Working with What We Have

Running OMOP-Based Network Studies in Asia-Pacific
(based on GLP-1 study)

Evelyn Goh | National University of Singapore , Saw Swee Hock School of Public Health

2 Apr 2026 | OHDSI APAC Scientific Forum, Apr 2026

In 2024, we ran the GLP-1 study

In 2024, we ran the GLP-1 study

- Is use of GLP-1 associated with acute liver injury?
  - Compared it with DPP4-i
  - Assessed via the standard OHDSI pipeline (Strategus, CohortMethod, etc.)

In 2024, we ran the GLP-1 study

- Is use of GLP-1 associated with acute liver injury?
  - Compared it with DPP4-i
  - Assessed via the standard OHDSI pipeline (Strategus, CohortMethod, etc.)

Partway through, we realised that there were some database-related issues.

APAC cohorts kept failing diagnostics

- Of 15 cohorts, 9 passed
  - N. America: 8/9
  - APAC-ANZ: 1/4
  - EU: 0/2

Cohort sizes got even poorer

- Before diagnostics, cohort sizes were already poor
  - N. America: 96.8%
  - APAC-ANZ: 2.7%
  - EU: 0.5%

- After diagnostics, N.Am made up over 99% of the data.
  - The remaining 1% came from APAC-ANZ JMDC (Japan)

OHDSI Frameworks are built around N. Am DBs

Source name	Country	COVARIATE BALANCE (ASDM < 0.1)	EMPIRICAL EQUIPOISE (PS overlaps > 0.25)	RESIDUAL BIAS (EASE < 0.25)	Overall
France Disease Analyzer	France	Fail	Fail	Pass	Fail
German Disease Analyzer	Germany				
Yonsei University Severance CDM	Korea	Fail	Fail	Pass	Fail
Japan Medical Data Center (JMDC)	Japan	Pass	Pass	Pass	Pass
Taipei Medical University	Taiwan	Fail	Fail	Pass	Pass
LPD Australia	Australia	Fail	Fail	Pass	NOT EVALUATED
MarketScan Multi-State Medicaid	U.S.	Pass	Pass	Pass	Pass
MarketScan Medicare Supplemental (MDCR)	U.S.	Pass	Pass	Pass	Pass
MarketScan Commercial Claims (CCAE)	U.S.	Pass	Pass	Pass	Pass
Optum EHR	U.S.	Pass	Pass	Pass	Pass
HealthVerity CC	U.S.	Pass	Pass	Pass	Pass
Iqvia LRx-US9-LAAD	U.S.	Pass	Pass	Pass	Pass
US Department of Veterans Affairs (VA)	U.S.	Pass	Pass	Pass	Pass
PharMetrics	U.S.	Pass	Pass	Pass	Pass
OPTUM Extended DOD	U.S.				

Objectives (Long-Term)

Objectives (Long-Term)

- Conduct a meta-study (study of studies) that have databases from multiple regions

Objectives (Long-Term)

- Conduct a meta-study (study of studies) that have databases from multiple regions
- Assess how regional databases may differ from N. Am databases

Objectives (Long-Term)

- Conduct a meta-study (study of studies) that have databases from multiple regions
- Assess how regional databases may differ from N. Am databases
- If later found necessary, develop a set of guidelines, strategies, or frameworks on how to work with regional databases

Objectives (Short-Term – i.e., today)

Objectives (Short-Term – i.e., today)

- Present preliminary results based on one study
  - GLP-1 study

Objectives (Short-Term – i.e., today)

- Present preliminary results based on one study
  - GLP-1 study

- Open the floor for input, discussion for:
  - Exploratory methodology
  - Access to other studies, possibly running a meta-study package
  - Getting more people involved to guide the study
    - Knowledge of OHDSI statistical workings
    - Knowledge of OHDSI tech stack

Quick (?) Methodology

Quick (?) Methodology

- No, there is no protocol (yet)
  - Mostly basic statistical associations

Quick (?) Methodology

- No, there is no protocol (yet)
  - Mostly basic statistical associations

- Of the three main diagnostics criteria
  - Covariate balance
  - Empirical equipoise
  - EASE

