

Jumping on the Real-World data (RWD)/Real World Evidence (RWE) Bandwagon?

A Guide For a Statistical Programmer to Understand Payer Claims Data



Rachana Agarwal
Servier Pharmaceuticals

What is Real World Data(RWD) ?

Real world data is data related to patients health status and any healthcare that may be administered. RWD can come from various sources like:

- **Electronic health records(EHRs)**
- **Claims and Billing data**
- **Product and disease registries**
- Data gathered from **mobile devices, wearables and biosensors**

- Electronic Health Records
- Claims and Billing Activities
- Product and disease registries
- In-home use settings
- Data gathered from other sources that can inform on health care



What is Real World Evidence (RWE) ?



RWE is the clinical evidence regarding the usage and potential benefits or risks in using a product or procedure that is derived from analyzing the Real world data(RWD).



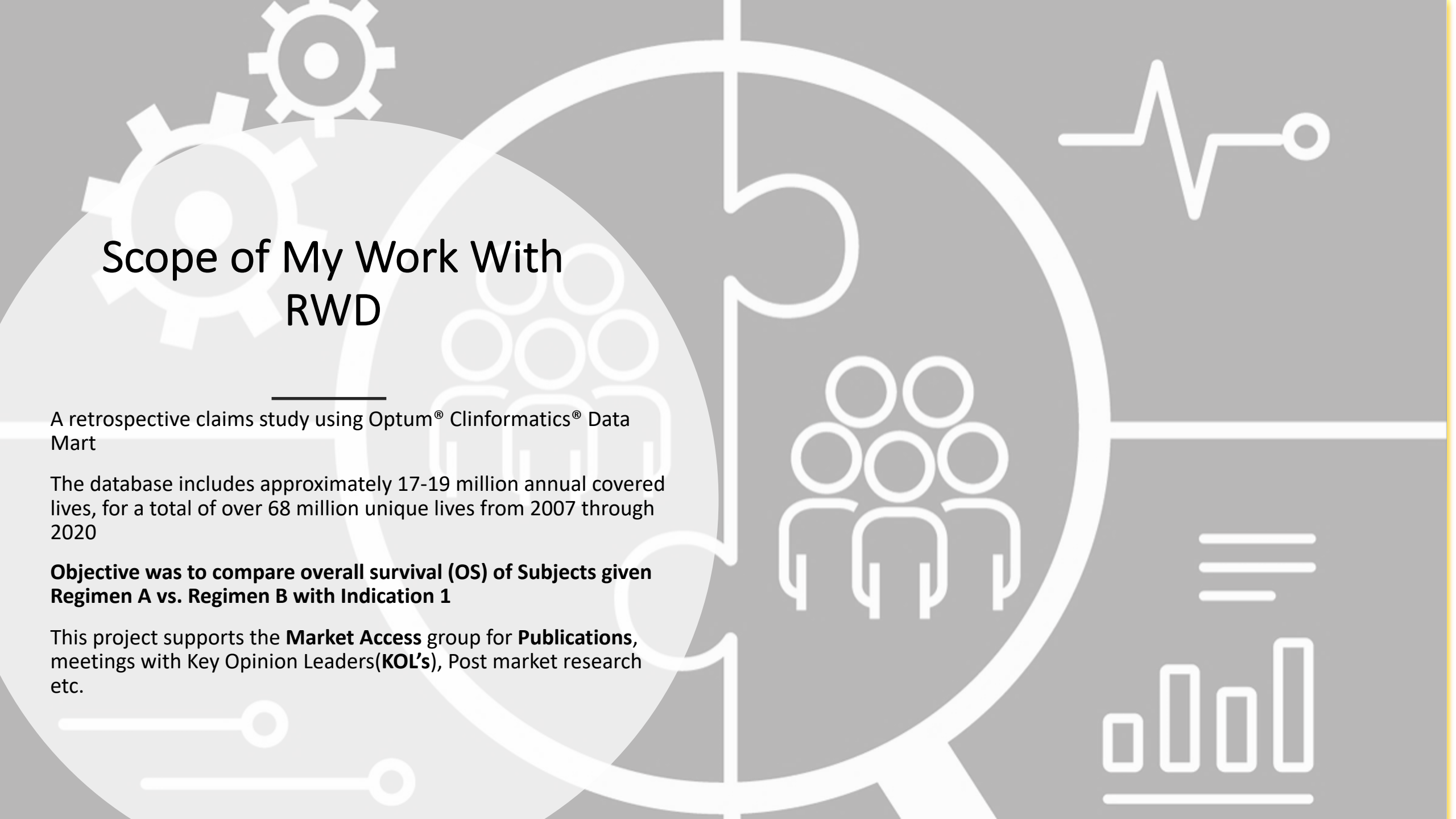
Used appropriately , **RWD** may generate **alternative evidence** when Randomized clinical trials(RCTs) are not feasible or improve trial efficiency and trial result interpretability



RWD has been commonly used to support **label extension** or **life cycle management** and is also being increasingly used to support **decision making** during regulatory **submissions**



Different study designs can be used including but not limited to, large **sample trials**, **observational studies** prospectively or retrospectively and **randomized trials** also.



