Natural language processing approaches to extracting patient functioning from clinical data

Denis Newman-Griffis

Stanford Center for Population Health Sciences
March 16, 2018

NIH National Institutes of Health Clinical Center

The Ohio State University

Washington University in St. Louis

Social Security USA Administration

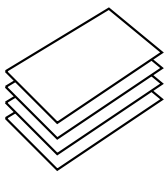
Outline

Motivating example: Disability adjudication support

New claimant applying for disability

Images from http://openclipart.org

Motivating example: Disability adjudication support



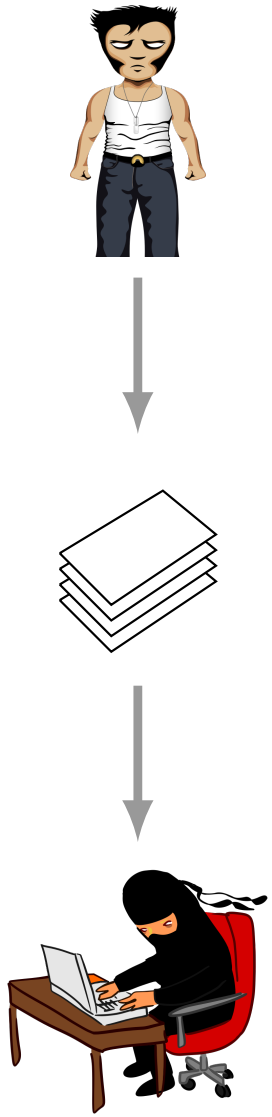
New claimant applying for disability

Application includes

- Allegations
- *(Optionally)* Self-collected medical evidence

Images from http://openclipart.org

Motivating example: Disability adjudication support



New claimant applying for disability

Application includes

- Allegations
- *(Optionally)* Self-collected medical evidence

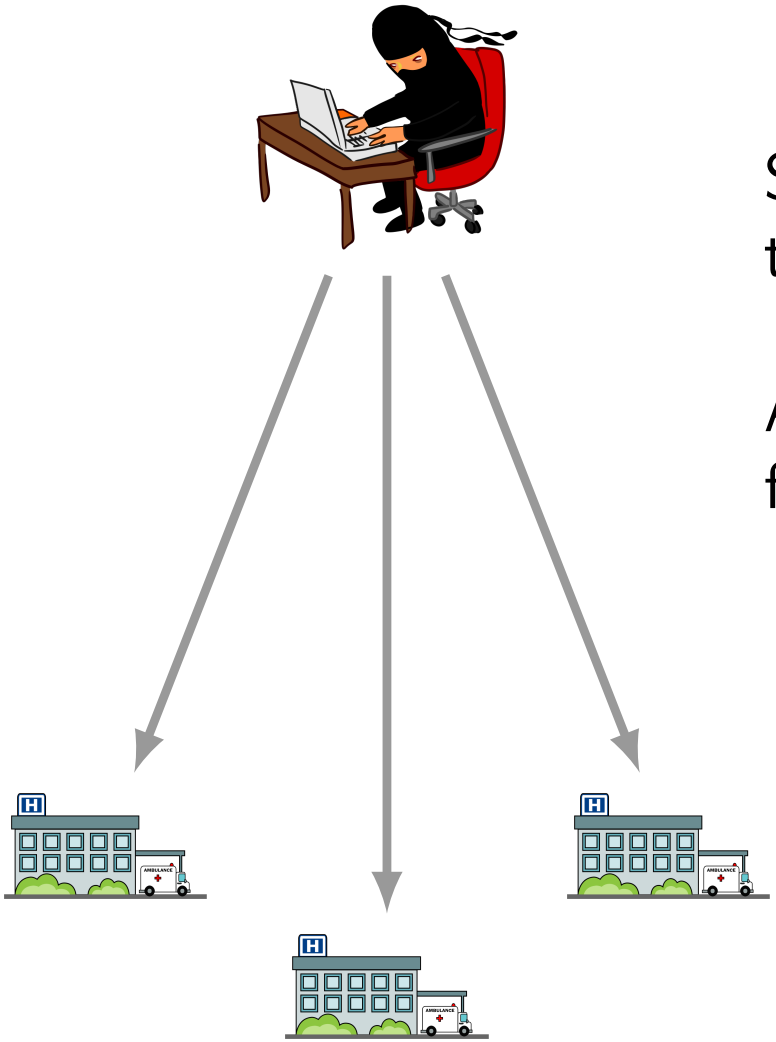
Images from http://openclipart.org

Motivating example: Disability adjudication support

SSA needs further evidence to decide the case