Quick (?) Methodology

- No, there is no protocol (yet)
  - Mostly basic statistical associations

- Of the three main diagnostics criteria
  - **Covariate balance – fails here**
  - **Empirical equipoise – fails here**
  - EASE

Quick (?) Methodology

- No, there is no protocol (yet)
  - Mostly basic statistical associations

- Of the three main diagnostics criteria
  - **Covariate balance – fails here – analysed for today**
  - **Empirical equipoise – fails here**
  - EASE

Quick (?) Methodology

- No, there is no protocol (yet)
  - Mostly basic statistical associations

- Today, we are looking at covariate balance in two studies
  - Distribution of covariates across regions
  - Observable patterns

Quick (?) Methodology

- No, there is no protocol (yet)
  - Mostly basic statistical associations

- Today, we are looking at covariate balance in two studies
  - Distribution of covariates across regions
  - Observable patterns

OK let’s start

And once again a reminder that this was “winging it”

Covariate balance in OHDSI studies

- OHDSI automatically generates covariates grouped by domains:
  - Condition occurrence concepts
  - Drug exposure concepts
  - Procedure concepts
  - Observation/measurement concepts
  - Demographics

- Uses FeatureExtraction
  - Define which domains are eligible
  - Define which time windows to use
  - Generate all eligible concepts then let regularized regression select

For a DB to pass covariate balance

- We set a threshold
  - ASDM = 0.1
  - If a cohort is over the threshold, it passes
  - If a cohort is under the threshold, it fails

For a DB to pass covariate balance

- We set a threshold
  - ASDM = 0.1
  - If a cohort is over the threshold, it passes
  - If a cohort is under the threshold, it fails

- Why did so many regional DBs fail the ASDM threshold?

To assess why

- At covariate threshold = 0.1
- For each cohort:
  - How many covariates?
- For each covariate:
  - Which covariates have cohort counts below threshold?
  - How much cohort count does each covariate have?

What I did

- Went to the ShinyApps for GLP1 and FQ (2023 SOS) studies
  - Downloaded raw data
  - Cohort counts, diagnostics, estimations, etc.

What I did

- Went to the ShinyApps for GLP1 and FQ (2023 SOS) studies
  - Downloaded raw data
  - Cohort counts, diagnostics, estimations, etc.

- Ran a local script to analyse (as we will next see)

Results from GLP1 study

1/2

Quick reminder

- Is use of GLP-1 associated with acute liver injury?
  - Compared it with DPP4-i
  - Assessed via the standard OHDSI pipeline (Strategus, CohortMethod, etc.)

GLP1 Study: Cohort Counts (before diagnostics)

Source name	Country	Region	Cohort count (ALI)	
France Disease Analyzer	France	Europe	1,360	
German Disease Analyzer	Germany	Europe	43,827	
			45187	**0.5%**
Yonsei University Severance CDM	Korea	APAC-ANZ	114,873	
Japan Medical Data Center (JMDC)	Japan	APAC-ANZ	111,180	
Taipei Medical University	Taiwan	APAC-ANZ	15,612	
LPD Australia	Australia	APAC-ANZ	36	
			241,701	**2.7%**
MarketScan Multi-State Medicaid	U.S.	N.Am	294,140	
MarketScan Medicare Supplemental (MDCR)	U.S.	N.Am	115,383	
MarketScan Commercial Claims (CCAE)	U.S.	N.Am	622,610	
Optum EHR	U.S.	N.Am	570,271	
HealthVerity CC	U.S.	N.Am	1,602,008	
Iqvia LRx-US9-LAAD	U.S.	N.Am	4,226,700	
US Department of Veterans Affairs (VA)	U.S.	N.Am	270,175	
PharMetrics	U.S.	N.Am	394,002	
OPTUM Extended DOD	U.S.	N.Am	736,315	
			8,831,604	**96.8%**

