OHDSI Treatment Pathways Protocol:

Exploration of initial treatment patterns in cancer patients with hypertension, type 2 diabetes mellitus, and depression

**Authors:**

George Hripcsak, MD, Columbia University

RuiJun Chen, MD, Columbia University

Thomas Falconer, MS, Columbia University

Patrick Ryan, PhD, Janssen Research and Columbia University

Gurvaneet Randhawa MD, MPH, National Cancer Institute

<<Other authors added who contribute results from their data and satisfy ICMJE guidelines>>

**Date:** December 18, 2017

**Acknowledgment:** The analysis is based in part on work from the Observational Health Sciences and Informatics collaborative. OHDSI (<http://ohdsi.org>) is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics.

The authors declare the following disclosures: Dr. Ryan is an employee of Janssen Research & Development.

Table of Contents

[Background 3](#_Toc405127685)

[Objective 3](#_Toc405127686)

[Data sources 3](#_Toc405127687)

[Population 6](#_Toc405127688)

[Hypertension 7](#_Toc405127689)

[Type 2 diabetes mellitus 8](#_Toc405127690)

[Depression 8](#_Toc405127691)

[Methods 9](#_Toc405127692)

[References 12](#_Toc405127693)

# Background

Patients with cancer represent a unique, vulnerable population who are also pre-disposed to developing a number of co-morbid conditions, which may often be chronic and require long-term treatment involving multiple therapies over time. While extensive research and clinical trials are often carried out for the neoplastic diseases themselves, fewer studies are performed on the chronic co-morbid conditions these patients may develop. Numerous treatment guidelines exist for chronic conditions, but it is unclear whether such guidelines are appropriate or followed for cancer patients and how real-world cancer patients are treated for such diseases.

Building on our recent characterization of treatment pathways for chronic disease, we aim to fill the gap of understanding the prevalence real-world pathways for the treatment of chronic co-morbid conditions in patients with cancer. Little is known about the specific sequence of treatments that patients follow upon initiation of therapy, particularly for this unique population. Understanding these pathways will establish context around questions of drug utilization, effectiveness, and adherence in this uniquely vulnerable population, and also demonstrate the feasibility of carrying out large-scale network studies in patients with cancer. The OHDSI data network will be used to conduct a systematic analysis of three chronic diseases in patients with cancer: hypertension, type 2 diabetes mellitus, and depression, to summarize the treatment pathways encountered by these cancer patients during the first 12 months of therapy.

# Objective

Among patients with cancer, characterize the prevalence, variety, and distribution of different initial treatment pathways for 3 chronic diseases: hypertension, type II diabetes, and depression. We will systematically summarize the treatment pathways observed among patients who have a diagnosis of cancer and at least 12 months of continuous observation and persistent treatment following initiation of the first treatment for one of the three conditions. We will stratify the results by year to evaluate temporal trends, and will further stratify by data-source to determine if treatment pathways vary by population, geography, and data capture process.

# Data sources

The analyses will be performed across a network of observational healthcare databases. All databases have been transformed into the OMOP Common Data Model, version 5. The complete specification for OMOP Common Data Model, version 5 is available at: <https://github.com/OHDSI/CommonDataModel>. The following databases will be included in this analysis:

* Columbia University / New York Presbyterian (NYP)
* Truven MarketScan Commercial Claims and Encounters (CCAE)
* Truven MarketScan Medicare Supplemental Beneficiaries (MDCR)
* Truven MarketScan Multi-state Medicaid (MDCD)
* Optum ClinFormatics (Optum)
* Clinical Practice Research Datalink (CPRD)
* Regenstrief Institute / Indiana Network for Patient Care (INPC)
* Stanford University (STRIDE)
* <<add others who agree to participate>>

Truven MarketScan Commercial Claims and Encounters (CCAE)

CCAE is an administrative health claims database for active employees, early retirees, COBRA continues, and their dependents insured by employer-sponsored plans (individuals in plans or product lines with fee-for-service plans and fully capitated or partially capitated plans). As of 30November2014, CCAE contained 117m patients with patient-level observations from Jan2000 through Jul2014. Source codes used in CCAE include: conditions- ICD-9-CM; drugs: NDC, HCPCS, ICD-9-CM; procedures: CPT-4, HCPCS, ICD-9-CM; lab: LOINC.

