Welcome to the Sentinel Innovation Center Webinar Series

The webinar will begin momentarily

Please visit www.sentinelinitiative.org for recordings of past sessions and details on upcoming webinars.

Note: closed-captioning for today's webinar will be available on the recording posted at the link above.

Sentinel

Harmonizing Electronic Health Records from Heterogeneous Systems via Automated Translation of Medical Concepts

Xu Shi
Department of Biostatistics, University of Michigan

August 19, 2020

Andrew Beam
Assistant Professor
Dept. of Epidemiology
Harvard University

Tianxi Cai
Professor
Dept. of Biostatistics
Dept of Biomedical Informatics
Harvard University

Patrick Heagerty
Professor
Dept. of Biostatistics
University of Washington

Xiaoou Li
Assistant Professor
School of Statistics
University of Minnesota

Hristina Pashova
Biostatistics
Corcept Therapeutics

W UW Medicine
SCHOOL OF MEDICINE

HDSI | Harvard Data Science Initiative

HARVARD MEDICAL SCHOOL | DEPARTMENT OF Biomedical Informatics

Demand for health information exchange

- **Goodbye "meaningful use", hello "promoting interoperability"**
  - Centers for Medicare & Medicaid Services (CMS) renamed EHR incentive program
  - To advance integration and sharing of healthcare data

BREAKING: CMS Finalizes "Promoting Interoperability" Rule

August 2, 2018 by Rajiv Leventhal | Reprints

The federal agency has finalized 90-day reporting periods for 2019 and 2020, while requiring 2015 CEHRT starting in 2019

Just three months after issuing a proposal, the Centers for Medicare & Medicaid Services (CMS) has finalized a rule late this afternoon that will overhaul the meaningful use program with a core emphasis on advancing health data exchange among providers.

The final rule issued today makes updates to Medicare payment policies and rates under the Inpatient Prospective Payment System (IPPS) and the Long-Term Care Hospital (LTCH) Prospective Payment System (PPS) that will incentivize value-based, quality care at these facilities.

"We're excited to make these changes to ensure care will focus on the patient, not on needless paperwork," CMS Administrator Seema Verma said in a statement. "We've listened to patients and their doctors who urged us to remove the obstacles getting in the way of quality care and positive health outcomes. Today's final rule reflects public feedback on CMS proposals issued in April, and the agency's patient-driven priorities of improving the quality and safety of care, advancing health information exchange and usability, and removing outdated or redundant regulations on healthcare providers to make way for innovation and greater value."

According to CMS, the rule applies to about 3,300 acute care hospitals and 420 long-term care hospitals, and will take effect Oct. 1

---

https://www.healthcare-informatics.com/article/value-based-care/breaking-cms-finalizes-promoting-interoperability-rule

Semantic interoperability: EHRs do not talk to each other

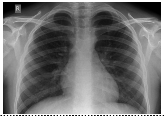
- **786.2 Cough (ICD-9)**
- **780.61 Fever (ICD-9)**
- **71010 Chest X-ray (CPT)**

lab_type	result	units
HEMOGRAM PLATELET COUNT	154	x10 3/uL
HEMOGRAM RED BLOOD CELL COUNT	4.59	x10 6 /uL
HEMOGRAM RED CELL DISTRIBUTION WIDTH	20.4	%
HEMOGRAM WHITE BLOOD CELL COUNT	6.6	x10 3 /uL

--- PHYSICIAN NOTE ---

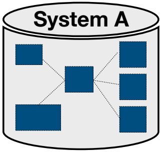
History of Present Illness

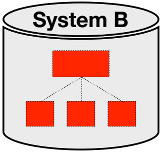
Presenting problem started 5 days ago. History comes from patient. Able to get a good history. Presents with symptoms suggestive of a lower GI bleed. This is a new problem, with no prior history of similar episodes. Symptoms developed over several days. Describes stool as black in color. Passing mucoid stools. Streaks of blood noted in stool. Saw gross blood in the bowel movement. Not on iron or Pepto bismol. Estimated blood loss is less than 50 cc. No history of prior GI bleeding. No history orthostatic symptoms, excessive fatigue, or syncope.



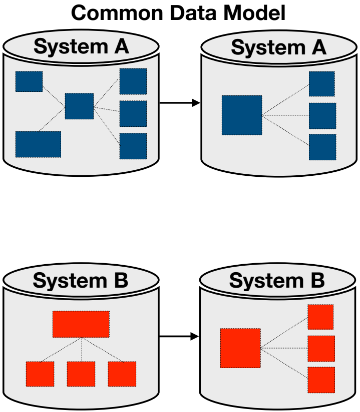
- **Standardized medical code for billing** ☺
  - Common language across healthcare providers and insurers
- **Inconsistent coding in practice** ☹
  - System A use 786.05: **shortness of breath**
  - System B use 786.09: other **dyspnea** and respiratory abnormality

What are potential challenges?

