



SOS Week 2 Tutorial: Data Diagnostics



KUD for Today



KNOW

What is Database Diagnostics?



UNDERSTAND

How to create the profile for a given database

How to fill out the data diagnostics inputs



DO

Interpret the data diagnostics output for one SOS example



Agenda

Overview of Data Diagnostics

- Explain what data diagnostics is and how it works

Build a Database Profile together

- Demo of the `executeDbProfile` function and the outputs to share

Data Diagnostic Study Question Inputs

- What elements of a study question you need to know to run data diagnostics
- Show the inputs for the 4 study questions

Go through the output

- Demo the shiny app and how to interpret data diagnostic results



Agenda

Overview of Data Diagnostics

- Explain what data diagnostics is and how it works

Build a Database Profile together

- Demo of the `executeDbProfile` function and the outputs to share

Data Diagnostic Study Question Inputs

- What elements of a study question you need to know to run data diagnostics
- Show the inputs for the 4 study questions

Go through the output

- Demo the shiny app and how to interpret data diagnostic results



What is Data Diagnostics?



R package that allows us to determine



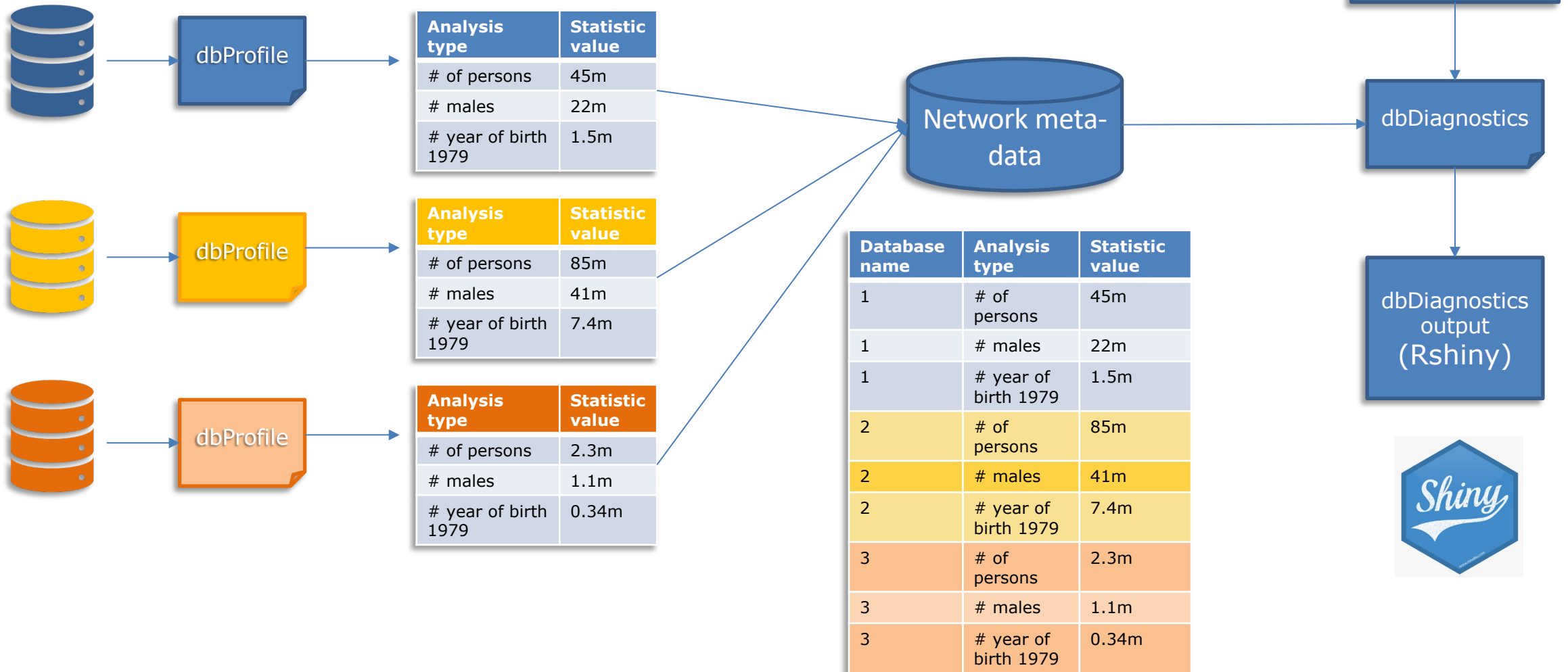
which databases have the elements required to answer a research question



using only a set of aggregated summary statistics

Data Diagnostics Process

Source meta-data





Database Profile Summary Statistics

List of Summary Statistics included in executeDbProfile

The executeDbProfile function will execute and/or export the following aggregate summary statistics and DQD checks:

Persons	total	Measurement	occurrence records, by measurement_concept_id	
	by gender		occurrence records, by measurement_source_concept_id	
	by year of birth		records with no value (numeric, string or concept)	
	by race		Condition occurrence	by condition_concept_id
	by ethnicity			by condition_source_concept_id
	with at least one day of observation in each month		Drug exposure	by drug_concept_id
	by observation period start month			by drug_source_concept_id
	by number of observation periods		Procedure occurrence records	by procedure_concept_id
	by length of observation period (in 30d increments)			by procedure_source_concept_id
	with at least one visit occurrence, by visit_concept_id		Observation occurrence records	by observation_concept_id
	distinct patients that overlap between specific domains - including death			by observation_source_concept_id
	with at least one concept_id, by measurement_concept_id		Device exposure records	by device_concept_id
	with at least one concept_id, by condition_concept_id			by device_source_concept_id
	with at least one concept_id, by procedure_concept_id		Distribution of numeric values	by measurement_concept_id and unit_concept_it
	with at least one concept_id, by drug_concept_id			
	with at least one concept_id, by device_concept_id			
with at least one concept_id, by observation_concept_id				