9,118,492

GLP1 Study: Cohort Counts (before diagnostics)

Source name	Country	Cohort count - GLP1	Cohort count - DPP4	Cohort count - after matching	Matched/GLP1	Matched/DPP4
France Disease Analyzer	France	1,863	2,284	621	33%	27%
German Disease Analyzer	Germany	15,532	25,173			
		17,395	**27,457**	**621**		
Yonsei University Severance CDM	Korea	2,537	15,471	571	23%	4%
Japan Medical Data Center (JMDC)	Japan	7,025	19,380	3,134	45%	16%
Taipei Medical University	Taiwan	1,969	15,711	938	48%	6%
LPD Australia	Australia	8	156			
		11,539	**50,718**	**4,643**		
MarketScan Multi-State Medicaid	U.S.	43,064	36,136	11,039	26%	31%
MarketScan Medicare Supplemental (MDCR)	U.S.	37,421	71,906	5,924	16%	8%
MarketScan Commercial Claims (CCAE)	U.S.	300,267	275,699	41,273	14%	15%
Optum EHR	U.S.	374,383	346,229	94,372	25%	27%
HealthVerity CC	U.S.	883,609	505,236	245,191	28%	49%
Iqvia LRx-US9-LAAD	U.S.	3,940,626	2,509,387	929,511	24%	37%
US Department of Veterans Affairs (VA)	U.S.	130,688	160,641	32,310	25%	20%
PharMetrics	U.S.	367,916	143,106	94,143	26%	66%
OPTUM Extended DOD	U.S.	243,056	204,402			
		6,321,030	**4,252,742**	**1,453,763**		

GLP1 Study: Cohort Counts (before diagnostics)

Source name	Country	Cohort count - GLP1	Cohort count - DPP4	Cohort count - after matching	Matched/GLP1	Matched/DPP4
France Disease Analyzer	France	1,863	2,284	621	33%	27%
German Disease Analyzer	Germany	15,532	25,173			
		17,395	**27,457**	**621**		
Yonsei University Severance CDM	Korea	2,537	15,471	571	23%	4%
Japan Medical Data Center (JMDC)	Japan	7,025	19,380	3,134	45%	16%
Taipei Medical University	Taiwan	1,969	15,711	938	48%	6%
LPD Australia	Australia	8	156			
		11,539	**50,718**	**4,643**		
MarketScan Multi-State Medicaid	U.S.	43,064	36,136	11,039	26%	31%
MarketScan Medicare Supplemental (MDCR)	U.S.	37,421	71,906	5,924	16%	8%
MarketScan Commercial Claims (CCAE)	U.S.	300,267	275,699	41,273	14%	15%
Optum EHR	U.S.	374,383	346,229	94,372	25%	27%
HealthVerity CC	U.S.	883,609	505,236	245,191	28%	49%
Iqvia LRx-US9-LAAD	U.S.	3,940,626	2,509,387	929,511	24%	37%
US Department of Veterans Affairs (VA)	U.S.	130,688	160,641	32,310	25%	20%
PharMetrics	U.S.	367,916	143,106	94,143	26%	66%
OPTUM Extended DOD	U.S.	243,056	204,402			
		6,321,030	**4,252,742**	**1,453,763**		