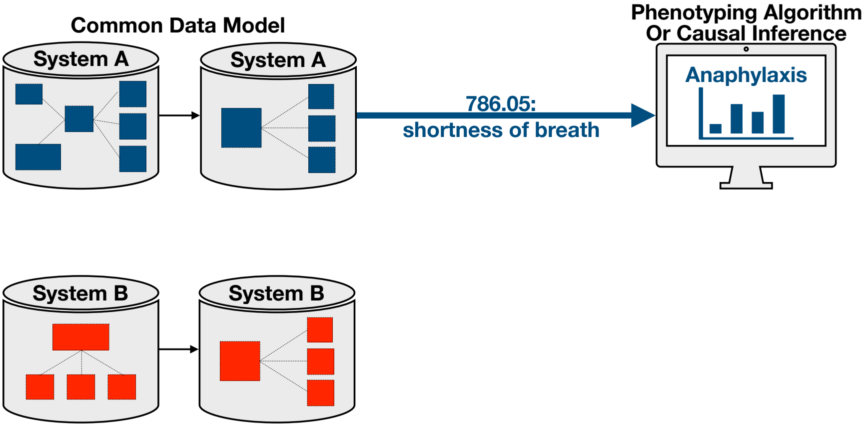




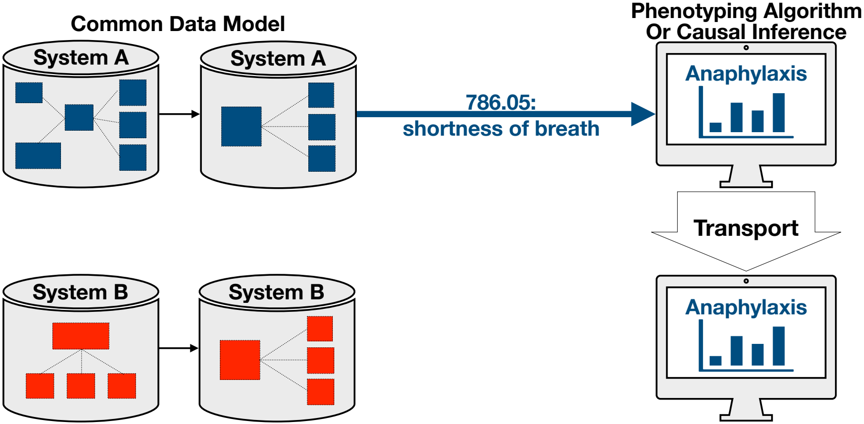
What are potential challenges?



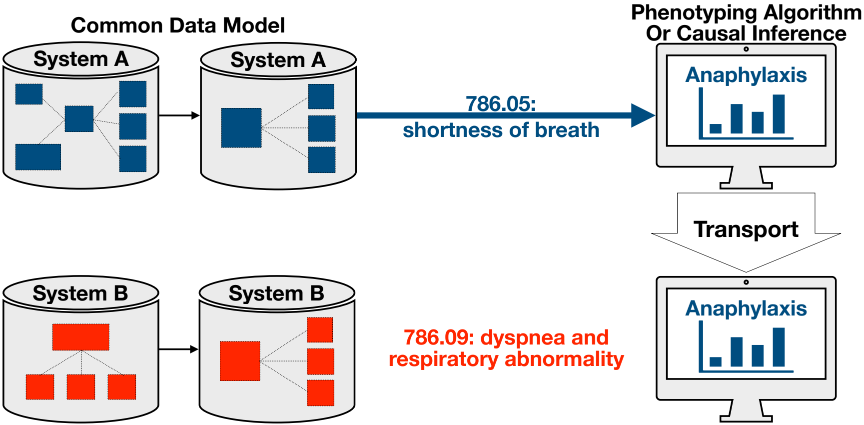
What are potential challenges?



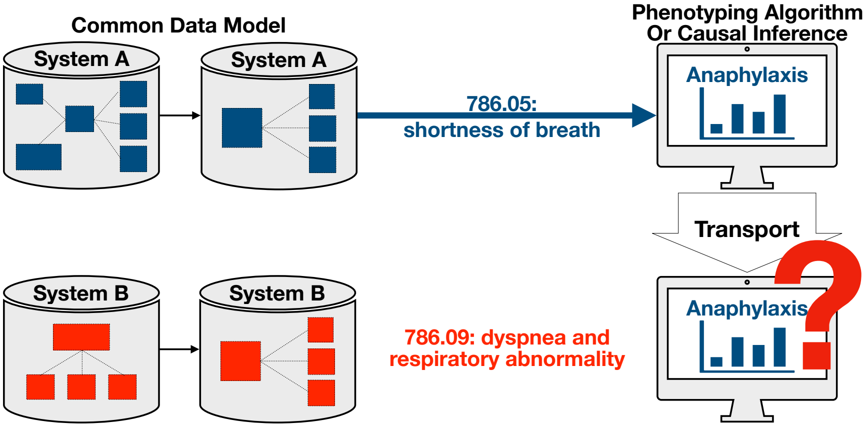
What are potential challenges?



What are potential challenges?



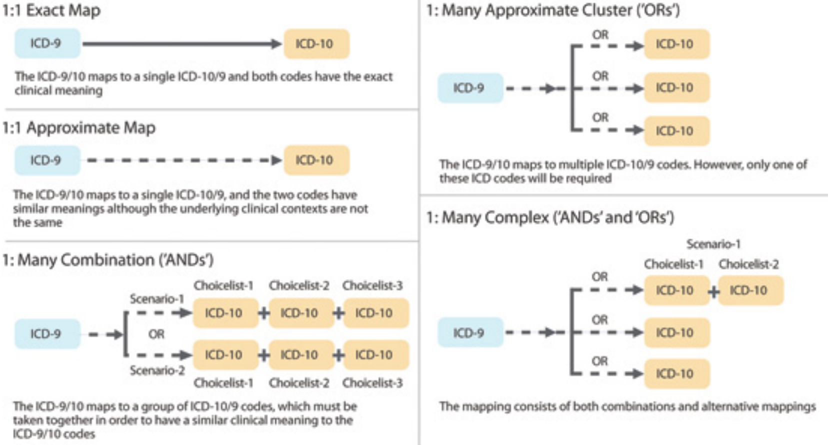
What are potential challenges?



