Disclosures & Disclaimers

- HJH: Nothing to disclose; Opinions and views expressed are my own and not those of Optum or UnitedHealth Group.
- PMK: Nothing to disclose; Opinions and views are my own and not those of the FDA.
- YN: Nothing to disclose; Opinions and views expressed are my own and not those of Syapse.
- NJR: Nothing to disclose; Opinions and views are my own and not those of the McKesson Life Sciences.

FRIENDS of CANCER RESEARCH

A Big Group, with Big Data and Big Lessons: Learnings from Friends of Cancer Research RWE Pilot 2.0

ISPOR Live Webinar

Wednesday, May 20, 2020

3:15 PM - 4:15 PM ET

FRIENDS of CANCER RESEARCH

Henry (Joe) Henk, PhD, OptumLabs

- Moderator
- Brief overview of the `Friends' RWE Pilot 2.0

Pallavi Mishra-Kalyani, PhD, U.S. FDA

- Inherent heterogeneity of data sources
- Emphasis on how data sources and sampling frames may account for differences observed in findings

Yanina Natanzon, PhD, MS, Syapse

- Describing the application of specific variable definitions to demonstrate how data heterogeneity across real world datasets (RWD) affects what attributes can be measured and analyzed, and thus potentially impact the results of the pilot

Nicholas J. Robert, MDCM, McKesson Life Sciences

- The path forward – how we can harmonize the various data sets so there are common terms/endpoints to leverage the growing amount and diversity of RWD while allowing for meaningful cross-analyses

FRIENDS of CANCER RESEARCH

Pilot 2.0: Establishing a Framework to Evaluate Real-World Endpoints

What are the real-world outcomes for aNSCLC patients treated with frontline therapies in usual care settings?	
Research Objectives	Objective 1: Description of demographic and clinical characteristics of patients with aNSCLC receiving Frontline doublet chemotherapy, PD-(L)1 monotherapy; or PD-(L)1 + doublet chemotherapy. Objective 2: Evaluate and compare rwOS, rwTTD, and rwTTNT among select frontline therapy pairs in aNSCLC
Study Design	Retrospective observational analyses of data derived from electronic health record (EHR), registries and administrative claims databases.

Data Partners:
ASCO CancerLinQ/Concerto HealthAI, COTA, Flatiron Health, IQVIA, Kaiser Permanente/CRN, Mayo Clinic/OptumLabs®, McKesson, SEER, Syapse, and Tempus

Key Collaborators:
FDA, NCI

Data Analytics/Graph Generation:
Aetion

For more information on this project, please go to https://www.focr.org/rwe

FRIENDS of CANCER RESEARCH

Harmonization versus Aggregation

Common SAP	Population identification
	Variable definitions
	Statistical analysis
	Uniform reporting template
Independently applied by each site	Dataset creation
	Statistical analysis
Presenting results	All data report to FOCR met CMS or HIPAA cell size suppression policy
	Reported results based on aggregate data and collocated by FOCR
	Survival figures based on reported life tables and collated by FOCR
Meetings, meetings, meetings	Bi-weekly tele-calls (March 2019 – present)
	In-person, ASCO 2019
	In-person, Washington DC (August 2019)
	Results: FOCR 8th Annual Blueprint for Breakthrough Forum (September 2019)

FRIENDS of CANCER RESEARCH

Data source types

Electronic Health Records (EHRs)

- Structured EMR fields
- Abstracted from unstructured notes
- Setting (community, integrated delivery networks, academic)
- Can include medical & pharmacy claims

SEER-MCR

- SEER
  - Cancer surveillance registries
  - Tumor characteristics
- Medicare
  - FFS
  - Administrative claims

ePA + claims data

- Electronic prior authorization (ePA) for chemotherapy
- Tumor characteristics
- Administrative claims

FRIENDS of CANCER RESEARCH

Inherent heterogeneity of data sources

How can data sources and sampling frames account for differences observed in findings?

FRIENDS of CANCER RESEARCH

Studies using Real-World Data Sources

- **RCTs are gold standard for comparing treatments as the process of randomization removes confounding by known and unknown factors**
  - However there are various sources of data in the real-world that could provide supplemental data to what is collected in RCTs
- **Various sources for Real-world Data are available including electronic health records, claims databases or other health plan data, and clinical registries**
  - The most robust and comprehensive real-world data will come from sources that can be linked to provide more complete information about patients
- **Each data source is designed to collect specific information: e.g. claims data are for billing and reimbursement purposes**
  - Generally not designed for research purposes, so not all types of data can be retroactively "fit" to answer a research question
  - Must consider qualities of each data type and underlying heterogeneity when trying to use it for research

FRIENDS of CANCER RESEARCH

Heterogeneity of Real-World Data Sources

Heterogeneity by type of data source

- Underlying populations are defined by network (academic vs community, region, insurance type/SES)
- Different data types designed to measure different characteristics
  - Data collection focuses on different elements of patient care
  - Example: identifying time-zero or baseline time for treatment effect estimation
    - For EHR: may be time of diagnosis
    - For claims data: may be time of first treatment administration
- Likely misclassification in each data source, but extent unknown and not comparable by data source
- Unstructured data is usually rich with information, but must be abstracted

Heterogeneity within a data source

- Large networks may have similar issues within a single data source in terms of multiple underlying population
- Other issues may include infrequent/inconsistent patient visit times/assessments, misclassification, and other measurement biases