GLP1 Study: Cohort Counts

Source name	Country	Cohort count - GLP1	Cohort count - DPP4	Cohort count - after matching	Matched/GLP1	Matched/DPP4
France Disease Analyzer	France	1,863	2,284	621	33%	27%
German Disease Analyzer	Germany	15,532	25,173			
		17,395	**27,457**	**621**		
Yonsei University Severance CDM	Korea	2,537	15,471	571	23%	4%
Japan Medical Data Center (JMDC)	Japan	7,025	19,380	3,134	45%	16%
Taipei Medical University	Taiwan	1,969	15,711	938	48%	6%
LPD Australia	Australia	8	156			
		11,539	**50,718**	**4,643**		
MarketScan Multi-State Medicaid	U.S.	43,064	36,136	11,039	26%	31%
MarketScan Medicare Supplemental (MDCR)	U.S.	37,421	71,906	5,924	16%	8%
MarketScan Commercial Claims (CCAE)	U.S.	300,267	275,699	41,273	14%	15%
Optum EHR	U.S.	374,383	346,229	94,372	25%	27%
HealthVerity CC	U.S.	883,609	505,236	245,191	28%	49%
Iqvia LRx-US9-LAAD	U.S.	3,940,626	2,509,387	929,511	24%	37%
US Department of Veterans Affairs (VA)	U.S.	130,688	160,641	32,310	25%	20%
PharMetrics	U.S.	367,916	143,106	94,143	26%	66%
OPTUM Extended DOD	U.S.	243,056	204,402			
		6,321,030	**4,252,742**	**1,453,763**		

GLP1 Study: Cohort Counts

Source name	Country	Cohort count - GLP1	Cohort count - DPP4	Cohort count - after matching	Matched/GLP1	Matched/DPP4
France Disease Analyzer	France	1,863	2,284	621	33%	27%
German Disease Analyzer	Germany	15,532	25,173			
		17,395	**27,457**	**621**		
Yonsei University Severance CDM	Korea	2,537	15,471	571	23%	4%
Japan Medical Data Center (JMDC)	Japan	7,025	19,380	3,134	45%	16%
Taipei Medical University	Taiwan	1,969	15,711	938	48%	6%
LPD Australia	Australia	8	156			
		11,539	**50,718**	**4,643**		
MarketScan Multi-State Medicaid	U.S.	43,064	36,136	11,039	26%	31%
MarketScan Medicare Supplemental (MDCR)	U.S.	37,421	71,906	5,924	16%	8%
MarketScan Commercial Claims (CCAE)	U.S.	300,267	275,699	41,273	14%	15%
Optum EHR	U.S.	374,383	346,229	94,372	25%	27%
HealthVerity CC	U.S.	883,609	505,236	245,191	28%	49%
Iqvia LRx-US9-LAAD	U.S.	3,940,626	2,509,387	929,511	24%	37%
US Department of Veterans Affairs (VA)	U.S.	130,688	160,641	32,310	25%	20%
PharMetrics	U.S.	367,916	143,106	94,143	26%	66%
OPTUM Extended DOD	U.S.	243,056	204,402			
		6,321,030	**4,252,742**	**1,453,763**		

Next we look at covariate coverage

- For each database,
  - When ASDM=0.1
  - What is the ratio of covariates with subject counts
    - <threshold (labelled -1)
    - <1,000
    - <10,000
    - <20,000
    - <50,000
    - <100,000
    - <500,000
    - <1M
    - >1M

Ratio of covariates with subject counts <x

STUDY: GLP1

COHORT: ACUTE LIVER INJURY

Cohort comparisons across database – all incidences of acute liver injury

	-1	<100	<1000	<10,000	<20,000	<50,000	<100,000	<500,000	<1,000,0	>1M
optum ehr count	0.5753	0.5753	0.5753	0.7158	0.8016	0.8838	0.9326	1	1	1
healthverity count	0.6675	0.6675	0.6675	0.6675	0.7791	0.878	0.9292	0.9952	1	1
lrx-us9-laad count	0.7553	0.7553	0.7553	0.7553	0.7553	0.8251	0.8885	0.9748	0.9935	1
pharmetrics count	0.6643	0.6643	0.6643	0.8724	0.925	0.9774	0.9955	1	1	1
ibm mdcr count	0.66	0.66	0.7005	0.942	0.9735	0.9964	1	1	1	1
ibm ccae count	0.7299	0.7299	0.7299	0.8543	0.9115	0.9629	0.9871	1	1	1
va-omop count	0.5729	0.5729	0.5729	0.8085	0.8774	0.9402	0.9738	1	1	1
ibm mdcd count	0.6727	0.6727	0.6727	0.8905	0.9391	0.9817	0.9964	1	1	1
optum extended dod count	0.6224	0.6224	0.6224	0.761	0.8492	0.9263	0.9687	1	1	1
lpdau count	0.9916	1	1	1	1	1	1	1	1	1
severance cdm count	0.6655	0.6655	0.7545	0.9446	0.9712	0.9991	1	1	1	1
tmudb count	0.7898	0.7898	0.9488	0.9999	1	1	1	1	1	1
jmdc count	0.7966	0.7966	0.8332	0.9621	0.9837	0.999	1	1	1	1
france da count	0.8936	0.9775	1	1	1	1	1	1	1	1
german da count	0.8792	0.8792	0.9455	0.9986	0.9999	1	1	1	1	1