- **Performance of phenotyping algorithm can dramatically drop**
- **Causal inference can fail due to incorrect confounding adjustment**

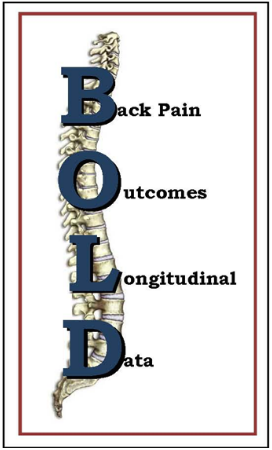
Manual mapping is imprecise

- **General equivalence mapping (GEM): ICD-9 (10k) ⇆ ICD-10 (60k)**
  - Approximate mappings with multiple scenarios: data merged with adhoc decisions



Data Driven Mapping of Medical Codes

The Back pain Outcomes using Longitudinal Data study

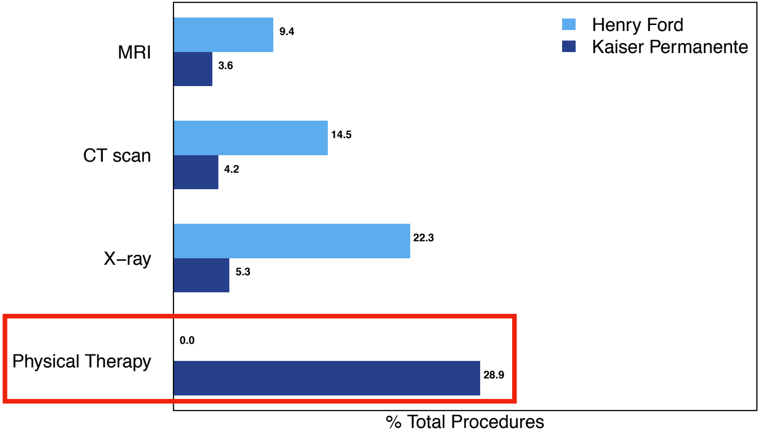


- **The elderly with back pain**
  - 5000 patients $\geq$ 65 years old
  - Cost-effectiveness of early diagnostic imaging
- **EHR data from three sites:**
  - Henry Ford Health System in Detroit
  - Kaiser Permanente Northern California
  - Harvard Vanguard in Boston

Jarvik JG, et. al. (2012), Jarvik JG, et. al. (2015)

Data quality check before pulling EHR data from study sites

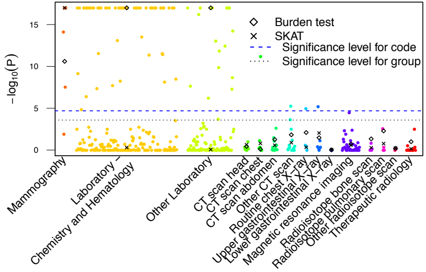
- **Compare use of CPT codes between study sites**

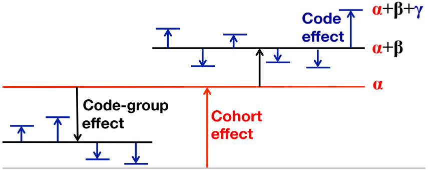


- **Question:** can we scan for variation in the endorsement of all medical codes to identify such data quality issue?

Detect and quantify coding differences under a hierarchical structure

- **Code grouping** e.g. PheWAS (phenome-wide association studies)
  CCS (Clinical Classifications Software)





Testing:
Group-wise association test
codes in a group ⇔ genetic variants in a region

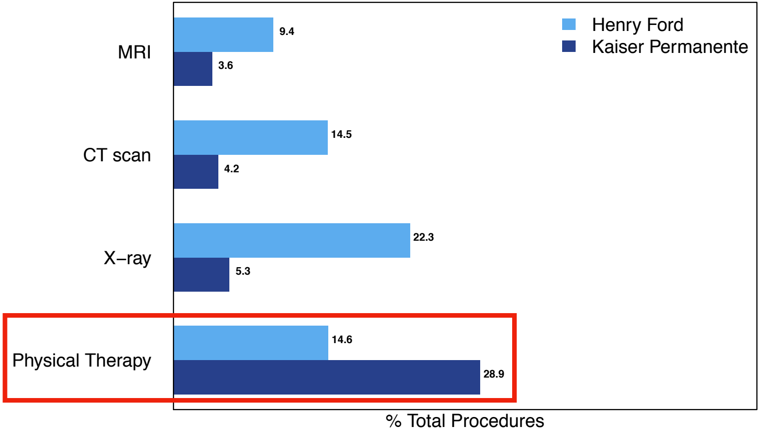
Estimation:
hierarchical shrinkage
post-regularization inference

CPT-SCAN: `https://xu-rita-shi.shinyapps.io/CPT_SCAN/`

Shi et. al. (2017)

Further investigation into observed differences in code endorsement

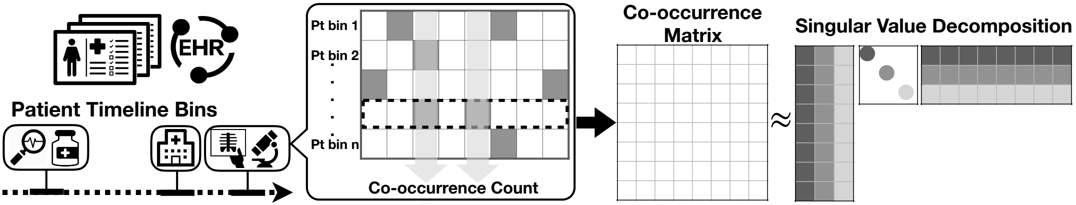
- **Compare use of CPT codes between study sites**



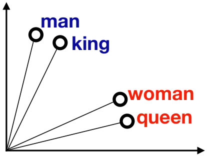
- Henry Ford uses a generic code "*HF0PT*" for physical therapy

Can data tell me “*HF0PT*” = “physical therapy”?

