Observational Study Protocol

Study Code	OBS-19-00690
Version	1.0
Date	10 July 2019

Exploration of patients with bladder cancer using the OHDSI data network and the OMOP Oncology CDM module

Sponsor: AstraZeneca

Author:

TABLE OF CONTENTS PAGE

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation or Special Term	Explanation
CCI	Charlson Comorbidity Index
CDM	Common data model
CI	confidence interval
CTLA-4	cytotoxic t lymphocyte-associated antigen 4
ECOG	Eastern Cooperative Oncology Group
EHR	Electronic health record
EMR	Electronic medical record
EMA	European Medicines Agency
GPP	good pharmacoepidemiology practice
IO	immuno-oncology
MAH	Market authorization holder
OHDSI	Observational Health Data Science and Informatics
OMOP	Observational Medical Outcomes Partnership
OS	overall survival
PD-1	programmed cell death 1
PD-L1	programmed death ligand 1
PFS	progression-free survival
SAP	statistical analysis plan
TTD	time to treatment discontinuation
US	United States

RESPONSIBLE PARTIES

Name	Professional Title	Role in Study	Affiliation	Email Address
Christian Reich	VP, Real World Analytics Solution, IQVIA		IQVIA	Christian.reich@iqvia.com
Asieh Golozar	Director, Epidemiology	Epidemiologist	OBU-Epidemiology and RWE	asieh.golozar@astrazeneca.com
MakanSarkeshik	Medical Affairs Lead		OBU Medical	Makan.sarkeshik@astrazeneca.com

PROTOCOL SYNOPSIS

Exploration of patients with bladder cancer using the OHDSI data network and the OMOP Oncology CDM module

Background/Rationale

This retrospective study aims to describe characteristics of patients, treatments patterns,and clinical outcomes in patients with advanced bladder cancer with and without liver metastasis.

Objectives

To utilize the standardized cancer data in OMOP CDM Oncology module to describe patient demographic and clinical characteristics, real-world treatment patterns(type of therapy, duration, lines of therapy, treatment sequencing) and clinical outcomesof patients diagnosed withadvanced bladder cancer with and without liver metastasis.

Methods

Study design: Retrospective observational cohort study

Data Source(s): At least five data sources participating in the OHDSI Network will be utilized for this study: Electronic health record (EHR) and registry data in OMOP CDM from Northwestern University, Memorial Sloan Kettering Cancer Center, Tufts University, Columbia University, IQVIA Oncology electronic medical record (EMR*).*

Study Population: Patients withadvanced bladder cancer in any of the participating centers.

Outcome(s):

Primary outcomes assessed following the date of initiation of first-line therapy and by each subsequent line of therapy will be:

- Overall survival(OS)
- Progression-free survival (PFS)

Other outcomes of interest following the date of initiation of first-line therapyinclude:

- First line treatment regimen distribution
- Time to treatment discontinuation (TTD)

- Time to secondline of treatment
- Second line treatment regimen distribution

Sample Size Estimations

Like with all observational studies, sample size is a readout, not a requirement. For estimation studies the sample size will determine whether or not the confidence interval of the resulting point estimate will include 1 (meaning, the Null hypothesis can't be rejected), but no such test is anticipated in this study.

Statistical Analysis

Demographics, clinical characteristics, and general treatment characteristics will be summarized using descriptive statistics. Time to event data will be evaluated by Kaplan-Meier curves (including a tabulation of the percentage of patients alive at pre-specified time points), together with two-sided 95% confidence intervals (CIs) for the median survival estimates.

Detailed treatment patterns will be summarized by line of therapy using similar approaches to thosedescribed above.
Full details on the statistical analyses will be documented in the Statistical Analysis Plan.

AMENDMENT HISTORY

Date	Section of study protocol	Amendment or update	Reason
<<>>		<<>>	
		N/A	

MILESTONES

Milestone	Planned date
Kickoff	1 August 2019
Feasibility of study based on available patients in OMOP CDM Oncology Module	30 August 2019
Queries developed and tested on IQVIA database	15 September 2019
Queries conducted	30 September 2019
Summary results delivered	30 October 2019

1. BACKGROUND AND RATIONALE

Urothelial carcinoma (UC), which accounts for >90% of all bladder cancers (Fleshner et al, 1996), is one of the 10 predominant malignancies worldwide (Powles, 2015).Despite recent improvements in clinical outcomes of patients with advanced bladder cancer following first- or second-line treatment with immune checkpoint inhibitors, little is known about their epidemiological characteristics of patients with bladder cancer who have liver metastasis at the time of diagnosis or develop liver metastasis during the course of disease. Sufficient data on this patient population are hard to obtain.