Images from http://openclipart.org

Motivating example: Disability adjudication support



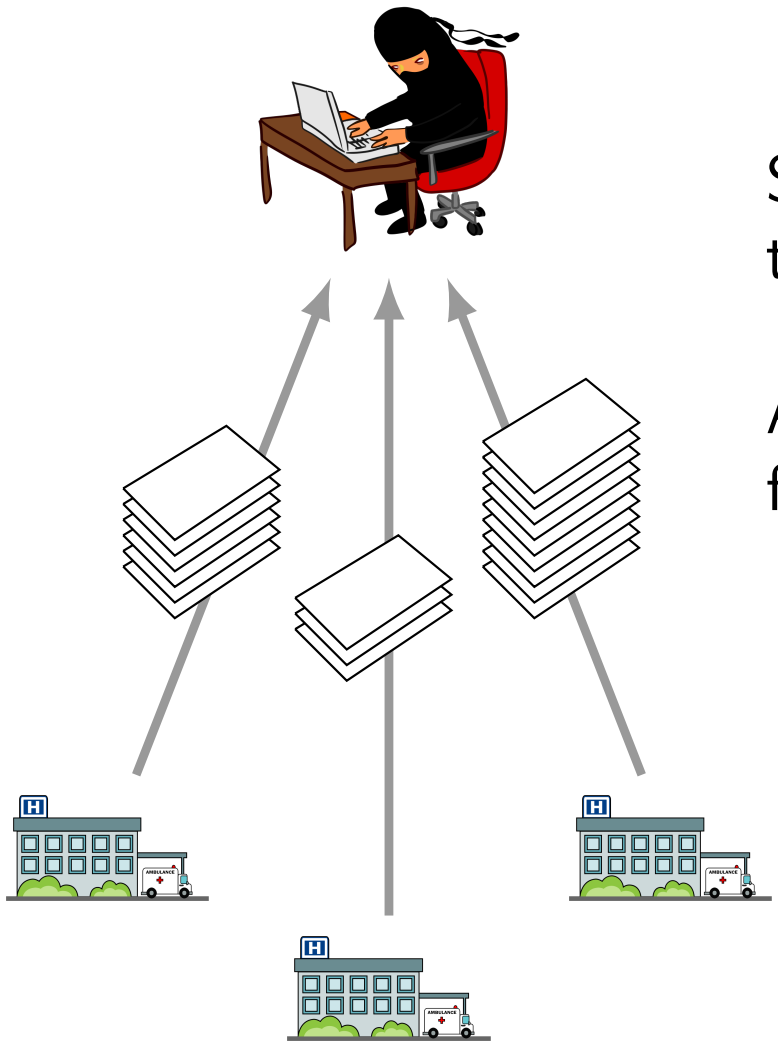
SSA needs further evidence to decide the case

Adjudicator contacts care facilities for related claimant records

- Here, assume 3 different ERs

Images from http://openclipart.org

Motivating example: Disability adjudication support



SSA needs further evidence to decide the case

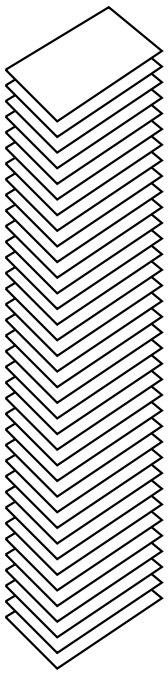
Adjudicator contacts care facilities for related claimant records

- Here, assume 3 different ERs

Care facilities send back hundreds of pages of records

Images from http://openclipart.org

Motivating example: Disability adjudication support

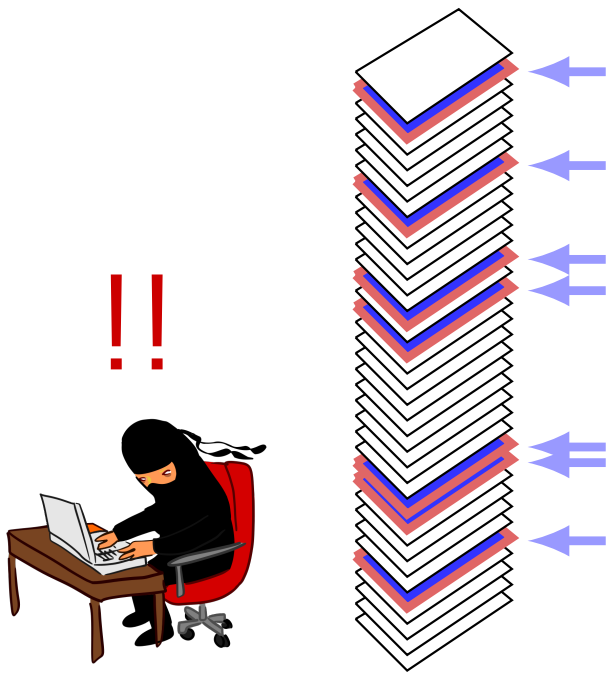


Adjudicator now has all necessary information

- Hundreds of pages of records

Images from http://openclipart.org

Motivating example: Disability adjudication support



Adjudicator now has all necessary information

- Hundreds of pages of records

Only a small subset are relevant to the case

- Much relevant information in free text observations
- This is where NLP comes in!