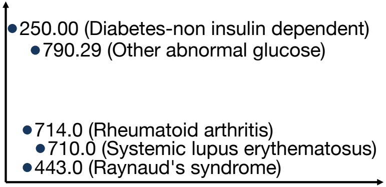
- **Co-occurrence: semantic information from the context**
  - “HF0PT” is surrounded by codes for pain-related diseases or treatments
  - “Physical therapy” often appears in such a context



Computers learn the meaning of a word from its context



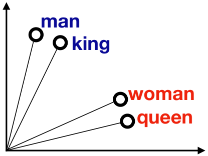
- `word2vec`: **represent a word as a vector**
  - Learn semantic relationship from co-occurrence
  - Words with similar meanings are close



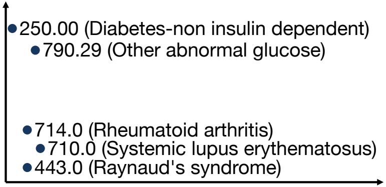
- `code2vec`: **represent a code as a vector**
  - Code ⇔ word; Healthcare system ⇔ language
  - Interpret meaning of codes in clinical practice setting

Mikolov et. al. (2013), Levy & Goldberg (2014), Choi et. al. (2016), Beam et. al. (2018)

Computers learn the meaning of a word from its context



- `word2vec`: **represent a word as a vector**
  - Learn semantic relationship from co-occurrence
  - Words with similar meanings are close



- `code2vec`: **represent a code as a vector**
  - Code ⇔ word; Healthcare system ⇔ language
  - Interpret meaning of codes in clinical practice setting

Question: can we infer a mapping between two sets of code-vectors learned from two healthcare systems, respectively?

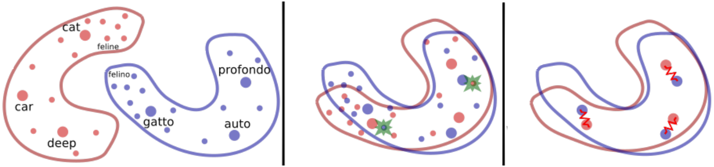
Mikolov et. al. (2013), Levy & Goldberg (2014), Choi et. al. (2016), Beam et. al. (2018)

From language translation to code mapping

- **Inconsistent objectives in language translation with `word2vec`**

Generate word vectors ➡ **Space alignment** ➡ **Infer a mapping**

max **inner product** | min ℓ_2 **distance** | max **cosine**



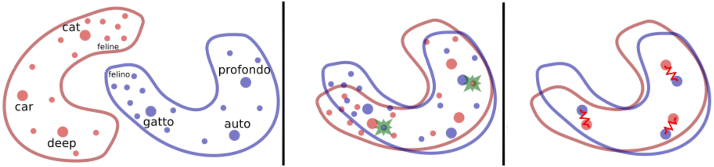
Xing et. al. (2015), Conneau et. al. (2018), Shi et. al. (2018)

From language translation to code mapping

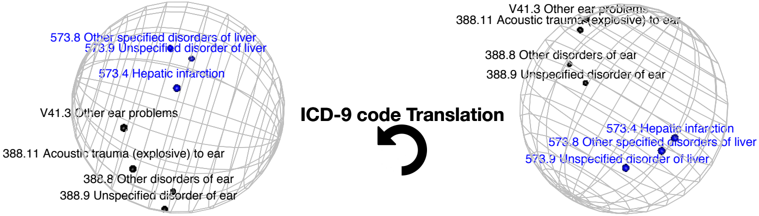
- **Inconsistent objectives in language translation with `word2vec`**

Generate word vectors ➡ **Space alignment** ➡ **Infer a mapping**

max **inner product** | min ℓ_2 **distance** | max **cosine**



- **Length normalization: semantic information is in the direction**



Partners HealthCare | **Veterans Health Administration**

Xing et. al. (2015), Conneau et. al. (2018), Shi et. al. (2018)

How do statisticians think about language translation?

$\mathbb{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]^\mathsf{T}_{n\times p}$, $\mathbb{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_n]^\mathsf{T}_{n\times p}$: n vectors, each $\mathbf{X}_i, \mathbf{Y}_i \in R^p$

- n: number of codes
- p: dimension of code-vectors

$\mathbb{Y}_{n\times p}$

1	1	$\mathbf{Y}_1$
2	2	$\mathbf{Y}_2$
3	3	$\mathbf{Y}_3$
4	4	$\mathbf{Y}_4$
5	5	$\mathbf{Y}_5$

$\mathbb{X}_{n\times p}$

1	1	$\mathbf{X}_1$
2	2	$\mathbf{X}_2$
3	3	$\mathbf{X}_3$
4	4	$\mathbf{X}_4$
5	5	$\mathbf{X}_5$

How do statisticians think about language translation?

- **Classical regression**

$$\mathbb{Y}_{n\times p} = \mathbb{X}_{n\times p}\mathbb{W}_{p\times p} + \mathbb{U}_{n\times p}$$

$\mathbf{Y}_i \sim \mathbf{X}_i$ correctly linked

$$\underset{\mathbb{Y}_{n\times p}}{\begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \\ 4 & 4 \\ 5 & 5 \end{bmatrix}} = \underset{\mathbb{X}_{n\times p}}{\begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \\ 4 & 4 \\ 5 & 5 \end{bmatrix}} \underset{\mathbb{W}_{p\times p}}{\begin{bmatrix} & \\ & \end{bmatrix}} + \mathbb{U}_{n\times p}$$

How do statisticians think about language translation?