The traditional approach to generating real world evidence (RWE) requires the following steps:

- Procurement of a commercially available databases
- Extraction and loading of the entire database to an in-house compute environment
- Analysis by in-house team of statistical programmers and epidemiologists

In cancer, this approach is particularly hard to implement for a number of reasons:

1. Databases are chronically difficult to obtain because cancers are typically treated in cancer centers which rarely make data available commercially. In addition, many countries create compliance hurdles against that approach.
2. Databases are hard to assess whether they are fit for purpose before proper feasibility assessment with respect to the question at hand
3. Data use agreements are very difficult and time consuming to draw up
4. The quality of the aggregated data and their sampling approach is opaque.
5. Significant resources are required for the technical infrastructure and team.

To address these challenges, IQVIA adopted the open, systematic and standardized approach of the Observational Health Data Science and Informatics collaborative (OHDSI). OHDSI provides Open Source resources to the community: the (observational medical outcomes partnership) OMOP Common Data Model (CDM), the OMOP Standardized Vocabularies, and a number of tools and methods to interrogate and analyze the data.

This allows to generate RWE through the OHDSI Research Network, whereby each participating institution retains their data, but makes them available for querying and statistical evaluation. All data are standardized to a common data model (format) and reference data (vocabularies). That allows for queries and studies to be developed with no access to the data where they eventually are executed. IQVIA as the coordinating center establishes data quality certification and adherence to conventions.

Typical observational studies base their cohort and outcome definitions on diagnostic codes, sometimes in combination with drug treatment, procedure occurrence or lab tests.Malignant neoplasms tend to have significantly more diagnostic make up: anatomical site including

local penetration, morphology, metastatic spread, affected lymph nodes, mechanical consequences of tumor growth and biomarkers. All of them together define the disease and have therapeutic consequences. However, typical observational data sources do not convey the necessary level of detail, and if they do, they collect individual results of diagnostic procedures but do not summarize to the above attributes (abstraction). In addition, cancer treatment tends to also be more complex compared to other disease modalities. Cancers are treated with chemotherapy regimens, often combined with targeted therapies or immunotherapies, surgery or radiotherapy. None of these attributes follow standard definitions that can be easily applied to retrospective observational data.

The OHDSI Initiative formed a Working Group to address these issues. It standardized the topology, morphology, spreading and other tumor markers, therapies and abstractions. It has the potential to allow distributed research in a network of databases all describing the patient populations in a standardized way without the need to extract the data, aggregate and interpret them at analysis time. If feasible, the OHDSI network has potentially a large number of patients available globally for systematic and standardized research.

1.1 Rationale

Durvalumab, an anti-PD-L1 monoclonal antibody developed by AstraZeneca/MedImmune, is currently being evaluated for the treatment of urothelial cancer alone or in combination with tremelimumab, a human immunoglobulin antibody against the cytotoxic T lymphocyte-associated antigen-4 (CTLA-4). To better understand unmet treatment needs, to identify special patient populations who might benefit from new treatments, and to collect baseline data that can help contextualize the impact of future novel therapies in specific patient population, characterizing the current treatment landscape of bladder cancer patients with and without liver metastasis in real-world setting is specifically needed.

This study aims at testing the feasibility of OHDSIapproach in generating validreal-world evidence on cancer in a network of databases.

2. OBJECTIVES AND HYPOTHESES

The aim of this study is to assess the feasibility of OHDSI approach in generating RWE on cancer. Identification of bladder cancer patients with and without liver metastasis and describing their characteristics, including age, sex, and race, comorbidities. their treatment pattern, and clinical outcomes in the real-world will be served as a case example. Anonymized, pre-existing data from a real-world datasets of patients with bladder cancer will be used to address the following primary objectives:

2.1 Primary Objectives

1. To assess the feasibility of OHDSI approach in generating valid oncology-related evidence from retrospective observational data. This will be reflected in the attrition of available outcome for the bladder cancer cohort.

2.2 Secondary Objectives

1. To describe demographic and clinical characteristics of patients diagnosed with bladder cancer with and without liver metastasis
2. To characterize detailed treatment patterns by line of therapy among patients with bladder cancer with and without liver metastasis:
   - Specific agents received
   - Time to treatment discontinuation (TTD) (duration of treatment)
   - Treatment received by PD-L1 status
3. To estimate overall survival (OS) following each line of therapy among patients with advanced bladder cancer with and without liver metastasis
4. To estimate progression free survival (PFS) among patients with advanced bladder cancer with and without liver metastasis

2.3 Study Subgroups

3. METHODOLOGY

3.1 Study Design – General Aspects

The study design is a retrospective cohort study of patients diagnosed with bladder cancer with and without metastasis.

Patients will be followed longitudinally until death or end of database unless lost to follow up (censored).

