all antidepressant classes. TCA's had the lowest overall PDC, but exhibited a wider variation among databases (0.23-0.60). The TCA PDC was consistent for Claims (0.23-0.28) but significantly higher for EMR (0.60).

Table 3: Characteristics of treatment patterns by data source

	Overall	CCMC	Medco	GE	PM	MDCD
Qualifying 1st line treatment (%)	17.4%	14.2%	9.6%	29.4%	16.2%	16.5%
Qualifying 1st line treatment (n)	221,802	89,801	12,569	39,062	68,363	12,007
First Line Treatment (%)						
MAOI	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
TCA	1.6%	1.4%	1.9%	1.6%	1.5%	3.3%
SSRI	73.6%	72.9%	72.5%	74.9%	74.3%	72.1%
SNRI	9.9%	10.3%	10.4%	11.1%	8.8%	7.9%
Other	14.9%	15.5%	15.2%	12.4%	15.4%	16.7%
Treatment Pattern (%)						
Discontinued	62.5%	65.3%	65.5%	45.7%	66.7%	69.1%
Continued	17.1%	14.7%	14.9%	32.4%	14.1%	5.3%
Switched	12.3%	12.0%	12.2%	13.9%	11.7%	12.7%
Augmented	8.1%	8.0%	7.4%	8.0%	7.5%	12.9%
TX days (mean)	231	193	214	414	201	112
TX days (med)	104	88	88	205	85	32
PDC	0.41	0.39	0.39	0.58	0.38	0.26

Figure 2: Treatment patterns by data source

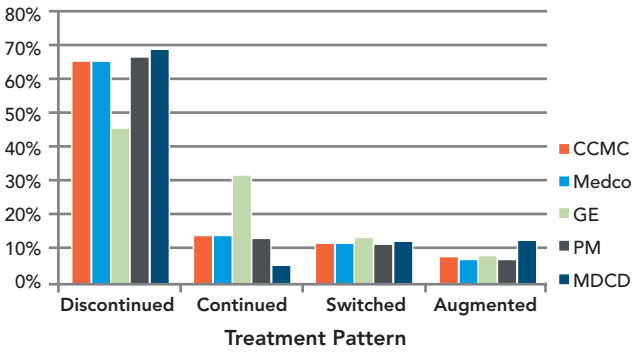


Figure 3: Median treatment days by antidepressant class and data source

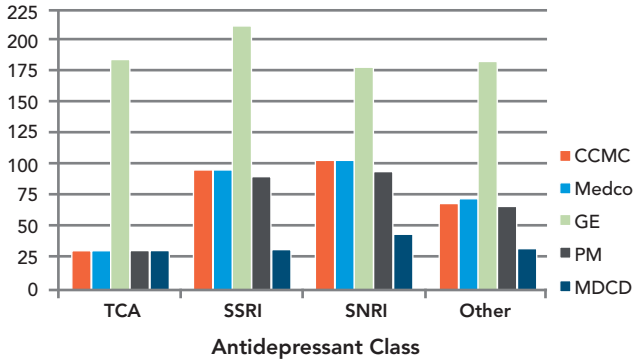
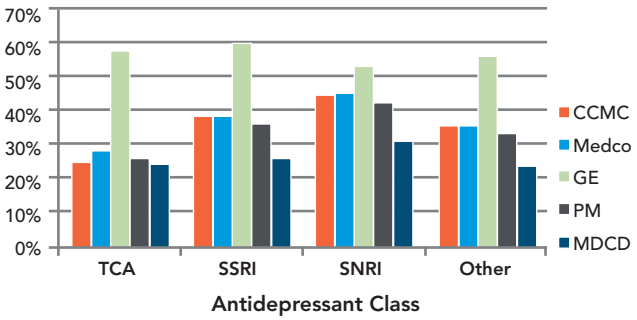


Figure 4: Proportion of days covered (PDC) by antidepressant class and data source



Discussion

The treatment patterns analysis was conducted on three sets of conceptually similar commercial claims data; therefore the consistent results across the CCMC, Medco and PM data were expected. The disparities seen in some of the Government Claims (MDCD) results (e.g., shorter treatment duration, less continued antidepressant use) may be the result of either population characteristics (e.g., MDCD were generally younger and may represent a subgroup that is less likely to comply with prescribed treatment) and/or different rules for medical reimbursement. The EMR database exhibited the most inconsistent results (more treated patients, longer treatments), likely reflecting fundamental differences in the underlying reason for data capture in this population (i.e., record of patient medical history as opposed to medical cost reimbursement). Despite these differences, the overall patterns of treatment across disparate databases and populations were strikingly similar, which may reflect the availability of American Psychiatric Association (APA) Treatment Guidelines for Patients with Major Depressive Disorders.[8]

