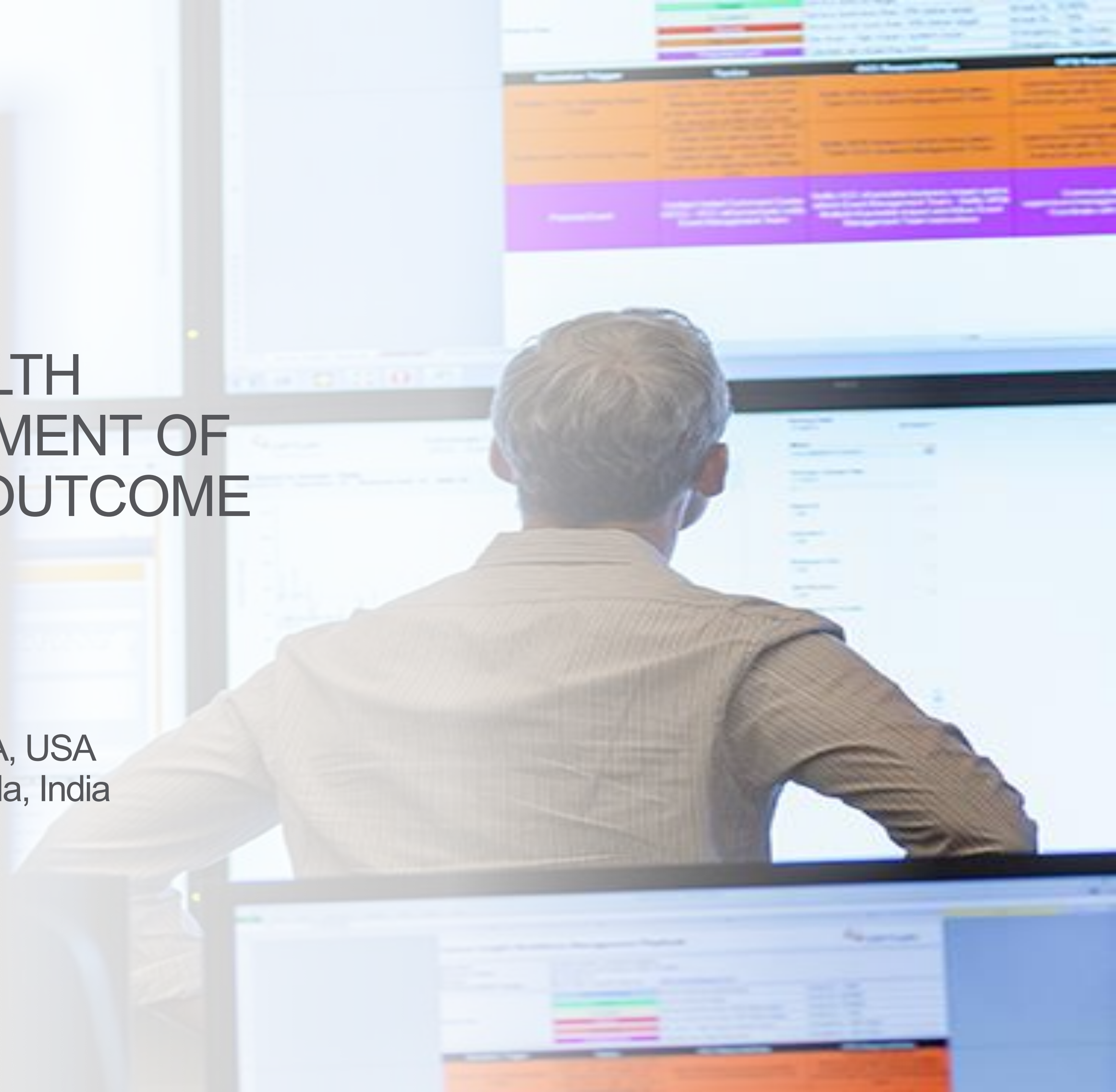
ROLE OF ELECTRONIC HEALTH RECORDS IN THE ADVANCEMENT OF REAL WORLD DATA BASED OUTCOME RESEARCH IN ONCOLOGY

Argentinis, E[1]; Mazumder D[2]; Dhangar I[2]