Images from http://openclipart.org

NLP in 30 seconds

Subfield of AI: looks at processing information that is described in human language

- Broadly: going from unstructured text to structured data
- Clinical applications in phenotyping, DDI, ADE detection, etc

NLP in 30 seconds

Subfield of AI: looks at processing information that is described in human language

- Broadly: going from unstructured text to structured data
- Clinical applications in phenotyping, DDI, ADE detection, etc

```
Medication list:
- 300 mg aspirin p.o. tid
- 150 ml ethanol p.o. daily
- 2 tab methotrexate MWF
```

NLP in 30 seconds

Subfield of AI: looks at processing information that is described in human language

- Broadly: going from unstructured text to structured data
- Clinical applications in phenotyping, DDI, ADE detection, etc

```
Medication list:
- 300 mg aspirin p.o. tid
- 150 ml ethanol p.o. daily
- 2 tab methotrexate MWF
```

$\Longrightarrow$

Medication	Dose	Freq
Aspirin	300mg	3x daily
Ethanol	150ml	1x daily
Methotrexate	2tab	MWF

Functioning and Diagnostic information

Diagnostic

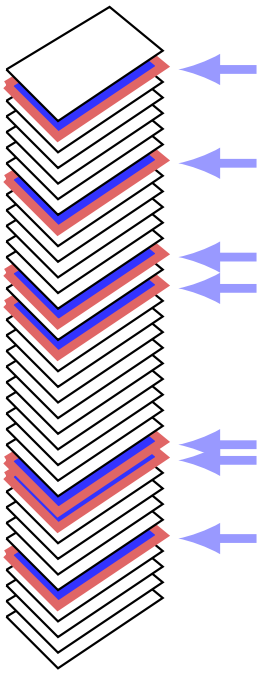
- Concerned with detecting, describing, and treating conditions
- Diagnoses, symptoms, procedures, measurements, etc
- Focus of most clinical NLP

Functioning

- Concerned with evaluating, describing, and rehabilitating impact of health conditions
- Activities (mobility, self care, communication, domestic life, etc), participation, rehabilitative care, goals, etc

Filtering to relevant data

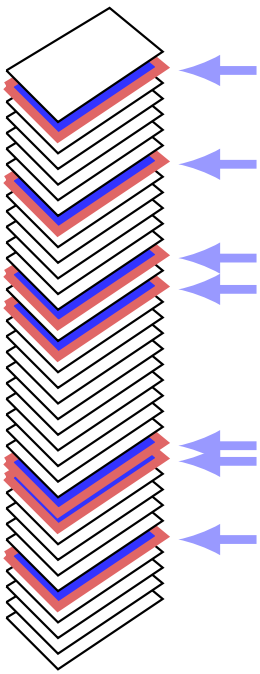
Goal: Identify functioning data in free text medical records relevant to alleged impairments



Key research questions:

Filtering to relevant data

Goal: Identify functioning data in free text medical records relevant to alleged impairments

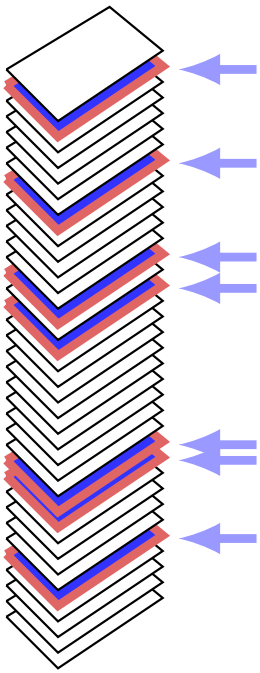


Key research questions:

1. What does functioning information look like in text?

Filtering to relevant data

Goal: Identify functioning data in free text medical records relevant to alleged impairments



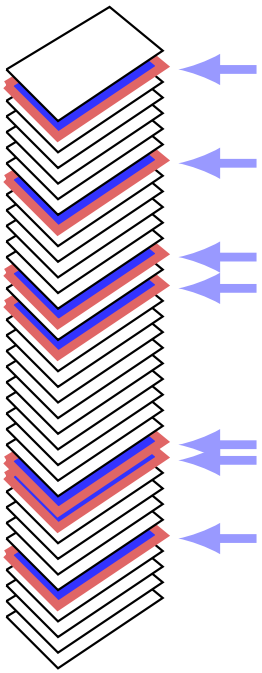
Key research questions:

1. What does functioning information look like in text?

2. What new and existing NLP techniques can we apply to get it?

Filtering to relevant data

Goal: Identify functioning data in free text medical records relevant to alleged impairments



Key research questions:

1. What does functioning information look like in text?
2. What new and existing NLP techniques can we apply to get it?
3. What challenges does functioning information pose?