Scope of My Work With RWD

A retrospective claims study using Optum® Clinformatics® Data Mart

The database includes approximately 17-19 million annual covered lives, for a total of over 68 million unique lives from 2007 through 2020

Objective was to compare overall survival (OS) of Subjects given Regimen A vs. Regimen B with Indication 1

This project supports the **Market Access** group for **Publications**, meetings with Key Opinion Leaders(**KOL's**), Post market research etc.

Optum Claims Data and Tables

- The Optum® Clinformatics® Data Mart is a database comprised of administrative health claims for members of a large nationally managed care company affiliated with Optum
- **Claims are submitted by the service providers and pharmacies**
- Claims are **verified, adjudicated, adjusted** and **de-identified** for patient privacy

Name of Data Table	Description
HRA	Health Risk Assessment
MEMBER_CONTINUOUS_ENROLLMENT	Member Eligibility
MEMBER_ENROLLMENT	Member Detail Eligibility
MEDICAL	Medical Services
MED_DIAGNOSIS	Medical Diagnosis
MED_PROCEDURE	Medical Procedure
CONFINEMENT	Inpatient Confinement
RX	Prescription Claims
LABRESULT	Lab Results
PROVIDER	Provider
PROVIDER_BRIDGE	Provider Bridge
DOD	Date of Death
LU_DIAGNOSIS	Diagnosis Lookup
LU_NDC	NDC Lookup
LU_PROCEDURE	Procedure Lookup
SES	Socio-Economic Status

References: Optum CDM User Manual

Challenges Navigating Claims Data

Challenges

- **Unfamiliarity** with data as its unlike clinical trial data
- **Size** of the database(7TB)
- Organization of the data, data is divided in **quarters**
- **Inherent limitation** of the claims data such as **incomplete, inaccurate or missing data due** to change in health plans or health care assistance programs

Lessons learnt

- Have enough **storage and processing capability** to hold 7TB data once is unpacked
- **Familiarize** with data tables and dictionaries
- **Store** Tables to save the processing time for subsequent runs
- **Organize** the Diagnosis codes, Procedure and NDC codes required and make a master files to avoid rework
- Use of **Macros** and brush up on that **Proc SQL!**

Case Study

- **Identify the subject population**
 - Subjects with **first** occurrence of the **Indication 1**, **no remission or relapse before the first diagnosis**
 - These subjects should be **enrolled in the health plan enrollment 6 month** prior to diagnosis date
- **Assign the Cohort**
 - Subjects who had Indication 1 and were given **Regimen A** drugs
 - Subjects who had Indication 1 and were given **Regimen B** drugs and were hospitalized and also had a lumbar puncture procedure

Identify the Subjects with Indication 1

- **Medical Diagnosis** and **Confinement** Datasets were used to identify the subjects with **first** diagnosis of the **Indication 1(ICD code)** and Enrolment tables used to confirm the 6 month **continuous enrolment** in the health plan
- International Classification of Diseases (**ICD**) codes are the most common type of diagnosis codes. ICD codes were developed by The World Health Organization (WHO)
- Look for ICD9 and ICD10 diagnosis codes for the Indication 1 and ignore subjects who had a remission or relapse prior to the diagnosis

Medical Diagnosis Table

PATID	PAT_PLANID	CLMID	DIAG	ICD_FLAG	FST_DT
10001	53051584796	2898883745	20270	9	2010-04-16
10001	53051584796	2899491546	20700	9	2010-04-20
10001	53051584796	2899690202	20250	10	2010-04-21
10001	53051584796	2902807744	27050	10	2010-05-10

Confinement

Patid	Admit_Date	Conf_Id	Diag1	Diag2	Diag3	Diag4	Diag5	Disch_Date
1001	2015-02-26	OMONR4K	5990	4280	4168	2761	25000	2015-03-03
1001	2015-08-08	OMONR4K	5990	5849	42731	29420	V4987	2015-08-11

Associated ICD-9-CM codes for the treatment of multiple myeloma

203 Multiple myeloma and immunoproliferative neoplasms
 203.0 Multiple myeloma
 Kahler's disease
 Myelomatosis
Excludes: solitary myeloma (238.6)

The following fifth-digit subclassification is for use with category 203:
 0 without mention of having achieved remission
 >Failed remission<
 1 in remission
 2 in relapse