[1]Optum, Advanced Data and Analytics, Boston, MA, USA
[2]Optum Global solutions, Advanced Analytics, Noida, India



OPTUM®

Introduction

- Following the 21st Century Cures Act, Real-world data (RWD) and real-world evidence (RWE), have evolved into a complementary source of evidence for traditional clinical trials.[1]
- Electronic health records (EHRs) contain the details of real world clinical encounters and treatment outcomes. Due to the nature of real world data, their use for regulatory decision making is contingent upon clear standards for the data and analytics methods to ensure accurate, unbiased findings.[1]
- High quality data and methods are critical in oncology research, given that patients are treated are more heterogeneously than those recruited into clinical trials. Real world patients often require~~s~~ treatment by multiple specialties, which makes the patient history complex and adds many factors that may impact outcomes.
- In this review, we sought to understand the oncology EHR research landscape in U.S. through a systematic review of the literature. We also sought to characterize the range of analytical questions and approaches used in EHR-based studies.

Methodology

- Research articles reporting use of any oncology EHR, published in Pubmed between Jan 1, 2015 and Aug 5, 2019 were included and analyzed. Scope of review was focused on commercially licensed datasets only.

1

Literature Searches

- ✓ Database: Pubmed
- ✓ Period: Jan 1, 2015 and Aug 5, 2019
- ✓ Full text publication

2

Eligibility

- ✓ Studies based on EHR
- ✓ Study type: retrospective/prospective/ cross sectional analysis of EHR
- ✓ Jurisdiction: US only
- ✓ Type of EHR: Commercially licenced

3

Analysis

Macro

- ✓ Objective(s)
- ✓ Study endpoints
- ✓ Analytical approaches

Micro

- ✓ Variables
- ✓ Structured vs unstructured data
- ✓ Data sources

The majority of the studies were published in the past 2 years

- A total of 3906 abstracts were screened, of which 35 were included following a full text review of 278 papers
- Twenty seven (77%) studies published in the past 2 years
- Studies ranged from basic descriptive studies to advanced analytics including correlation studies and predictive models

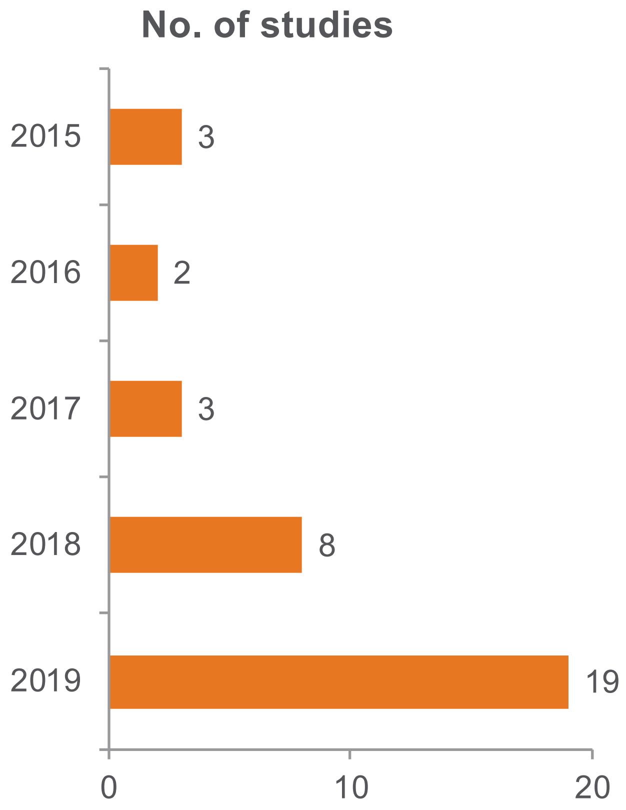
Abstracts screened
3906

Exclusion (3628)
Animal /in vitro: 58
Case study 49
Letter/editorial 19
Disease not of interest 383
Ex US 684
Not a RWE study: 449
Claims based study ***1652***
Registry 334