The ETL specification for transforming CCAE into the OMOP CDM is available at: <http://omop.org/cdm>.

ACHILLES has been used to characterize the database and provide a data quality assessment. The ACHILLES summary is available internally within Janssen at: <http://hix.jnj.com/achilles/#/truven_ccae/dashboard>.

Truven MarketScan Medicare Supplemental Beneficiaries (MDCR)

MDCR is an administrative health claims database for Medicare-eligible active and retired employees and their Medicare-eligible dependents from employer-sponsored supplemental plans (predominantly fee-for-service plans). Only plans where both the Medicare-paid amounts and the employer-paid amounts were available and evident on the claims were selected for this database.. As of 30November2014, MDCR contained 9m patients with patient-level observations from Jan2000 through Jul2014. Source codes used in MDCR include: conditions- ICD-9-CM; drugs: NDC, HCPCS, ICD-9-CM; procedures: CPT-4, HCPCS, ICD-9-CM; lab: LOINC.

The ETL specification for transforming MDCR into the OMOP CDM is available at: <http://omop.org/cdm>.

ACHILLES has been used to characterize the database and provide a data quality assessment. The ACHILLES summary is available internally within Janssen at: <http://hix.jnj.com/achilles/#/truven_mdcr/dashboard>.

Truven MarketScan Multi-state Medicaid (MDCD)

MDCD is an administrative health claims database for the pooled healthcare experience of Medicaid enrollees from multiple states. As of 30November2014, MDCD contained 16m patients with patient-level observations from Jan2006 through Dec2012. Source codes used in MDCD include: conditions- ICD-9-CM; drugs: NDC, HCPCS, ICD-9-CM; procedures: CPT-4, HCPCS, ICD-9-CM; lab: LOINC.

The ETL specification for transforming MDCD into the OMOP CDM is available at: <http://omop.org/cdm>.

ACHILLES has been used to characterize the database and provide a data quality assessment. The ACHILLES summary is available internally within Janssen at: <http://hix.jnj.com/achilles/#/truven_mdcd/dashboard>.

Optum ClinFormatics (Optum)

Optum is an administrative health claims database for members of United Healthcare, who enrolled in commercial plans (including ASO, 36.31M), Medicaid (prior to July 2010, 1.25M) and Legacy Medicare Choice (prior to January 2006, 0.36M) with both medical and prescription drug coverage. As of 30November2014, Optum contained 38m patients with patient-level observations from Oct2005 through Dec2013. Source codes used in Optum include: conditions- ICD-9-CM; drugs: NDC, HCPCS, ICD-9-CM; procedures: CPT-4, HCPCS, ICD-9-CM; lab: LOINC.

The ETL specification for transforming Optum into the OMOP CDM is available at: <http://omop.org/cdm>.

ACHILLES has been used to characterize the database and provide a data quality assessment. The ACHILLES summary is available internally within Janssen at: <http://hix.jnj.com/achilles/#/optum/dashboard>.

Clinical Practice Research Datalink (CPRD)

CPRD is an anonymized longitudinal electronic health records from primary care practices in UK. Patient management system with many aspects of patient care covered, including diagnoses, prescriptions, signs and symptoms, procedures, labs, lifestyle factors, clinical and administrative/social data. As of 30November2014, CPRD contained 11m patients with patient-level observations from Jan1988 through Nov2013. Source codes used in CPRD include: conditions- Read; drugs: Multilex; procedures: OPCS.