The following standard OMOP concepts are used for readouts:

- Gender Concepts
- Medication and roll-up to Ingredients and ATC drug classes
- Hierarchical SNOMED Conditions

OMOP Oncology Episodes (Primary disease diagnosis, Treatment and Remission, stable disease, survival calculations) will be calculated from the cancer cohort and, if available, stratified by relevant tumor attributes.

3.1.1 Data Source(s)

This study will be conducted in data sources owned by IQVIA, which have previously been converted to the OMOP CDM Oncology Module. The table below lists all these data sources.

Data source	Description
Northwestern University	All data assets are based on Cancer Center EHR in combination with SEER cancer registry reporting, subsequently converted to the OMOP CDM.
Memorial Sloan Kettering Cancer Center	
Tufts University	
Columbia University	
IQVIA Oncology EMR	

3.2 Data management

Prior to this study, the original data sources were converted to the OMOP CDM through a process known as ETL (extraction, transformation, loading). This process defines the rules for how source codes are mapped to standardized concepts in OMOP while maintaining quality of information. The data is refreshed periodically responding to updates by the data source.

The OMOP process maps source codes to a standardized set of concepts depending on their domain. There are five key standardized domains: condition, drug, measurement procedure and observation. Source codes under the condition and procedure domain map to SNOMED vocabulary. Source codes under the drug domain map to RxNorm vocabulary. Source codes under the measurement domain map to LOINC vocabulary. Source codes under the observation domain will map to either SNOMED, RxNorm or LOINC, depending on the context. Cancer attributes are encoded using concepts form ICD-O-3 (topology, morphology), NAACCR (other tumor attributes) and HemOnc (regimens).

Data sources participating in our network study will not provide ETL information, however these data sources have been validated to ensure that the OMOP data set matches the quality of information provided in the original source.

3.3 Data analysis

Queries are going to be developed by IQVIA and tested at the Oncology EMR database. They are then distributed to the participating sites, executed there and the results returned to IQVIA. IQVIA will then generate summary reports.

3.4 Study Population

Patients diagnosed with advancedbladder cancer with and without liver metastasis.

3.4.1 Study time frame

All available patients are considered for inclusion

3.5 Inclusion Criteria

Two patient populations will be identified:

a) Bladder cancer cohort:

Patients with primary diagnosis of bladder cancer.

b) Bladder cancer with liver metastasis cohort:

Patients with diagnosis of primary bladder cancer who have liver metastases as readout from imaging or pathology report or patients with diagnosis of primary bladder cancer with liver metastases.

Cohort a) will capture all patients, with and without metastasis and not yet diagnosed patients. Patient populations strictly without metastasis are more difficult to ascertain if the diagnostic workup has not established that (presence information is more robust than absence information). Patients in both cohorts should be 18 years or older at initial diagnosis.

3.6 Exclusion Criteria

NA

3.7 Participant Follow-up

Patients will be followed from the date of diagnosis of their bladder cancer to the end of follow-up (date of death for patients who have died, and date of last visit prior to data cut-off for those still living at the end of the study data). All treatments received during this follow-up window, and clinical outcomes will be assessed.

4. VARIABLES AND EPIDEMIOLOGICAL MEASUREMENTS

4.1 Exposures

4.2 Outcomes

4.2.1 Overall Survival (OS)

4.2.2 Progression-free survival (PFS)

PFSfor each patient will be defined as the time (in days)untilthe earliest record of actual disease progression, or death from any cause, starting from the index date of the first, second or subsequent lines of treatment.

4.2.3 Other Outcomes

4.2.3.1 Time to Treatment Discontinuation (TTD)

TTD for each patient will be defined as the time (in days) from the date of initiation through the end of treatment within a line of therapy, or death from any cause.

4.2.3.2 Time to Second-Line Treatment

Time to second-linetreatment will be defined for each patient as the time (in days) following initiation of first-linetreatment until the start date of the second-line therapy(patients who die without a subsequent line will have their data censored at death).

4.2.3.3 Treatment related information

Specific details on systemic treatment will be captured by treatment line, restricting to the start date through end date of each line of therapy. Therapy lines will be referred to as first-line, second-line, and later.

Detailed information on the available outcomes will be available after feasibility assessment.

4.3 Other Variables and Covariates

Patient characteristicsincluding age at diagnosis, sex, stage of cancer at diagnosis and comorbidities will be assessed from baseline period (up to the 12-month period prior to and including the index date) from a summary of all attributes. If multiple values are present in the data during this period for a given variable, the value closest to the index date will be used.