Full text reviewed
278

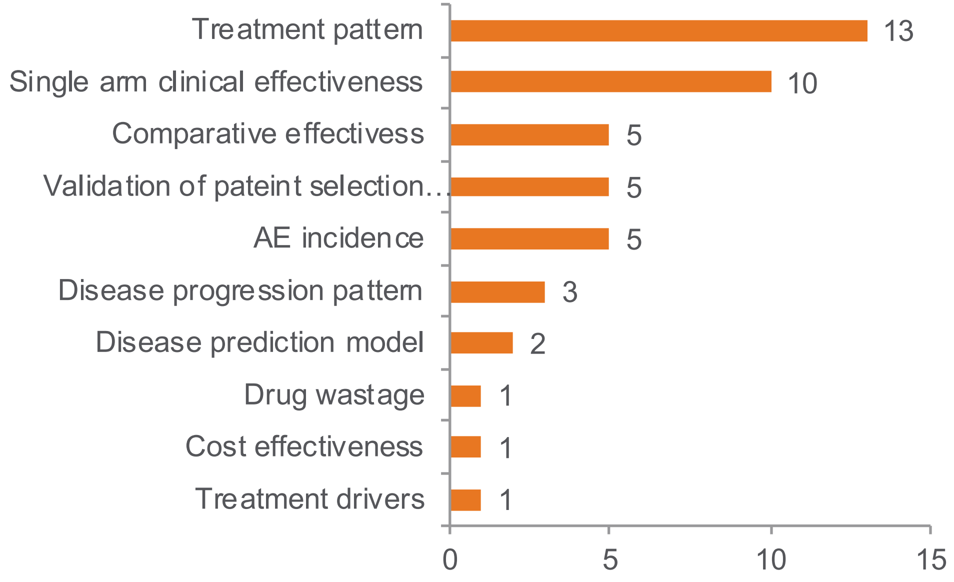
Exclusion (243)
Not a RWE study: 79
Claims based study: 12
Institutional study: 152

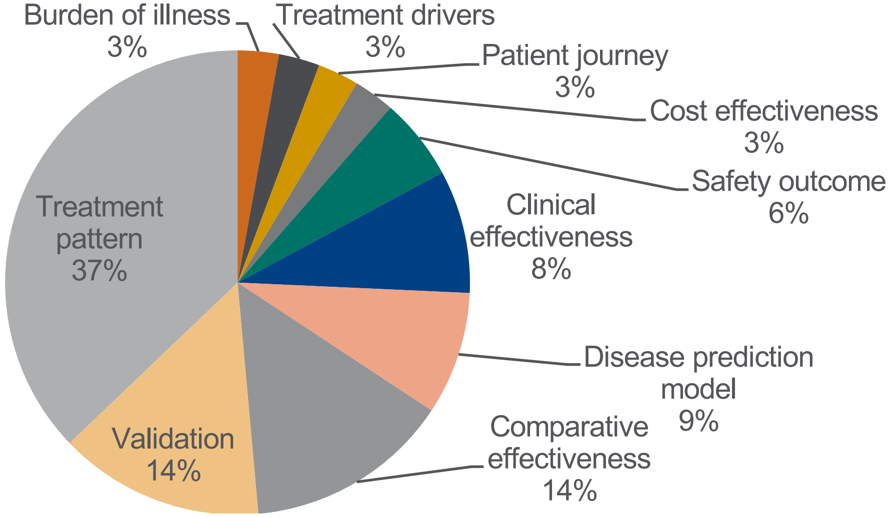
Included for analysis
35



Treatment pattern was most frequently studied, followed by single-arm treatment outcome studies assessing the effectiveness of an intervention or strategy

- Many publications had multiple objectives, with treatment pattern being the most frequently studied (13), followed by single-arm clinical effectiveness analysis (10); however these two were mostly studied together (*data shown later in the presentation*)





- Considering the primary objective, treatment pattern was most frequently studied (37%), followed by comparative effectiveness (14%), data extraction validations (14%), single-arm clinical effectiveness (9%), and others
- This suggests the strength of EHR data in capturing clinical outcomes and indicators of response to therapy

Many variables are found in the EHR for outcomes research

Some concepts are not found in a standardized EHR field

Cohort defining Variables

- Demographic variables (Age, Gender, Race, BMI)
- Risk factors (History of smoking, comorbidities)
- Clinical Variables (Stage, Biomarker/ mutation, histology, Labs, tumor size and grade)