- **Classical regression**

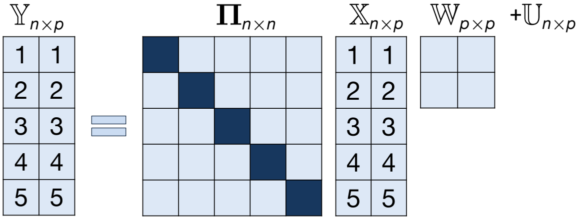
$$\mathbb{Y}_{n\times p} = \mathbb{X}_{n\times p}\mathbb{W}_{p\times p} + \mathbb{U}_{n\times p}$$

$\mathbf{Y}_i \sim \mathbf{X}_i$ correctly linked

- **Shuffled regression**

$$\mathbb{Y}_{n\times p} = \mathbf{\Pi}_{n\times n}\mathbb{X}_{n\times p}\mathbb{W}_{p\times p} + \mathbb{U}_{n\times p}$$

$\mathbf{Y}_i \sim \mathbf{X}_i$ may not correspond



Introduce a mapping matrix Π (the "dictionary")

no mismatch if $\Pi = \mathbb{I}$ is an identity matrix

How do statisticians think about language translation?

- **Classical regression**

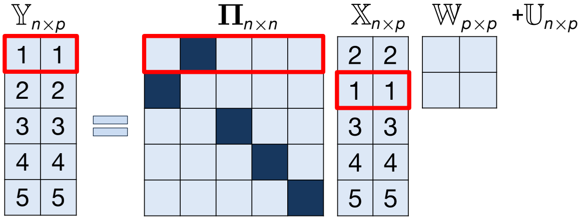
  $\mathbb{Y}_{n\times p} = \mathbb{X}_{n\times p}\mathbb{W}_{p\times p}+\mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ correctly linked

- **Shuffled regression**

  $\mathbb{Y}_{n\times p} = \mathbf{\Pi}_{n\times n}\mathbb{X}_{n\times p}\mathbb{W}_{p\times p}+\mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ may not correspond



Each row of Π is like a pointer:

match: $\mathbf{\Pi}_{i\cdot} = \mathbb{I}_{i\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_i$; **mismatch:** $\mathbf{\Pi}_{i\cdot} = \mathbb{I}_{j\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_j$

How do statisticians think about language translation?

- **Classical regression**

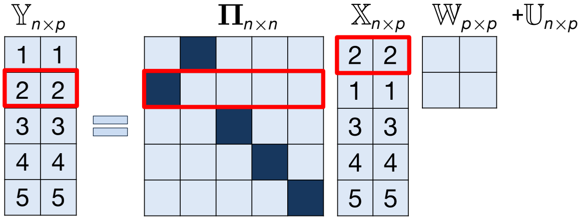
  $\mathbb{Y}_{n\times p} = \mathbb{X}_{n\times p}\mathbb{W}_{p\times p} + \mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ correctly linked

- **Shuffled regression**

  $\mathbb{Y}_{n\times p} = \mathbf{\Pi}_{n\times n}\mathbb{X}_{n\times p}\mathbb{W}_{p\times p} + \mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ may not correspond



Each row of Π is like a pointer:

match: $\Pi_{i\cdot} = \mathbb{I}_{i\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_i$; **mismatch:** $\Pi_{i\cdot} = \mathbb{I}_{j\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_j$

How do statisticians think about language translation?

- **Classical regression**

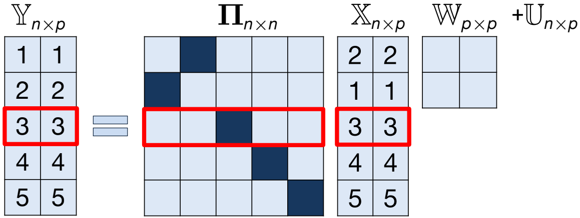
  $\mathbb{Y}_{n\times p} = \mathbb{X}_{n\times p}\mathbb{W}_{p\times p} + \mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ correctly linked

- **Shuffled regression**

  $\mathbb{Y}_{n\times p} = \mathbf{\Pi}_{n\times n}\mathbb{X}_{n\times p}\mathbb{W}_{p\times p} + \mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ may not correspond



Each row of Π is like a pointer:

match: $\Pi_{i\cdot} = \mathbb{I}_{i\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_i$; **mismatch:** $\Pi_{i\cdot} = \mathbb{I}_{j\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_j$

How do statisticians think about language translation?

- **Classical regression**

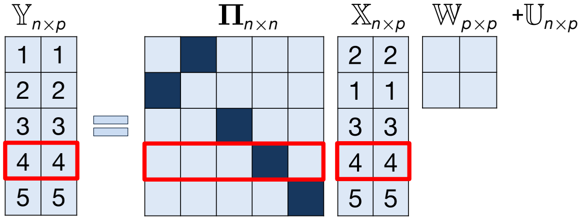
  $\mathbb{Y}_{n\times p} = \mathbb{X}_{n\times p}\mathbb{W}_{p\times p} + \mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ correctly linked

- **Shuffled regression**

  $\mathbb{Y}_{n\times p} = \mathbf{\Pi}_{n\times n}\mathbb{X}_{n\times p}\mathbb{W}_{p\times p} + \mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ may not correspond



Each row of Π is like a pointer:

match: $\Pi_{i\cdot} = \mathbb{I}_{i\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_i$; **mismatch:** $\Pi_{i\cdot} = \mathbb{I}_{j\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_j$

How do statisticians think about language translation?