5. STATISTICAL ANALYSIS PLAN

5.1 Statistical Methods – General Aspects

A global statistical analysis plan (SAP) will be developed to describe variable definitions for cohort selection, exposures, outcomes, covariates, and subgroups of interest. All analytic methods will be detailed, and a full set of table shells will be included. The SAP will be finalized before database lock.

Data management and analyses will be conducted using SQL and R.

Each of the variables defined above will be summarized descriptively for the overall study cohort. Analyses of treatment patterns and outcomes will also present results separately by line of therapy.The primary objectives of the study focus on descriptive analyses.

5.1.1 Primary Objective(s):Calculation of Epidemiological Measure(s) of Interest

Demographics, clinical characteristics, and general treatment characteristics will be summarized descriptively. Continuous study measures (e.g., age, body mass index) will be reported descriptively with mean, standard deviation, median, minimum, and maximum; when appropriate, these measures will also be presented in categories. Frequencies and percentages will be reported for categorical measures (e.g., number and proportion of patients with biomarker testing).

Detailed treatment patterns will be summarized by line of therapy. Time-to-event outcomes such asOS and PFSwill be evaluated descriptively by Kaplan-Meier curves (including a tabulation of the percentage of patients alive at pre-specified time points), together with two-sided 95% CIs for the median survival estimates.

5.1.2 Secondary Objective(s):Calculation of Epidemiological Measure(s) of Interest

Time to discontinuation, and time to next line of therapy will be evaluated descriptively by Kaplan-Meier curves, together with two-sided 95% CIs for the median survival estimates. When between-group comparisons are made, regression techniques that are appropriate for the form of the data will be used.

5.1.3 Potential Objective(s):Calculation of Epidemiological Measure(s) of Interest

N/A

5.2 Bias

5.2.1 Methods to Minimize Bias

The present study will be carried out using data recorded in a collection of EHR systems. The patient records are expected to be incomplete in many respects and may have erroneous entries. Treatment provided in hospitals or any other setting outside of the oncology clinic will not be seen;this may result in misclassification of clinical event types, such as treatment and outcomes. Comorbidities that can influence treatment given in the oncology clinic but that are themselves treated outside of the oncology setting are tend to be somewhat under-reported in the database, preventing adequate adjustment for such factors in the analyses. However, the primary purpose of the study is to describe a real-world cohort of patients with advanced bladder cancer in the institutions participating in the study, and the results should be used with caution when generalizing to the overall population.

5.2.2 Adjustment for Multiple Comparisons

Because the present study is descriptive in nature, and all between-group comparisons can be considered exploratory, no adjustment will be made for multiple comparisons.

5.2.3 Strengths and Limitations

The OMOP CDM Oncology Module is available in V1 and being tested for real-world use cases. This study is one of these tests. Therefore, we expect this process to be iterative: If issues are found the underlying cause (data conversion, vocabularies, definitions) is fixed, the data re-freshed (if appropriate) and the query re-run.

Depending on the nature of the source population in the database, our analysis is going to hold only for the studied cohort. The results cannot be extrapolated to the entire population suffering from the disease without to bias such as selection bias. This is because the data sources might not be fully representative for the disease. Acknowledging for this limitation, we believe the network study approach, particularly with OHDSI, allows us to put the results of each individual data source in context with one another and increases the generalizability of the findings. The OHDSI component allows us to evaluate this diverse array of healthcare data sources systematically.

5.3 Sample Size and Power Calculations

Like with all OHDSI Network Studies, study size and feasibility is established at part of the study. All patients who meet the inclusion and exclusion criteria during the study period will be in included in the study and no *a priori* power analyses will be conducted.

6. STUDY CONDUCT AND REGULATORY DETAILS

6.1 Study Conduct

6.2 Protection of Human Subjects

This study will be conducted in accordance with good pharmacoepidemiology practice (GPP). In this investigation we will use a medical record linkage database where the information of patients is anonymized and there is no need to obtain informed consent from patients. We will comply with all applicable data protection, security and privacy laws, rules and regulations with respect to the collection, production, use, processing, storage, transfer, modification, deletion, and/or disclosure of any information related to this study under this Agreement.

6.3 Collection and Reporting of Adverse Events/Adverse Drug Reactions

No reporting of adverse event data is requiredof studies using secondary data.

7. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

The results of this observational study are intended to be published in a peer-reviewed journal and as abstracts/presentations at medical congresses under the oversight of the market authorization holder (MAH).

8. LIST OF REFERENCES

9. APPENDICES

10. ATTACHMENTS

11. SIGNATURES

ASTRAZENECA SIGNATURE(S)

<<Study Description>>

<<This Observational Study Protocol >><<has/have>> been subjected to an internal AstraZeneca review>>
I agree to the terms of this Study protocol.

AstraZeneca representative

<<Name, title>>

Date
(Day Month Year)

<<Email address and telephone number>>