Associated ICD-10-CM codes (effective October 1, 2014) for the treatment of multiple myeloma

C90 Multiple myeloma and malignant plasma cell neoplasms
 C90.00 Multiple myeloma not having achieved remission
 C90.01 Multiple myeloma in remission
 C90.02 Multiple myeloma in relapse

Create a Master Medical Claims

Master Medical Claims										
Patid	Ndc	Proc_Cd	Fst_Dt	Rvnu_Cd	Pos	Std_Cost	Conf_Id	Bill_Type	Tos_Cd	Tos_Ext
20001		96450	4/26/2017		21	195.46	ZMZN4MLKN4KKM		PROF.PROOTH	PROF.PROOTH.CHEMO..
20001		96450	5/2/2017		21	195.46	ZMZN4MLKN4KKM	131	PROF.PROOTH	PROF.PROOTH.CHEMO
20001	NONE	J2505	5/4/2017	636	22	6650.28		131	ANC.DRUGAD	ANC.DRUGAD.6603.66218

- Create Master claims files by for the subjects in the population set and Medical Claims data using the **NDC** codes and **HCPCS** codes for **all the drugs** and procedures **that can be administered as part of the Regimens for ease of processing**
- **NDC** stands for National Drug Code and is a unique 10 or 11 digit, universal product identifier for human drugs in the United States
- **Healthcare Common Procedure Coding System (HCPCS)** is a collection of **standardized codes that represent medical procedures**, supplies, products and services. The codes are used to facilitate the **processing of health insurance claims by Medicare and other insurers**

10-Digit Format on Package	10-Digit Format Example	11-Digit Format	11-Digit Format Example
4-4-2	9999-9999-99	<u>5</u> -4-2	<u>0</u> 9999-9999-99
5-3-2	99999-999-99	5- <u>4</u> -2	99999- <u>0</u> 999-99
5-4-1	99999-9999-9	5-4- <u>2</u>	99999-9999- <u>0</u> 9

Treatment	HCPCS	NDC
Vincristine Sulfate Liposome injection	J9371	72893-008-03
Vincristine	J9370	20536-322-01
	J9375	72893-008-03
	J9380	0703-4402-11

Medical Claims Data							
PATID	FST_DT	NDC	PAID_DT	POS	PROC_CD	PROVCA	TOS_CD
10001	3/16/2016	NONE	4/7/2016	22	J9000	1	ANC.DRU
10001	1/3/2016	NONE	1/29/2016	22	J9001	1	ANC.DRU
10001	2/2/2016	NONE	2/24/2016	22	J9002	1	ANC.DRU