- **Classical regression**

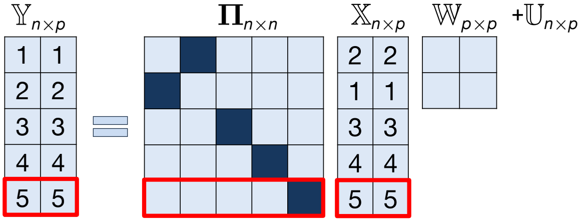
  $\mathbb{Y}_{n\times p} = \mathbb{X}_{n\times p}\mathbb{W}_{p\times p}+\mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ correctly linked

- **Shuffled regression**

  $\mathbb{Y}_{n\times p} =\mathbf{\Pi}_{n\times n}\mathbb{X}_{n\times p}\mathbb{W}_{p\times p}+\mathbb{U}_{n\times p}$

  $\mathbf{Y}_i \sim \mathbf{X}_i$ may not correspond

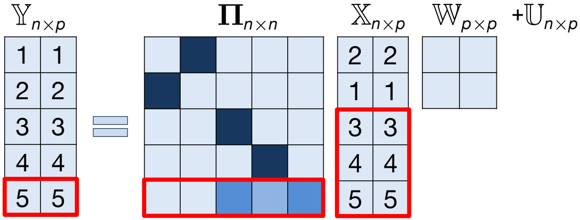


Each row of Π is like a pointer:

match: $\Pi_{i\cdot} = \mathbb{I}_{i\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_i$; **mismatch:** $\Pi_{i\cdot} = \mathbb{I}_{j\cdot} \Rightarrow \mathbf{Y}_i \sim \mathbf{X}_j$

Formulating the problem: mismatched spherical data

- $\mathbf{\Pi}$ encodes 1-to-1 and 1-to-many mapping

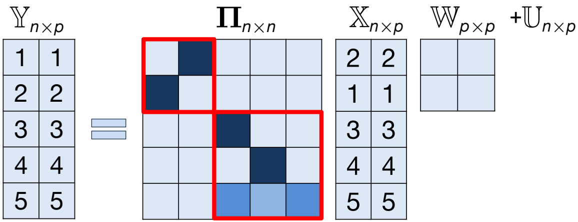


Allow for 1-to-many mapping

weight vector: $\mathbf{\Pi}_{i\cdot} = \omega$

The statistical problem: mismatched spherical data

- Π encodes 1-to-1 and 1-to-many mapping
- Assume Π is block diagonal

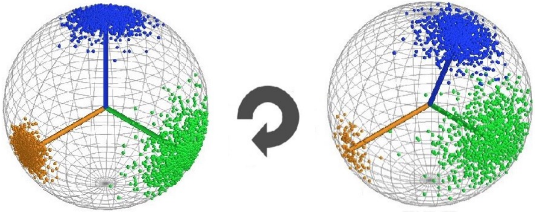


Incorporate code-group information

mismatch only occurs within group

The statistical problem: mismatched spherical data

- $\mathbf{\Pi}$ encodes 1-to-1 and 1-to-many mapping
- Assume $\mathbf{\Pi}$ is block diagonal
- $\mathbb{W}$ is an orthogonal matrix s.t. $\|\mathbb{W}\mathbf{X}_i\| = \|\mathbf{Y}_i\| = 1$

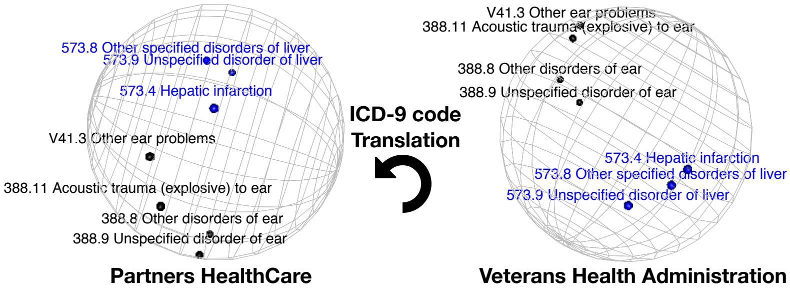


$\mathbb{W}$ rotates $\mathbb{X}$ on the sphere

Align spherical language spaces

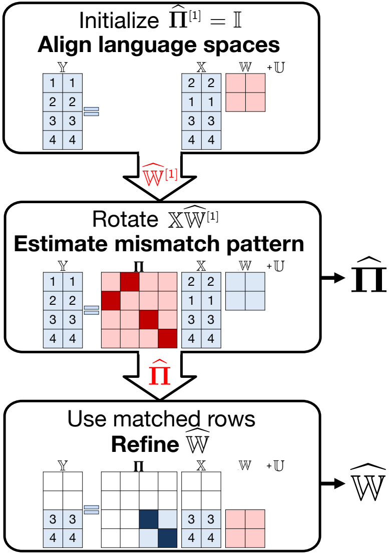
The statistical problem: mismatched spherical data

- Π encodes 1-to-1 and 1-to-many mapping
- Assume Π is block diagonal
- $\mathbb{W}$ is an orthogonal matrix s.t. $\|\mathbb{W}\mathbf{X}_i\| = \|\mathbf{Y}_i\| = 1$

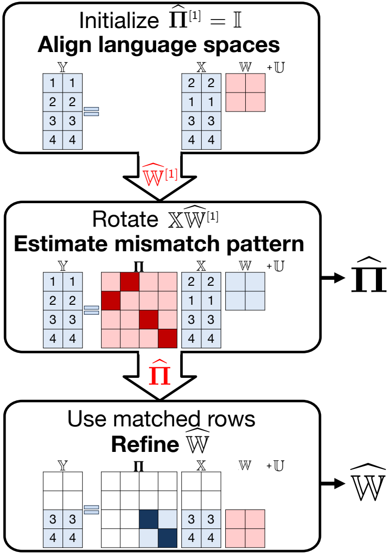


Goal: estimate $(\Pi, \mathbb{W})$ using mismatched spherical data

iSphereMAP: iterative Spherical regression MAPping



iSphereMAP: iterative Spherical regression MAPping

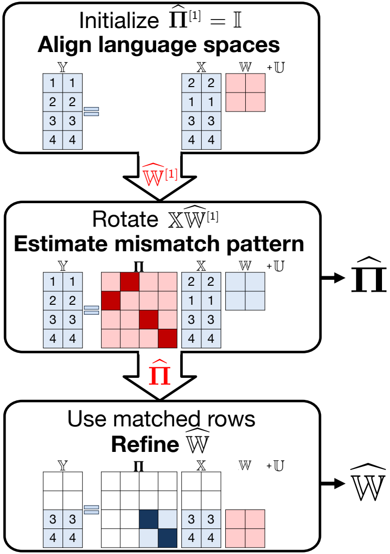


- **Find rotation via spherical regression**

$$\widehat{\mathbb{W}}^{[1]} = \underset{\mathbb{W}:\mathbb{W}\mathbb{W}^\top=\mathbb{I}_p}{\text{argmin}} \|\mathbb{Y} - \mathbb{X}\mathbb{W}\|_F^2 = UV^\top$$