FRIENDS of CANCER RESEARCH

Types of Bias resulting from Heterogeneity

Type of Bias	Description	Example
Selection Bias	• Failure of data to be representative of underlying patient population, or different populations across different networks of data • Unbalanced selection over time (differential censoring)	Patients choose their treatment, and this choice may be affected by health status
Confounding	• Induced by characteristics associated with treatment and endpoint • Difficult to know true confounders (potential differences on unmeasured characteristics)	Underlying supportive care may be different depending on network or within network
Misclassification	• Inaccuracies in reported data • Data entry/coding error/uncaptured information • Endpoint assessment from incomplete/different information or at inconsistent intervals	Measurements of tumor response may not be at the same intervals across or between networks

FRIENDS of CANCER RESEARCH

Regulatory Guidance and Resources

- **FDA has provided many resources for the use of various types of data and trial design**
- **May 2001: E10 Guidance for Industry[1] describes strategies for choosing a control group for clinical trials intended to demonstrate efficacy**
- **December 2018: FDA provided the Framework for the Real-World Evidence Program[2], which includes some information on how real-world data and evidence will be incorporated into regulatory decision making**

[1] https://www.fda.gov/downloads/guidances/ucm073139.pdf
[2] https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/ RealWorldEvidence/UCM627769.pdf

FRIENDS of CANCER RESEARCH

Applying common variable definitions

How does data heterogeneity across real-world datasets affects what attributes can be measured and analyzed, and thus potentially impact the results of the pilot?

FRIENDS of CANCER RESEARCH

Common variable definitions across datasets

The need

- Definitions that can be applied to heterogeneous RWD sources to generate comparable results

The approach

- **Describe**
  - Availability of each variable
  - Data source(s) for each variable
  - Method(s) of collection for each variable
  - Granularity of each variable
- **Compromise(s)**
- **Apply**

Common variable definitions across datasets

The approach

- **Describe:**
  1. Availability of each variable:
    - Demographics (Age/Race/Ethnicity vs Median household income/Insurance)
    - Smoking status
    - Molecular somatic testing/timing of testing
    - Performance status
    - Liver/kidney function

Common variable definitions across datasets

- **Describe:**
  2. Data source(s) for each variable:
  - EHR
    - one vs multiple
    - specialty vs general use
  - Claims
    - commercial
    - Medicare
    - pharmacy
  - EHR/cancer registry
  - EHR/claims
  - Cancer registry/claims
  - Other sources
    - Molecular testing laboratories
    - Mortality

FRIENDS of CANCER RESEARCH

Common variable definitions across datasets

- **Describe:**
  3. Method(s) of collection for each variable:
  - Structured
    - Enterprise data warehouse
    - HL7 messages (ADT, ORU, ORM ect)
  - Abstraction by
    - Physicians
    - Oncology nurses
    - Certified tumor registrars
    - NAACCER files
  - Data models
    - Common (OMOP, Sentinel, PCORnet ect)
    - Custom

FRIENDS of CANCER RESEARCH

Common variable definitions across datasets

- **Describe:**
  - *4. Granularity of each variable:*
    - Smoking status
      - Pack-years
      - Ever/never
      - ICD10 codes for nicotine dependence
    - Histology
      - Pathologist report
      - ICD-O-3
      - ICD10
    - Kidney/liver function
      - Clinical labs
      - ICD10
    - Molecular testing
      - Sequencing-level info
      - Specific variant information with interpretation
      - Clinical reports with variant calls, medical oncologist summaries

FRIENDS of CANCER RESEARCH

Common variable definitions across datasets

The compromise(s):

1. Create definitions that are adapted to **different** data sources but establish **the same** clinical intent
   - Exp. Claim-based (patient with continuous enrollment) vs EHR-based definitions (patients with encounters)

2. Require **minimal granularity** necessary for **clinically relevant** inference
   - Exp. Smoking pack- years vs ever/never
   - Exp. Specific NSCLC histologies vs squamous/non-squamous

3. Allow for **explicitly different** definitions for some variables
   - Exp. Kidney/Liver laboratory values vs ICD codes

FRIENDS of CANCER RESEARCH

What is our next step? The goal is to identify a path forward

How can we harmonize the various data sets so there are common terms/endpoints to leverage the growing amount and diversity in RWD while allowing for meaningful cross-analysis

FRIENDS of CANCER RESEARCH

The challenge of apples and oranges

- Different data sources
- Different definitions
- Different endpoints

FRIENDS of CANCER RESEARCH

RECIST: A model to consider

- Response evaluation criteria in solid tumors (RECIST)
- Rules that define tumor response in solid tumors
- Using different sources e.g. computed tomography (CT), physical examination
- Developed with the cooperation of multiple oncology organizations
- Used across the spectrum of clinical cancer trials to provide meaningful cross analysis

Playbook model

- Establish common definitions e.g. date of metastatic diagnosis
- Establish common methods e.g. determination of mortality
- Share these definitions and methods in the public forum with a publication

End product: A RWD language

- A language that can be used regardless of data source
- A dictionary of common terms
- The ability by using this RWD language to provide meaningful cross analysis

"Building a team is just the beginning, keeping it together means progress and working together is success"

- Henry David Thoreau

Contact information

Friends of Cancer Research

Laura Lasiter (llasiter@focr.org)

Panel

Henry (Joe) Henk (henry.henk@optum.com)

Pallavi Mishra-Kalyani (Pallavi.Mishra-Kalyani@fda.hhs.gov)

Yanina Natanzon (yanina.natanzon@syapse.com)

Nicholas J. Robert (Nicholas.RobertMD@Mckesson.com)

Q&A