The ETL specification for transforming CPRD into the OMOP CDM is available at: <http://omop.org/cdm>.

ACHILLES has been used to characterize the database and provide a data quality assessment. The ACHILLES summary is available internally within Janssen at: <http://hix.jnj.com/achilles/#/cprd/dashboard>.

Indiana Network for Patient Care (INPC)

The INPC is a clinical dataset derived from the Indiana Health Information exchange, incorporating a variety of sources including hospitals, outpatient clinics, freestanding radiology and laboratory centers, public health organizations, and drug dispensing data from SureScripts As of 15 January 2014, INPC CDM contained 11.9m patients with patient-level observations from Jan2004 through Dec2014.

The ETL specification for transforming INPC into the OMOP CDM is available at: <http://www.ohdsi.org/web/wiki/doku.php?id=documentation:example_etls>

*Database X*

*Database X description*

The ETL specification for transforming *Database X* into the OMOP CDM is available at: *ETL\_specification\_URL*

ACHILLES has been used to characterize the database and provide a data quality assessment. The ACHILLES summary is available at: *URL to ACHILLES*.

# Population

The general approach to defining the population of interest of cancer patients with a chronic disease who have persistent exposure for at least 12 months following treatment initiation is illustrated in Figure 1:

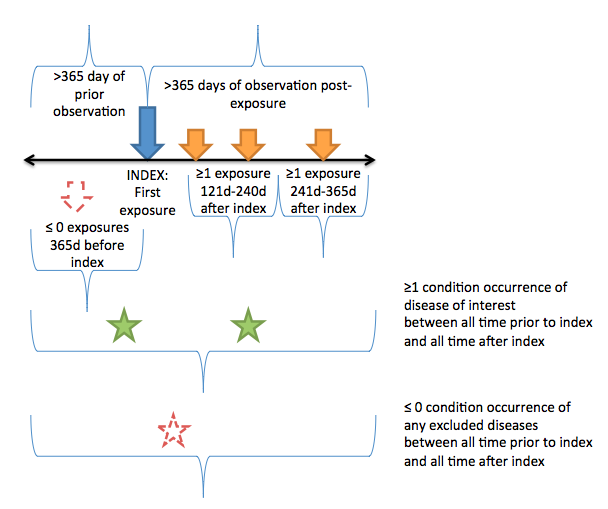


Figure 1: Treatment pathway population criteria

## Hypertension

The analysis will be performed in a cohort of patients who satisfy the following criteria:

Index rule defining the index date:

* First exposure to an antihypertensive treatment with at least 365 days of observation prior to exposure and at least 365 days of observation after the exposure

Inclusion rules based on the index date:

* At least one condition occurrence of neoplastic disease, excluding descendent conditions of benign neoplastic disease or lipoma, between all days before index date and all days after index date (based on SNOMED codes: 55342001-Neoplastic Disease; exclude 20376005-Benign neoplastic disease; 254827004-Lipomatous tumor)
* At least one condition occurrence of hypertension between all days before index date and all days after index date
* At most zero condition occurrences of pregnancy between all days before index date and all days after index date
* At most zero drug exposures of antihypertensive treatment between all days before index date and 1 day before index date
* At least one drug exposure of antihypertensive treatment between 121 days after index date and 240 days after index date
* At least one drug exposure of antihypertensive treatment between 241 days after index date and 365 days after index date

## Type 2 diabetes mellitus

The analysis will be performed in a cohort of patients who satisfy the following criteria:

Index rule defining the index date:

* First exposure to an antihyperglycemia treatment with at least 365 days of observation prior to exposure and at least 365 days of observation after the exposure

Inclusion rules based on the index date:

* At least one condition occurrence of neoplastic disease, excluding descendent conditions of benign neoplastic disease or lipoma, between all days before index date and all days after index date (based on SNOMED codes: 55342001-Neoplastic Disease; exclude 20376005-Benign neoplastic disease; 254827004-Lipomatous tumor)
* At least one condition occurrence of Type 2 diabetes mellitus between all days before index date and all days after index date
* At most zero condition occurrences of pregnancy between all days before index date and all days after index date
* At most zero condition occurrences of Type 1 diabetes mellitus between all days before index date and all days after index date
* At most zero drug exposures of antihyperglycemia treatment between all days before index date and 1 day before index date
* At least one drug exposure of antihyperglycemia treatment between 121 days after index date and 240 days after index date
* At least one drug exposure of antihyperglycemia treatment between 241 days after index date and 365 days after index date

## Depression

The analysis will be performed in a cohort of patients who satisfy the following criteria:

Index rule defining the index date:

* First exposure to an antidepressant treatment with at least 365 days of observation prior to exposure and at least 365 days of observation after the exposure

Inclusion rules based on the index date:

* At least one condition occurrence of neoplastic disease, excluding descendent conditions of benign neoplastic disease or lipoma, between all days before index date and all days after index date (based on SNOMED codes: 55342001-Neoplastic Disease; exclude 20376005-Benign neoplastic disease; 254827004-Lipomatous tumor)
* At least one condition occurrence of Type 2 diabetes mellitus between all days before index date and all days after index date
* At most zero condition occurrences of pregnancy between all days before index date and all days after index date
* At most zero condition occurrences of bipolar disorder or schizophrenia between all days before index date and all days after index date
* At most zero drug exposures of antidepressant treatment between all days before index date and 1 day before index date
* At least one drug exposure of antidepressant treatment between 121 days after index date and 240 days after index date
* At least one drug exposure of antidepressant treatment between 241 days after index date and 365 days after index date

Concepts and associated descendant source codes for each concept used to define the population are provided in Appendix 1: Population concepts.



# Methods

For each disease, we will create a cohort meeting the population inclusion criteria, as detailed above. For each patient in the qualifying cohorts, we will identify the sequence of treatments exposed during the 12 months after first exposure. Treatments will be summarized at the RxNorm ingredient level, and the DRUG\_ERA table will be used to identify exposures. The sequence will be determined by ordering the dates of first exposure to each qualifying ingredient for the disease. All patients will have at least one treatment (since to qualify for the cohort, they had at least one first exposure). If a person was maintained on the same treatment for the entire 12 month period, meaning all drug eras during the 12 month window were for the same ingredient and no other ingredient for the same disease was observed during that same interval, then the person would be classified as having a 1-drug treatment sequence. A person who switched between treatments only once during their interval, would result in a 2-drug treatment sequence. We summarize all sequence combinations that are less than 20 drugs. Note, this definition of treatment pathway does not characterize patients who may switch back to a treatment used previously (that is, an ingredient will only be listed once within a sequence, since we only use the first exposure to each ingredient). Also, this treatment pathway will not distinguish between switching vs. augmentation behavior, as the sequence of ingredients does not summarize whether the ingredients are being used concomitantly.

Once each patient’s treatment sequence is constructed, we will count the number of unique persons with the same treatment sequence. The counts will be produced within each source for each disease. We will also stratify the counts by the index year of the first exposure.

Data partners will generally supply sequence counts down to 1 to the coordinating center, and this will be used for analysis but will NOT be shared publically. Any release of results as part of publication with truncate sequences such that all counts are at least 5. Nevertheless, the analysis code provides each data partner the option to suppress any summary statistics below a minimum cell count number. The default is currently 1 (no suppression).

The table shells for the four results tables of summary statistics are provided below with illustrative results generated during the initial feasibility assessment.

These summary results will then be used to create tabular and graphical summaries of the evidence across the OHDSI data network to characterize the prevalence of treatment pathways by source and year. As a descriptive summary analysis with no pre-specified hypotheses, no statistical tests will be performed. All results will be placed in the public domain and a summary report will be generated and submitted for publication upon completion.