where $\mathbb{X}^\top\mathbb{Y} = UDV^\top$

iSphereMAP: iterative Spherical regression MAPping



- **Find rotation via spherical regression**

$$\widehat{\mathbb{W}}^{[1]} = \underset{\mathbb{W}:\mathbb{W}\mathbb{W}^{\intercal}=\mathbb{I}_p}{\text{argmin}} \|\mathbb{Y} - \mathbb{X}\mathbb{W}\|_F^2 = UV^{\intercal}$$

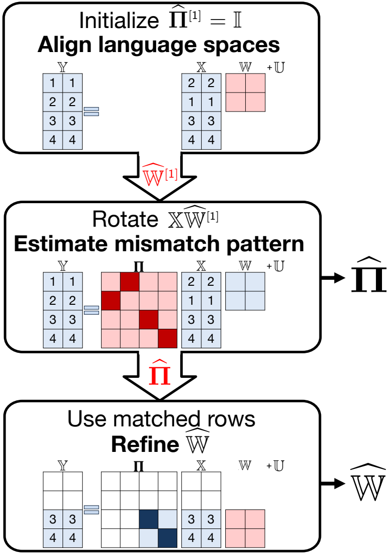
where $\mathbb{X}^{\intercal}\mathbb{Y} = UDV^{\intercal}$

- **Match a code to its nearest neighbor(s)**

$$\widetilde{\Pi}^k = \text{argmin} \|\widetilde{\mathbb{Y}}_k - \widetilde{\mathbb{X}}_k \Pi^{\intercal}\|_F^2$$

where $\widetilde{\mathbb{Y}}_k = \mathbb{Y}_k^{\intercal}$, $\widetilde{\mathbb{X}}_k = (\mathbb{X}_k \widehat{\mathbb{W}}^{[1]})^{\intercal}$

iSphereMAP: iterative Spherical regression MAPping



- **Find rotation via spherical regression**

$$\widehat{\mathbb{W}}^{[1]} = \underset{\mathbb{W}:\mathbb{W}\mathbb{W}^\top=\mathbb{I}_p}{\operatorname{argmin}} \|\mathbb{Y}-\mathbb{X}\mathbb{W}\|_F^2 = UV^\top$$

where $\mathbb{X}^\top\mathbb{Y} = UDV^\top$

- **Match a code to its nearest neighbor(s)**

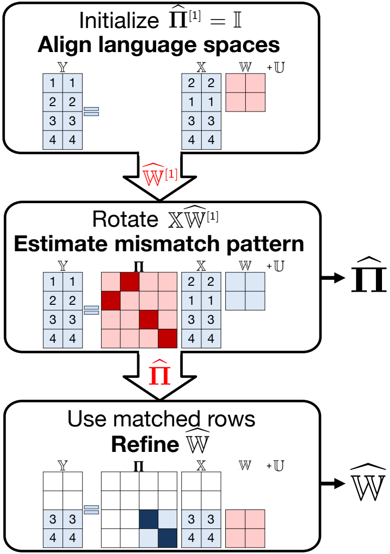
$$\widetilde{\Pi}^k = \operatorname{argmin} \|\widetilde{\mathbb{Y}}_k - \widetilde{\mathbb{X}}_k \Pi^\top\|_F^2$$

where $\widetilde{\mathbb{Y}}_k = \mathbb{Y}_k^\top$, $\widetilde{\mathbb{X}}_k = (\mathbb{X}_k\widehat{\mathbb{W}}^{[1]})^\top$

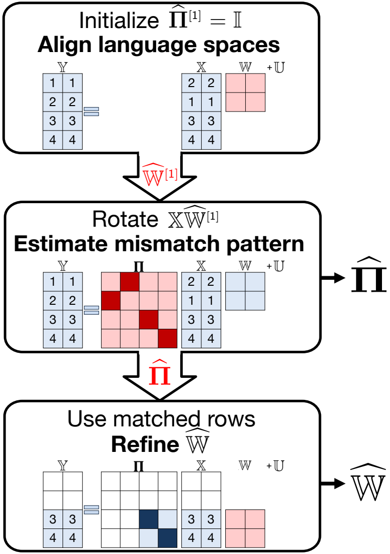
- **Refine rotation using matched data**

$$\widehat{\mathbb{W}} = \underset{\mathbb{W}:\mathbb{W}\mathbb{W}^\top=\mathbb{I}_p}{\operatorname{argmin}} \|\mathbb{Y}_{\text{match}} - \mathbb{X}_{\text{match}}\mathbb{W}\|_F^2$$

Theoretical guarantees



Theoretical guarantees

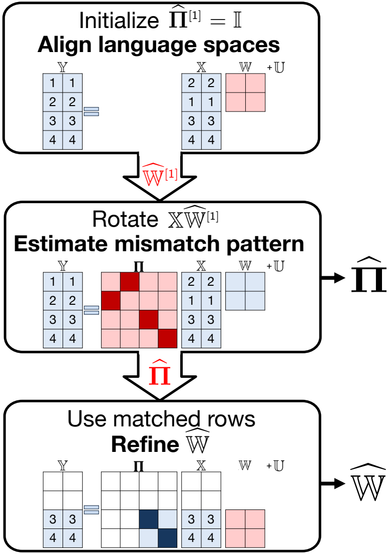


- **Is alignment insensitive to mismatch?**

  $\|\widehat{\mathbb{W}}^{[1]} - \mathbb{W}\|_F = O_p(\text{inherent noise} + \textbf{mismatch})$