Agenda

Overview of Data Diagnostics

- Explain what data diagnostics is and how it works

Build a Database Profile together

- Demo of the `executeDbProfile` function and the outputs to share

Data Diagnostic Study Question Inputs

- What elements of a study question you need to know to run data diagnostics
- Show the inputs for the 4 study questions

Go through the output

- Demo the shiny app and how to interpret data diagnostic results



Let's Build a Database Profile!





Agenda

Overview of Data Diagnostics

- Explain what data diagnostics is and how it works

Build a Database Profile together

- Demo of the `executeDbProfile` function and the outputs to share

Data Diagnostic Study Question Inputs

- What elements of a study question you need to know to run data diagnostics
- Show the inputs for the 4 study questions

Go through the output

- Demo the shiny app and how to interpret data diagnostic results



Data Diagnostics User Inputs

All Available Inputs

Age range: min ___ – max ___

Gender: Male Female Unknown

Race: White Black Asian Unknown

Ethnicity: Hispanic Not Hispanic Unknown

Calendar time: Start Date _____ - End Date _____

Minimum longitudinal follow-up: ___ days

Target: <concept list>

Comparator: <concept list>

Indication: <concept list>

Outcome: <concept list>

Data requirements		All	Any
Domains	Condition	x	x
	Drug	x	x
	Procedure	x	x
	Measurement	x	x
	Measurement with value	na	x
	Device	x	x
	Observation	x	x
	Death	x	x
Visits	Inpatient	x	x
	Emergency Room	x	x
	Outpatient	x	x



Data Diagnostics User Inputs

Aflibercept vs. Bevacizumab for Blinding Diseases with ESRD Outcome

Age range: min 18 – max NA

Gender: Male Female

Race: NULL (meaning there are no restrictions)

Ethnicity: NULL

Calendar time: Start Date NULL - End Date NULL

Minimum longitudinal follow-up: 365 days

Target: aflibercept

Comparator: bevacizumab

Indication: blinding diseases

Outcome: ESRD

Data requirements		All	Any
Domains	Condition	x	
	Drug	x	
	Procedure		
	Measurement		
	Measurement with value	<i>na</i>	
	Device		
	Observation		
	Death		
Visits	Inpatient		
	Emergency Room		
	Outpatient		



Data Diagnostics User Inputs

Aflibercept vs. Ranibizumab for Blinding Diseases with ESRD Outcome

Age range: min 18 – max NA

Gender: Male Female

Race: NULL (meaning there are no restrictions)

Ethnicity: NULL

Calendar time: Start Date NULL - End Date NULL

Minimum longitudinal follow-up: 365 days

Target: aflibercept

Comparator: ranibizumab

Indication: blinding diseases

Outcome: ESRD

Data requirements		All	Any
Domains	Condition	x	
	Drug	x	
	Procedure		
	Measurement		
	Measurement with value	<i>na</i>	
	Device		
	Observation		
	Death		
Visits	Inpatient		
	Emergency Room		
	Outpatient		



Data Diagnostics User Inputs

Ranibizumab vs. Bevacizumab for Blinding Diseases with ESRD Outcome

Age range: min 18 – max NA

Gender: Male Female

Race: NULL (meaning there are no restrictions)

Ethnicity: NULL

Calendar time: Start Date NULL - End Date NULL

Minimum longitudinal follow-up: 365 days

Target: ranibizumab

Comparator: bevacizumab

Indication: blinding diseases

Outcome: ESRD

Data requirements		All	Any
Domains	Condition	x	
	Drug	x	
	Procedure		
	Measurement		
	Measurement with value	<i>na</i>	
	Device		
	Observation		
	Death		
Visits	Inpatient		
	Emergency Room		
	Outpatient		



All SOS Study Question Inputs

	antiVEGF Study	Fluoroquinolone	MS Biologics	Risankizumab
Age restrictions	> 18	> 35	None	None
Gender(s)	male, female	male, female	male, female	male, female
Race(s)	All	All	All	All
Ethnicities(s)	All	All	All	All
Study Dates	Any	Jan 2010 – Dec 2019	Any	>= Apr 2019
Required Lookback	365 days	365 days	365 days	365 days
Required Domains	condition, drug	condition, drug	condition, drug	condition, drug
Desired Domains	NULL	procedure	NULL	NULL
Required Visits	NULL	NULL	NULL	NULL
Desired Visits	NULL	IP	NULL	IP
Target	Aflibercept	Fluoroquinolone	MS biologics	Risankizumab
Comparator	Bevacizumab	Cephalosporin	Disease modifying treatments	Biologics for PS
Indication	Blinding Diseases	UTI	Multiple Sclerosis	Psoriasis
Outcome	ESRD	Aortic Aneurysm	PML	Ischemic Stroke
Analysis Variants	3	6	2	3



Let's Run Data Diagnostics!





KUD for Today



KNOW

What is Database Diagnostics?



UNDERSTAND

How to create the profile for a given database

How to fill out the data diagnostics inputs



DO

Interpret the data diagnostics output for one SOS example