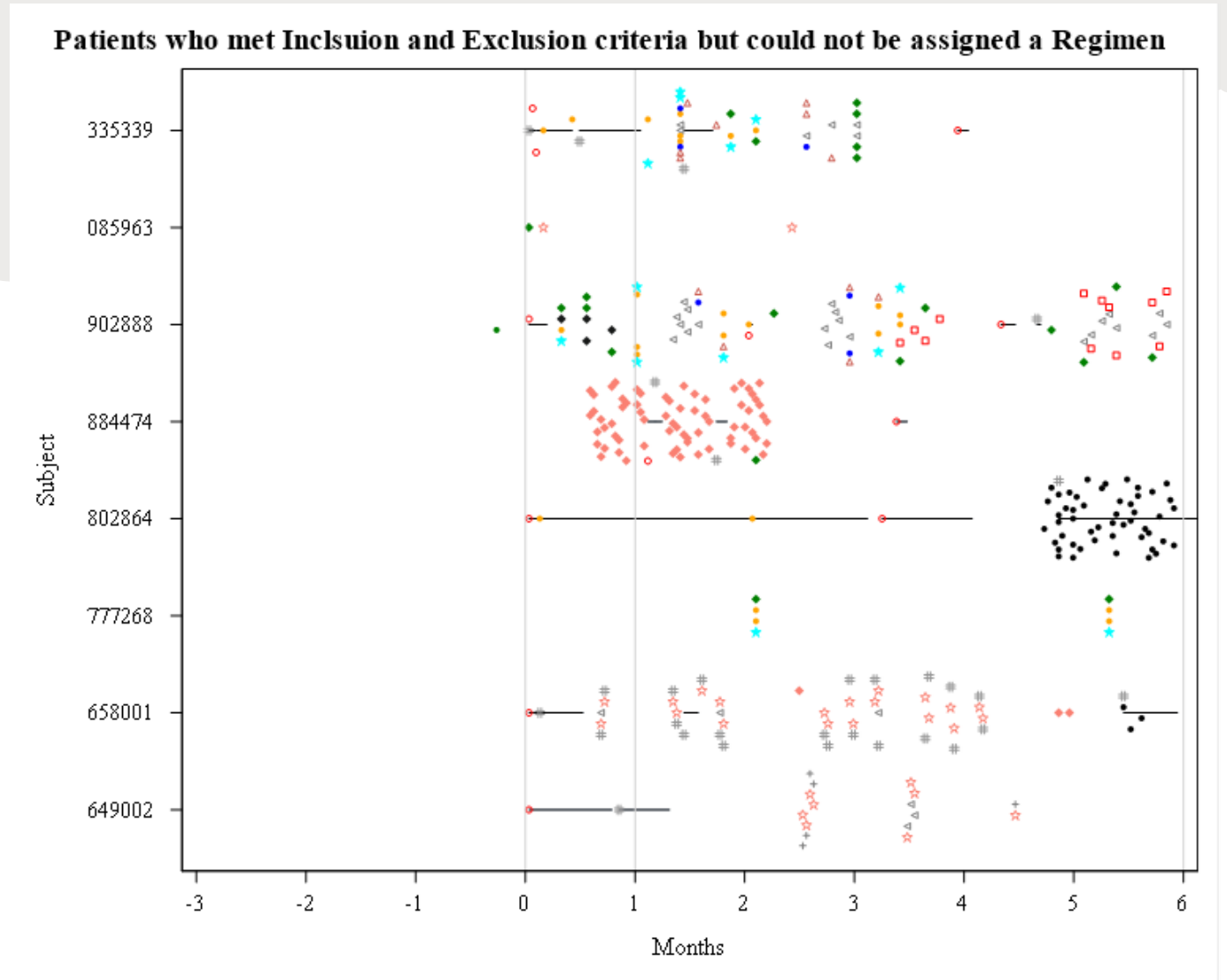
C o h o r t d e f i n e d			
P a t i d	A L L D I A G _ D A T E	I N D E X _ D A T E	C o h o r t
1 0 0 0 1	1 4 F e b 2 0 0 9	1 4 F e b 2 0 0 9	R e g i m e n A
1 0 0 0 2	0 8 J u l 2 0 1 1	0 5 J u l 2 0 1 1	
1 0 0 0 3	1 7 A p r 2 0 1 8	1 5 J u l 2 0 1 6	R e g i m e n A
1 0 0 0 4	1 5 A p r 2 0 0 8	1 5 A p r 2 0 0 8	R e g i m e n B
1 0 0 0 5	0 4 J a n 2 0 0 7	0 4 J a n 2 0 0 7	

Identify Subjects Who Were Given Regimen A or B

- **Regimen A or B** for these subjects is identified through presence of relevant **NDC** or **HCPCS** codes
- Subjects included in **Regimen B** will also have a **hospitalization** and a **Lumbar puncture** procedure.
- Hospitalizations are identified through the **Confinement data**
- Lumbar puncture is identified by the **HCPCS** code in **Medical claims data** and the HCPS code for it is (62270, 62328, 62272, 62329, 96450)
- **Missing Cohorts** for many subjects using the earlier algorithms

Scatter Plots to identify new Algorithms for cohort assignment

- Plots displaying all drugs and procedures given to the patients
- New Algorithms developed with the help of KOL's using these subject level scatter plots
- A Subject Level Analysis Dataset was created to support the statistical analysis



Examples of Analysis For Clinical Outcomes

Cohort and Censoring data														
Patid	INDEX_DATE	Cohort	PLAN_SDT	PLAN_EDT	Gender	Race	plan_edt2	dod	paramcd	death	aval	avalm	avalw	cnsr
10001	14Feb2009	Regimen B	01Jul2000	31Mar2003	M	W	30Jun2014		OS	0	1963	64.47	280.29	1
10002	14Feb2009	Regimen A	01Aug2007	30Sep2009	M	W	30Jun2014		OS	0	1963	64.47	280.29	1
10004	14Feb2009	Regimen A	01Jul2012	30Jun2014	M	W	30Jun2014		OS	0	1963	64.47	280.29	1
10005	05Jul2011	Regimen B	01May2000	31Dec2014	F	W	31Dec2014		OS	0	1276	41.88	182.14	1
10006	15Jul2016	Regimen B	01Jul2000	31May2003	M	W	31May2018	01May2018	OS	1	656	21.52	93.57	0

- **Time to Event dataset** is created using the death table and enrollment tables to calculate the time to event and censor subjects
- Kaplan Meier plots, Forest Plots and calculate Hazard Ratios are produced
- **Unlike Traditional Clinical trials**, it takes extra steps to Identify the population set and treatment assignments using RWD
- Despite the challenges it is **possible to analyze RWD using SAS** with **some planning and out of the box thinking**

Thank You!



Rachana Agarwal

Servier Pharmaceuticals

Rachana.agarwal@servier.com