Table 1: Number of persons, by source and disease

|  |  |  |
| --- | --- | --- |
| **SOURCE** | **Disease** | **num\_persons** |
| CCAE | Depression |  |
| MDCD | Depression |  |
| MDCR | Depression |  |
| OPTUM | Depression |  |
| CCAE | HTN |  |
| MDCD | HTN |  |
| MDCR | HTN |  |
| OPTUM | HTN |  |
| CCAE | T2DM |  |
| MDCD | T2DM |  |
| MDCR | T2DM |  |
| OPTUM | T2DM |  |

Table 2: Number of persons, by source and disease by year

|  |  |  |  |
| --- | --- | --- | --- |
| **SOURCE** | **Disease** | **index\_year** | **num\_persons** |
| CCAE | Depression | 2000 |  |
| CCAE | Depression | 2001 |  |
| CCAE | Depression | 2002 |  |
| CCAE | Depression | 2003 |  |
| CCAE | Depression | 2004 |  |
| CCAE | Depression | 2005 |  |
| CCAE | Depression | 2006 |  |
| CCAE | Depression | 2007 |  |
| CCAE | Depression | 2008 |  |
| CCAE | Depression | 2009 |  |
| CCAE | Depression | 2010 |  |
| CCAE | Depression | 2011 |  |

Table 3: Treatment pathway summary, by source and disease

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **SOURCE** | **Disease** | **d1\_concept\_id** | **d2\_concept\_id** | **d1\_concept\_name** | **d2\_concept\_name** | **num\_persons** |
| CCAE | Depression | 739138 | NULL | Sertraline | NULL |  |
| CCAE | Depression | 715939 | NULL | Escitalopram | NULL |  |
| CCAE | Depression | 797617 | NULL | Citalopram | NULL |  |
| CCAE | Depression | 755695 | NULL | Fluoxetine | NULL |  |
| CCAE | Depression | 750982 | NULL | Bupropion | NULL |  |
| CCAE | Depression | 743670 | NULL | venlafaxine | NULL |  |
| CCAE | Depression | 722031 | NULL | Paroxetine | NULL |  |
| CCAE | Depression | 715939 | 797617 | Escitalopram | Citalopram |  |
| CCAE | Depression | 715939 | 750982 | Escitalopram | Bupropion |  |
| CCAE | Depression | 715259 | NULL | duloxetine | NULL |  |
| CCAE | Depression | 739138 | 750982 | Sertraline | Bupropion |  |
| CCAE | Depression | 797617 | 750982 | Citalopram | Bupropion |  |

Table 4: Treatment pathway summary, by source and disease, by year

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **SOURCE** | **Disease** | **index\_year** | **d1\_concept\_id** | **d2\_concept\_id** | **d1\_concept\_name** | **d2\_concept\_name** | **num\_persons** |
| CCAE | Depression | 2001 | 739138 | NULL | Sertraline | NULL |  |
| CCAE | Depression | 2001 | 722031 | NULL | Paroxetine | NULL |  |
| CCAE | Depression | 2001 | 797617 | NULL | Citalopram | NULL |  |
| CCAE | Depression | 2001 | 755695 | NULL | Fluoxetine | NULL |  |
| CCAE | Depression | 2001 | 743670 | NULL | venlafaxine | NULL |  |
| CCAE | Depression | 2001 | 750982 | NULL | Bupropion | NULL |  |
| CCAE | Depression | 2001 | 739138 | 750982 | Sertraline | Bupropion |  |
| CCAE | Depression | 2001 | 722031 | 750982 | Paroxetine | Bupropion |  |
| CCAE | Depression | 2001 | 755695 | 750982 | Fluoxetine | Bupropion |  |
| CCAE | Depression | 2001 | 797617 | 715939 | Citalopram | Escitalopram |  |

# References

<<contributions welcome around treatment guidelines or any available literature on treatment pathways that should be used when we write up our findings>>