Ratio of covariates with subject counts <x

STUDY: GLP1

COHORT: GLP1 USERS

Cohort comparisons across database – all GLP1											
				-1	<100	<1000	<10,000	<20,000	<50,000	<100,000	<500,000	<1,000,0	>1M
optum ehr count				0.5924	0.5924	0.5924	0.7502	0.8253	0.9027	0.9455	1	1	1
healthverity count				0.6768	0.6768	0.6768	0.6964	0.7942	0.8877	0.9304	0.991	1	1
lrx-us9-laad count				0.7479	0.7479	0.7479	0.7479	0.7479	0.7778	0.8513	0.945	0.9709	1
pharmetrics count				0.6995	0.6995	0.6995	0.8345	0.8974	0.947	0.9728	1	1	1
ibm mdcr count				0.628	0.628	0.7893	0.9673	0.9876	1	1	1	1	1
ibm ccae count				0.6932	0.6932	0.6932	0.8571	0.9107	0.9583	0.9796	1	1	1
va-omop count				0.5996	0.5996	0.5996	0.8615	0.9136	0.9652	0.9917	1	1	1
ibm mdcd count				0.6374	0.6374	0.7648	0.9527	0.9827	1	1	1	1	1
optum extended dod count				0.5983	0.5983	0.5983	0.8224	0.8896	0.9518	0.9814	1	1	1
lpdau count				0.9885	1	1	1	1	1	1	1	1	1
severance cdm count				0.6628	0.8327	0.9671	1	1	1	1	1	1	1
tmudb count				0.7468	0.8949	0.9894	1	1	1	1	1	1	1
jmdc count				0.7305	0.7686	0.9412	1	1	1	1	1	1	1
france da count				0.8336	0.9176	0.9874	1	1	1	1	1	1	1
german da count				0.8493	0.8493	0.9402	0.9981	1	1	1	1	1	1

Ratio of covariates with subject counts <x

STUDY: GLP1

COHORT: DPP4 USERS

Cohort comparisons across database – all DPP4												
				-1	<100	<1000	<10,000	<20,000	<50,000	<100,000	<500,000	,000,000	>1M
optum ehr count				0.5985	0.5985	0.5985	0.7633	0.8372	0.9126	0.9498	1	1	1
healthverity count				0.7041	0.7041	0.7041	0.7998	0.8726	0.9316	0.9621	0.9989	1	1
lrx-us9-laad count				0.762	0.762	0.762	0.762	0.762	0.8395	0.8965	0.9663	0.985	1
pharmetrics count				0.7204	0.7204	0.7204	0.9201	0.9542	0.9821	0.9953	1	1	1
ibm mdcr count				0.6883	0.6883	0.7415	0.9495	0.9725	0.9942	1	1	1	1
ibm ccae count				0.7328	0.7328	0.7328	0.8876	0.929	0.9675	0.9844	1	1	1
va-omop count				0.6344	0.6344	0.6344	0.8585	0.912	0.9589	0.9868	1	1	1
ibm mdcd count				0.6694	0.6694	0.8093	0.9641	0.9888	0.9999	1	1	1	1
optum extended dod count				0.6449	0.6449	0.6449	0.868	0.9201	0.9661	0.989	1	1	1
lpdau count				0.8768	0.9331	1	1	1	0.9999	1	1	1	1
severance cdm count				0.6046	0.6046	0.8465	0.9854	1	0.9999	1	1	1	1
tmudb count				0.7428	0.7428	0.9102	0.9954	1	0.9999	1	1	1	1
jmdc count				0.7739	0.7739	0.902	0.9904	1	0.9999	1	1	1	1
france da count				0.8524	0.9222	0.9854	1	1	0.9999	1	1	1	1
german da count				0.8455	0.8455	0.9229	0.9891	0.9985	0.9999	1	1	1	1