  Consistency requires sparse mismatch

Theoretical guarantees



- **Is alignment insensitive to mismatch?**

  $\|\widehat{\mathbb{W}}^{[1]} - \mathbb{W}\|_F = O_p(\text{inherent noise} + \textbf{mismatch})$

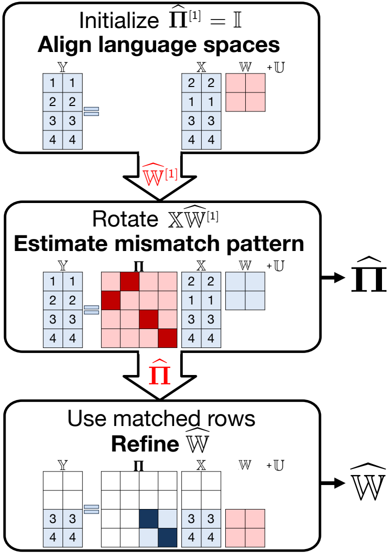
  Consistency requires sparse mismatch

- **Is code mapping correct?**

  Correctly map $\mathbf{Y}_i$ to $\mathbf{X}_j$ if one-to-one;

  Consistently estimate the weight if one-to-many

Theoretical guarantees



- **Is alignment insensitive to mismatch?**

  $\|\widehat{\mathbb{W}}^{[1]} - \mathbb{W}\|_F = O_p($inherent noise $+$ **mismatch**$)$

  Consistency requires sparse mismatch

- **Is code mapping correct?**

  Correctly map $\mathbf{Y}_i$ to $\mathbf{X}_j$ if one-to-one;

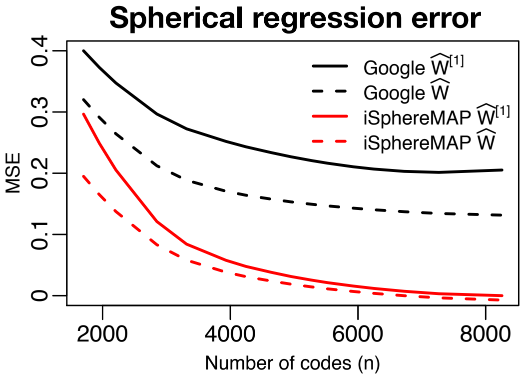
  Consistently estimate the weight if one-to-many

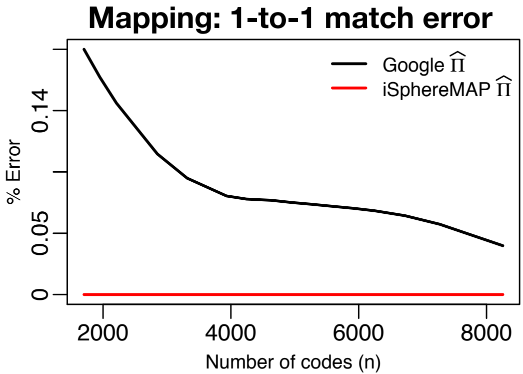
- **Can we better estimate $\mathbb{W}$?**

  $\|\widehat{\mathbb{W}} - \mathbb{W}\|_F = O_p($inherent noise$)$

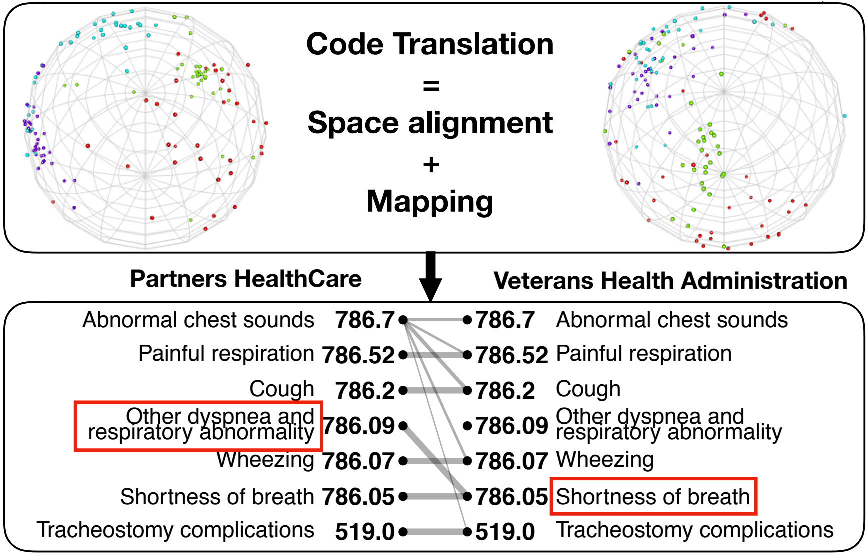
  As good as if **no mismatch** is present

Simulation: iSphereMAP vs Mikolov et. al. 2013 (Google)





Example: ICD-9 code translation between two systems



Example: ICD9-to-10 mapping for suicide and self-inflicted injuries (SSI)

Manual mapping (GEM)

Data driven (iSphereMAP)

Take home messages

- **EHRs need to be “semantically” translated before being fed into a phenotyping algorithm or statistical model**
- **Manually curated mappings are imprecise and error prone**
- **Data driven mappings are scalable and automated**
    - Based on summary of co-occurrence: does not require individual level data
    - Unsupervised: does not rely on training labels

Thank you!

Questions?

shixu@umich.edu