These analyses are subject to common limitations in observational data. Commercial claims data, such as CCMC, Medco and PM are primarily used for administrative purposes, enabling healthcare providers to obtain reimbursement for services provided. As a result, issues such as diagnostic miscoding are possible. Government claims data (MDCD) also are predominately used for administrative purposes, but the populations serviced differ from those of commercial claims. In the EMR (GE) data, diagnostic miscoding, or the absence of diagnostic coding, is potentially greater as these data are not used for reimbursement purposes. Additionally, as it relates to the EMR (GE) data, only prescriptions written is available (whereas prescriptions filled is available in claims data) and days' supply is usually inferred based on National Drug Code (NDC) information. These factors likely lead to the differing treatment patterns observed in the GE data.

Despite these database limitations, we have provided an example of a collaborative analysis of treatment patterns in patients diagnosed with depression, conducted on five disparate observational databases. This research provides a relatively simple, yet applicable illustration of how standardized analytics provides an efficient way of enabling meaningful comparisons across disparate data sources. In addition to the demographic and treatment pattern analyses presented, this general approach can be applied to a variety of retrospective observational analyses (e.g., incidence estimation, health outcomes, drug safety/adverse events, burden of illness, etc.).

The potential benefits of CDM implementation, however, go well beyond individual analysis applications. For example, database epidemiology on rare diseases or orphan drugs is often hindered by inadequate sample size from any single retrospective data source. As such, there has historically been a heavy reliance on patient registries and/or the use of multiple retrospective data sources; both of which result in logistically complicated and costly projects. The ability to efficiently combine data from several disparate data sources using a standardized

format and vocabulary changes this as the CDM enables the easy implementation of either pooled or database stratified analyses (which as we demonstrate above may be necessary due to inherent differences in data capture processes and/or underlying population characteristics that are important for interpreting results produced from each database). Furthermore, the general concept of the CDM can be expanded to multiple, similar patient registries (i.e., multiple registries that focus on similar disease and have many conceptually common data fields), in essence enabling the creation of a "master" registry.

The ability to conduct these types of data processing and analyses tasks in a single, standardized manner will certainly minimize (and potentially eliminate) many of the historic limitations inherent in retrospective observational studies.

For more information, please contact Gary.Schneider@evidera.com, Stephanie.Reisinger@evidera.com, or Matthew.Reynolds@evidera.com.

REFERENCES

[1] OMOP Research Archive. Available at: http://omop.org/ResearchArchive. Accessed April 1, 2015.

[2] Murray RE, Ryan PB, Reisinger SJ. Design and Validation of a Data Simulation Model for Longitudinal Healthcare Data. *AMIA Annu Symp Proc.* 2011; 2011:1176-1185.

[3] OHDSI Collaborators. Available at: http:// http://www.ohdsi.org/who-we-are/collaborators. Accessed April 1, 2015.

[4] Reisinger S, Powell G, Dreyfus B, Schneider G. Use of Common Data Model To Meaningfully Compare Treatment Patterns for Depression Among Disparate Databases. Poster presented at the ISPOR 18th Annual International Meeting, New Orleans, LA, May 2013.

[5] Reisinger S, Powell G, Dreyfus B, Schneider G. Use of Common Data Model to Meaningfully Compare Patients Diagnosed and Treated for Depression Among Disparate Databases. Oral presentation at the 29th International Conference on Pharmacoepidemiology & Therapeutic Risk Management, Montreal, Canada, August 2013.

[6] Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, DeFalco FJ, et al. Feasibility and Utility of Applications of the Common Data Model to Multiple, Disparate Observational Health Databases. *JAMIA* 2015 Feb 10. Available at: http://dx.doi.org/10.1093/jamia/ocu023. Accessed April 3, 2015.

[7] OMOP CDM and Vocabularies Version 4. Available at: http://omop.org/CDM. Accessed April 1, 2015.

[8] American Psychiatric Association (APA). Practice Guideline for the Treatment of Patients with Major Depressive Disorder. 3rd ed. Arlington (VA): American Psychiatric Association (APA); 2010 Oct. 152 p.