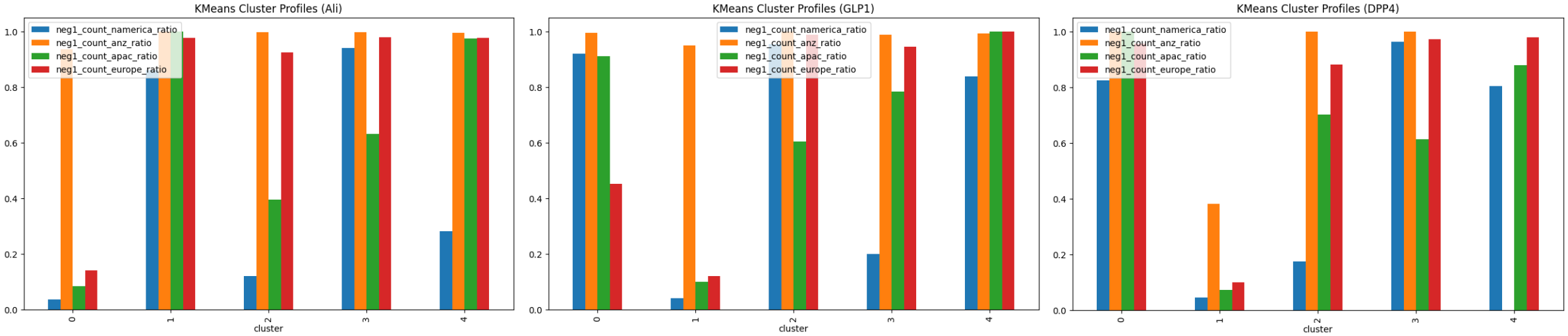
Then we looked for clusters of patterns

- Based on ratio of covariate availability

$$\frac{n(databases\ with\ covariate)}{n(all\ regional\ databases)}$$

- Then conducted kMeans clustering
  - average (mean) of coverage ratio for all covariates inside that cluster

Ran for each cohort (ALI, GLP1, DPP4)



Higher bar

=

more missingness in that region for covariates in that cluster

Across the 5 clusters, we found some patterns

	ALI		GLP1		DPP4	
	Covariates are…	**n**	**Covariates are…**	**n**	**Covariates are…**	**n**
0	Mostly -1 in ANZ only	2,091	Severely missing in NAm, ANZ, APAC Moderately missing in EU	637	Severely missing across NAm, ANZ, APAC, EU	4,800
1	Almost always -1 everywhere	7,567	Severely missing in ANZ Mostly available in NAm, APAC, SU	1,732	Mostly available in NAm, APAC, EU Moderately missing in ANZ	1,676
2	Severely missing in ANZ/EU Moderate missing in APAC Mostly present in N.AM	1,370	Severely missing in NAm, ANZ, EU Moderately missing in APAC	3,976	Mostly available in NAm Severely missing in ANZ, APAC, EU	1,970
3	Severely missing in N.Am, ANZ, EU Moderately missing in APAC	5,883	Mostly available in NAm Severely missing in ANZ, APAC, EU	2,403	Severely missing in NAm, ANZ, EU Moderately missing from APAC	4,736
4	Severely missing in ANZ, APAC, EU Moderately missing in N.Am	3,113	Severely missing in NAm, ANZ, APAC, EU	4,188	Entirely available in ANZ Severely missing in NAm, APAC, EU	669
		20,024		**12,936**		**13,851**