Treatment Variables

- Type of therapy (Chemo/ targeted; mono / combo)
- Line of therapy (1L, 2L, 3L…)
- Dosing/ duration
- Switch/ discontinuation/ dose change

Outcome Variables

- Median Overall survival/ Life years gained
- Progression free survival/ Disease progression
- Clinical response/ Recurrence score

Derived cost variables

- Estimated cost of therapeutic interventions
- Healthcare resource utilization
- Wastage of drugs (in terms of healthcare costs)

a. *Lines of Therapy can be complicated to identify, Oral therapies sometimes not captured*

b. *Cancer status: Whether the patient is improving, responding to drug, progressing, in remission etc. is not captured in standard fields*

c. *For cost analyses, one needs a linked data set with EHR and claims for each patient*

Variables are listed in the order of occurrence

Structured Data are easier to access but can limit our ability understand clinical rationale, necessitating use of unstructured data from notes

- EHR has two key source of information, (a) structured data sourced through designated EHR fields predesigned in EHR platforms and (b) unstructured data lying in physician notes and lab/radiology reports
- Structured data are easily accessible through standard query of software used to extract data, but unstructured data require advanced technologies like NLP and machine learning to generate precise insights

	Type of Variables	Way to access	Objectives addressed
Structured Data	➢ **Demographic** – Age, Gender, Ethnicity, Geography ➢ **Clinical** – Comorbidity, diagnosis, history ➢ **Treatment** – Diagnostic and treatment procedures, Lab measures, Biomarker test ordered, dosing, pharmacy dispenses ➢ **Outcome** - Survival	➢ Using standard extraction query from predefined EHR tables	➢ Understand impact of a procedure or intervention on survival ➢ Switch or discontinuation ➢ Treatment algorithms
Unstructured Data	➢ **Clinical** – Symptom, stages, history, radiology, histology, performance status, Qualitative Lab results, Biomarker test result, adverse events, reasons for discontinuation ➢ **Outcome** – Response, remission, progression free survival	➢ Using NLP or through manual review of physician notes, lab results or radiology reports	➢ Association between a clinical/histological/ radiological variable and treatment decision/outcome ➢ Predict outcome/health state

OPTUM®

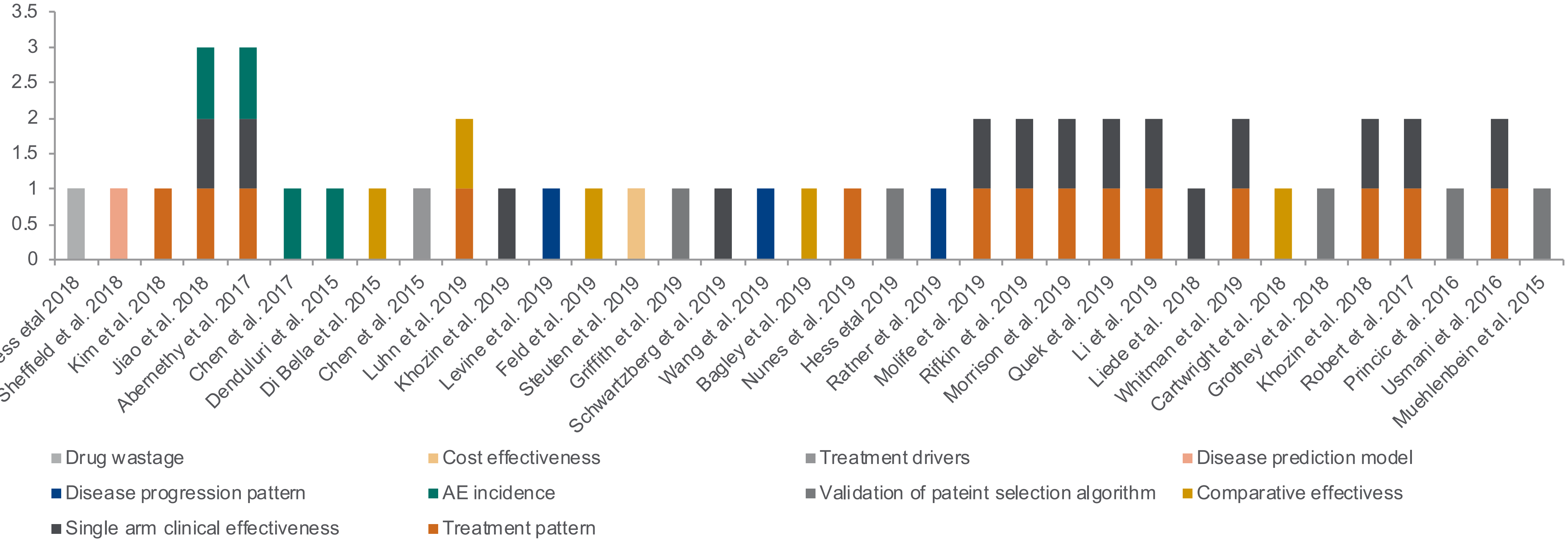
Majority of the analysis were based on structured data

