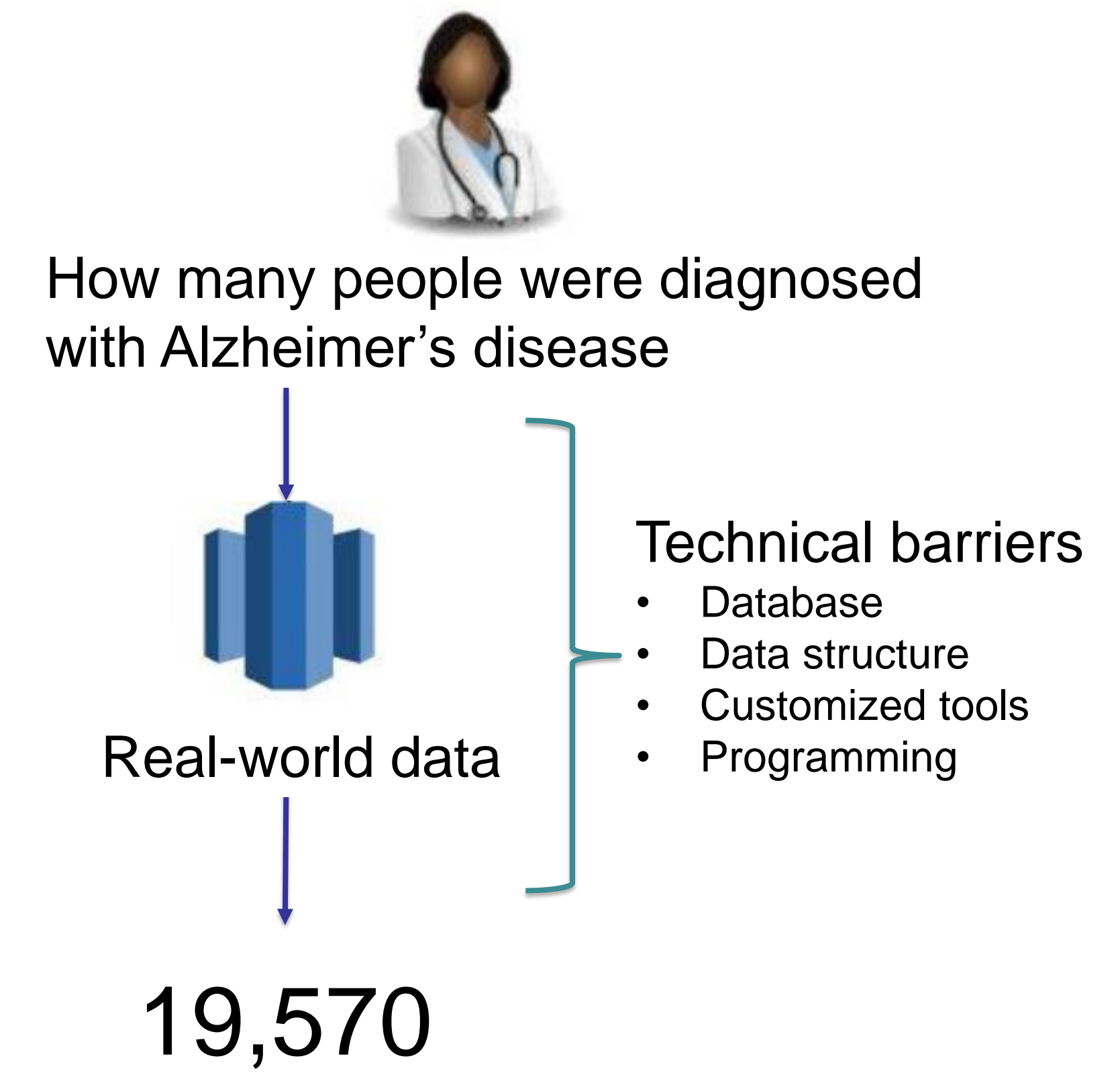


Machine Learning-assisted query and information retrieval system on real-world data

PRESENTER: Miguel Romero Calvo  
Yupeng Li

Introduction



Workflow

1. User Input

Main

Correct detection

Correct code map

Query: How many people were diagnosed with Alzheimer's disease

Detect

Execute

2. Entity recognition by Amazon Comprehend Medical (CM)

The following key entities have been detected:

How many people were diagnosed with Alzheimer's disease CONDITION

Drugs and Conditions will be respectively replaced by the following RxNorm & ICD10 codes:

How many people were diagnosed with G30.9 CONDITION

3. Code correction (optional)

- Add/remove code

Main

Correct detection

Correct code map

Add detection

Write name Aspirin 30Mg Category: DRUG Highlight

Remove detection

Name: Alzheimer's disease (category: Remove

- Change code

Main

Correct detection

Correct code map

Drug

Condition

Show drug info

Show condition info

Map to RxNorm code

Map to ICD10 code

Update drug

Update condition

TEXT: Alzheimer's disease

DISAMBIGUATED TO: G30.9

INFERRED OPTIONS

Score	ICD10 Code	Name
1. (0.739)	G30.9	Alzheimer's disease, unspecified
2. (0.631)	G20	Parkinson's disease
3. (0.625)	G30	Alzheimer's disease
4. (0.528)	G30.8	Other Alzheimer's disease
5. (0.514)	G30.0	Alzheimer's disease with early onset

Directly query OMOP-CDM datasets with natural language - making the data querying as easy as a google search



Take a picture or click the link to learn more details

<https://github.com/OHDSI/NOTOS>

4. Text pre-processing

How many people were diagnosed with <ARG-CONDITION><0>

5. Text-to-SQL translation

```
SELECT COUNT( DISTINCT con1.person_id ) FROM ( <SCHEMA>.condition_occurrence con1 JOIN <CONDITION-TEMPLATE><ARG-CONDITION><0> ON con1.condition_concept_id=concept_id);
```

6. SQL post-processing

```
SELECT COUNT( DISTINCT con1.person_id ) FROM (omop_synpuf_110k.condition_occurrence con1 JOIN ( SELECT descendant_concept_id AS concept_id FROM (SELECT * FROM (SELECT concept_id_2 FROM ( (SELECT concept_id FROM omop_synpuf_110k.concept WHERE vocabulary_id='ICD10CM' AND ( concept_code='G30.9' )) JOIN ( SELECT concept_id_1, concept_id_2 FROM omop_synpuf_110k.concept_relationship WHERE relationship_id='Maps to' ) ON concept_id=concept_id_1 ) ) JOIN omop_synpuf_110k.concept ON concept_id_2=concept_id) JOIN omop_synpuf_110k.concept_ancestor ON concept_id=ancestor_concept_id ) ON con1.condition_concept_id=concept_id) ;
```

7. SQL execution

Request run successfully

count
0 19570

Text-to-SQL model

- Compiled a list of high-frequency questions and the corresponding SQL queries
- Generated multiple real-world variations for each pre-defined questions
- Fine-tuned the T5 model for text-to-SQL translation

Model accuracy for in-scope questions

Metric	Validation	Test
Exact Match	0.9926	0.9920
Execution	0.9999	0.9999

Miguel Romero Calvo, Yupeng Li, Tesfagabir Meharizghi, Weilin Meng, Selvan Senthivel, Saman Sarraf, Lin Lee Cheong