>30% of each cohort have covariates -1

	ALI		GLP1		DPP4	
	Covariates are…	**n**	**Covariates are…**	**n**	**Covariates are…**	**n**
0	Mostly -1 in ANZ only	2,091	Severely missing in NAm, ANZ, APAC Moderately missing in EU	637	Severely missing across NAm, ANZ, APAC, EU	4,800
1	Almost always -1 everywhere	7,567	Severely missing in ANZ Mostly available in NAm, APAC, SU	1,732	Mostly available in NAm, APAC, EU Moderately missing in ANZ	1,676
2	Severely missing in ANZ/EU Moderate missing in APAC Mostly present in N.AM	1,370	Severely missing in NAm, ANZ, EU Moderately missing in APAC	3,976	Mostly available in NAm Severely missing in ANZ, APAC, EU	1,970
3	Severely missing in N.Am, ANZ, EU Moderately missing in APAC	5,883	Mostly available in NAm Severely missing in ANZ, APAC, EU	2,403	Severely missing in NAm, ANZ, EU Moderately missing from APAC	4,736
4	Severely missing in ANZ, APAC, EU Moderately missing in N.Am	3,113	Severely missing in NAm, ANZ, APAC, EU	4,188	Entirely available in ANZ Severely missing in NAm, APAC, EU	669
		20,024		**12,936**		**13,851**

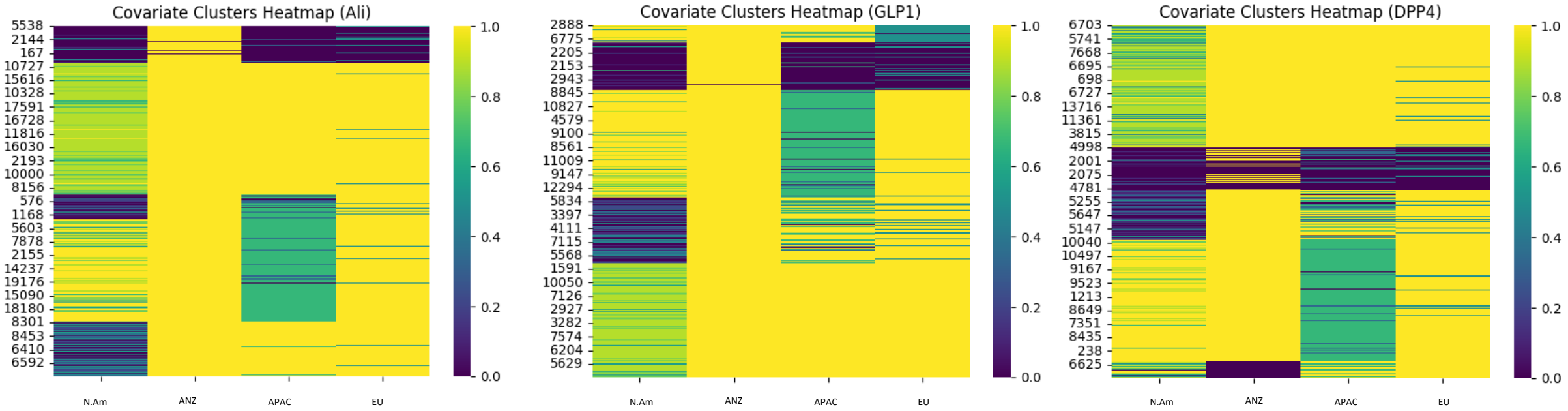
(2) 5-20% available in NAm but not anywhere else