- There is a clear relationship between study objectives and type of EHR source used
- Assessment of outcomes associated with a therapy, change in treatment can be analyzed mostly with structured data, while understanding underlying rationale, performance status, histologic and radiologic findings and metastatic status requires access to unstructured data

Objectives	Types of Metrics Used for Measurement
Burden of illness	Drug wastage, no. and type of visits, hospital days etc.
Single arm clinical effectiveness	Line of therapy, switch, overall survival, progression free survival, response rate, progression,
Comparative effectiveness	Adherence, all-cause discontinuation, genetic aberration, performance status, progression, overall survival, time to next treatment, time to treatment failure
Cost-effectiveness	Performance status, time to treatment failure, AEs, survival, resource utilization
Disease prediction model	Metastasis, progression, ECOG performance status
Patient journey	Progression pattern, time to progression
Safety	Treatment induced AEs, dose adjustment
Treatment drivers	Histological or radiological status, line of therapy, survival
Treatment pattern	Dosing frequency, schedule, treatment sequence, dose adjustment, switch, discontinuations, time to treatment, time to next treatment, supportive care use
Validation	RECIST, radiological or histological status, comorbidity
Note: Orange fonts indicate sourced from unstructured data	

Only few studies have targeted complex questions like treatment rationale or predicting disease type or onset.

- Although EHRs have greater clinical details regarding treatment rationale, disease progression and response to therapy, most studies focused on traditional analysis like clinical effectiveness of specific regimen or comparative effectiveness. This suggests that the EHR has yet to be harnessed for the rich insight it can offer in real world research.



OPTUM®

Both data types have their own limitations, warranting greater use of technology and analytical approaches to improve precision of real world insights

- Structured data provides standardized information, but is often missing data, limiting its utility for research
- Unstructured data are highly variable, and subject to interpretation
- Clinical language may vary across provider networks, practices and geography creating a need for language extraction methods that are adaptable, reliable, scalable and repeatable

Limitations of datasets as highlighted by authors

Structured Data

- ✓ Missing records: Labs, treatments, radiology details etc.
- ✓ Lack of Detail: Patient and family history
- ✓ Lack of insight: understanding of treatment rationale, or treatment change
- ✓ Limitation of source: Certain databases source data from oncology clinics only

Unstructured Data

- ✓ Bias: Human extraction depending on interpretation
- ✓ Lack of standardization: Abstraction and curation guides are not created based on a single accepted set of definitions
- ✓ Training: NLP technologies and human curators may be trained using different source text and training guides.
- ✓ Variability: Clinical notes are highly variable making them difficult to decipher using a common set of standards

Discussion

- In the recent years, interest in using RWD for research and specifically to supplement clinical trials has increased.[1]
- High quality RWD data is critical in oncology, as the patients recruited into cancer clinical trials often may not represent the real world population of cancer patients treated in routine practice[2]
- Our analysis suggests that there has been a notable increase in EHR publications in the past 2 years; however majority of studies focused on simpler analysis such as practice trends and clinical outcomes
- Structured data is the foundation of inputs used for real world evidence studies. However unstructured data offer rich insight into the underlying rational for treatment and insight into progression and response to therapy.
- Each dataset has limitations, warranting use of multiple data sources to supplement and correlate findings
- Finally, although oncology research attempts to design trials that are more representative of the “real world” population, it may take months or years to complete such studies as sample sizes would need to be increased and the recruitment effort could take longer.[2] Therefore the research community must find ways to improve precision and validity of studies using EHR data.

Thank You!