	ALI		GLP1		DPP4	
	Covariates are…	**n**	**Covariates are…**	**n**	**Covariates are…**	**n**
0	Mostly -1 in ANZ only	2,091	Severely missing in NAm, ANZ, APAC Moderately missing in EU	637	Severely missing across NAm, ANZ, APAC, EU	4,800
1	Almost always -1 everywhere	7,567	Severely missing in ANZ Mostly available in NAm, APAC, SU	1,732	Mostly available in NAm, APAC, EU Moderately missing in ANZ	1,676
2	Severely missing in ANZ/EU Moderate missing in APAC Mostly present in N.AM	1,370	Severely missing in NAm, ANZ, EU Moderately missing in APAC	3,976	Mostly available in NAm Severely missing in ANZ, APAC, EU	1,970
3	Severely missing in N.Am, ANZ, EU Moderately missing in APAC	5,883	Mostly available in NAm Severely missing in ANZ, APAC, EU	2,403	Severely missing in NAm, ANZ, EU Moderately missing from APAC	4,736
4	Severely missing in ANZ, APAC, EU Moderately missing in N.Am	3,113	Severely missing in NAm, ANZ, APAC, EU	4,188	Entirely available in ANZ Severely missing in NAm, APAC, EU	669
		20,024		**12,936**		**13,851**

(3) Missing everywhere except APAC

	ALI		GLP1		DPP4	
	Covariates are…	**n**	**Covariates are…**	**n**	**Covariates are…**	**n**
0	Mostly -1 in ANZ only	2,091	Severely missing in NAm, ANZ, APAC Moderately missing in EU	637	Severely missing across NAm, ANZ, APAC, EU	4,800
1	Almost always -1 everywhere	7,567	Severely missing in ANZ Mostly available in NAm, APAC, SU	1,732	Mostly available in NAm, APAC, EU Moderately missing in ANZ	1,676
2	Severely missing in ANZ/EU Moderate missing in APAC Mostly present in N.AM	1,370	Severely missing in NAm, ANZ, EU Moderately missing in APAC	3,976	Mostly available in NAm Severely missing in ANZ, APAC, EU	1,970
3	Severely missing in N.Am, ANZ, EU Moderately missing in APAC	5,883	Mostly available in NAm Severely missing in ANZ, APAC, EU	2,403	Severely missing in NAm, ANZ, EU Moderately missing from APAC	4,736
4	Severely missing in ANZ, APAC, EU Moderately missing in N.Am	3,113	Severely missing in NAm, ANZ, APAC, EU	4,188	Entirely available in ANZ Severely missing in NAm, APAC, EU	669
		20,024		**12,936**		**13,851**

Visualisation of availability



- Each line represents the likelihood of that covariate having data in each database
- Each column is a region
- Darker colour = more likely to be available in databases

Results from FQ study (?)

2/2

Attempting it on the 2023 FQ <> AAA SOS

- 9 databases – 8 N.Am, 1 JMDC
- The ones from the SOS challenge, not the nice big proper one with 15 databases from later

- Assess risk of AAA following use of fluoroquinolones

Attempting it on the 2023 FQ <> AAA SOS

- Did not have diagnostics info for JMDC so could not compare
- Really N. Am. vs Japan (and Japan was very small)

But what does it say anyway if I run a similar analysis for clusters?

Partially replicating patterns

- ~56-60% NaN across all dbs
  - Can’t confirm across regions
- JMDC NaN slightly higher (+3-4%) than N.Am for 3 of 4 cohorts

Tried with the proper bigger FQ <> AAA study

- Couldn’t download the data from ShinyApp
- So we probably need a proper study package to deploy if we want to do this properly

And so?

In summation

1. Many global network studies are primarily centered on N. Am data due to sheer overwhelming size
   - Even before diagnostics
2. It **seems** that regional cohorts are failing from two aspects: covariate balance, and empirical equipoise
3. When looking at covariate balance, a quick analysis from GLP1 study seems to support some missingness patterns

Next steps?

1. More exploration
   - EASE, negative control outcomes
   - Need to do covariate balance on more studies if possible
2. Need to firm up protocol after exploration
3. Need access to databases for initial exploration of other studies?
4. Need methodological and statistical validation

Recruiting

- One (1) OHDSI stats expert
- One (1) big project data owner who has:
  - Finished their project
  - Has access to database
  - Has multiple regions in the project (before and after diagnostics)

Thank you very much!

Evelyn Goh | e0983111@u.nus.edu