Outline

Defining functioning (informally)

Individuals engage in different tasks and life situations in daily life, and operate in different environments.

Functioning (roughly) describes the ability to do these things in a given health condition.

Defining functioning (informally)

Individuals engage in different tasks and life situations in daily life, and operate in different environments.

Functioning (roughly) describes the ability to do these things in a given health condition.

We consider this at the individual level!

But there are lots of interesting
population-level questions as well.

Functioning in context: examples

Health condition: severe back pain when standing/walking for more than 10 minutes

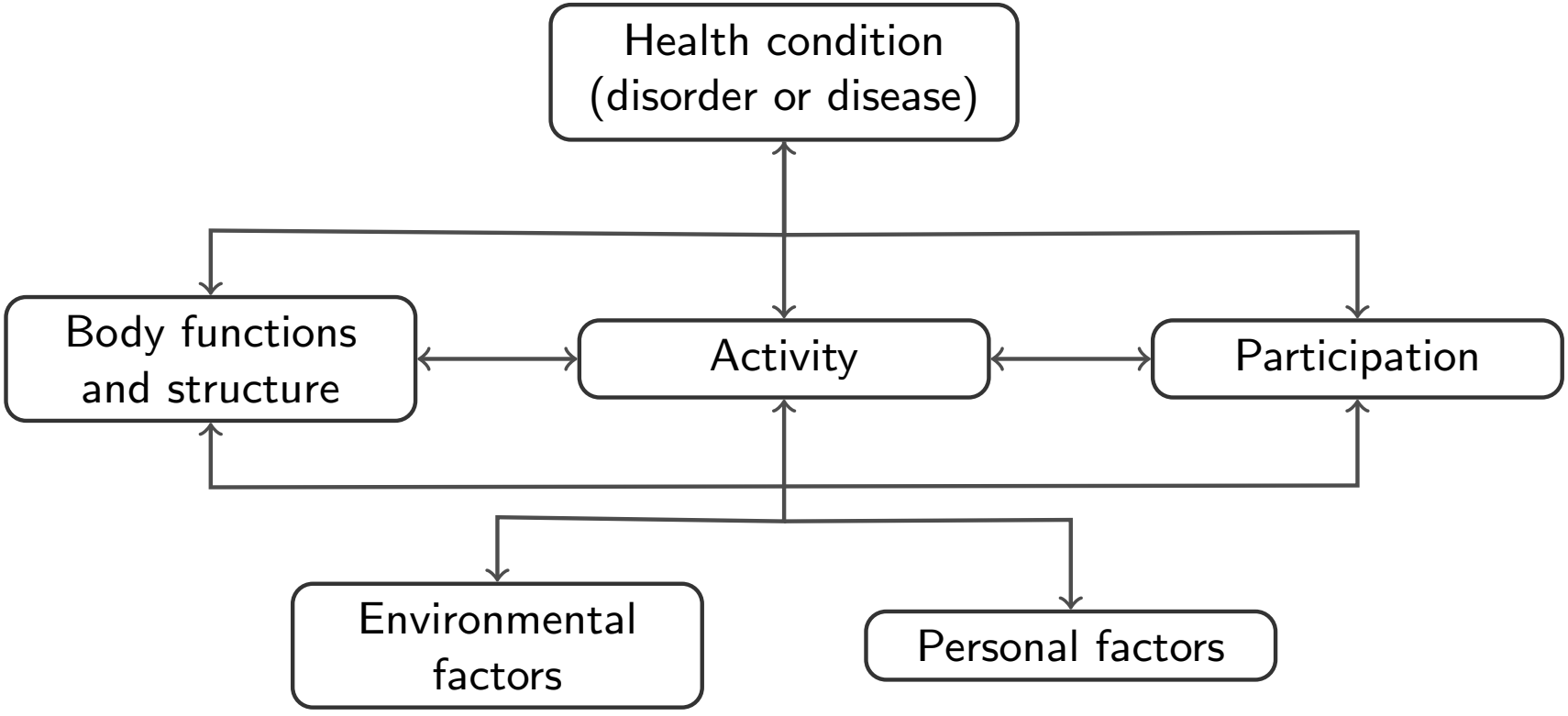
Effects on three hypothetical cases:

1. Retiree in assisted living facility, has existing transportation needs to church, store, etc
2. Packing manager in warehouse
3. Computer programmer who backpacks on the weekends

The ICF Framework

International Classification of Functioning, Disability, and Health

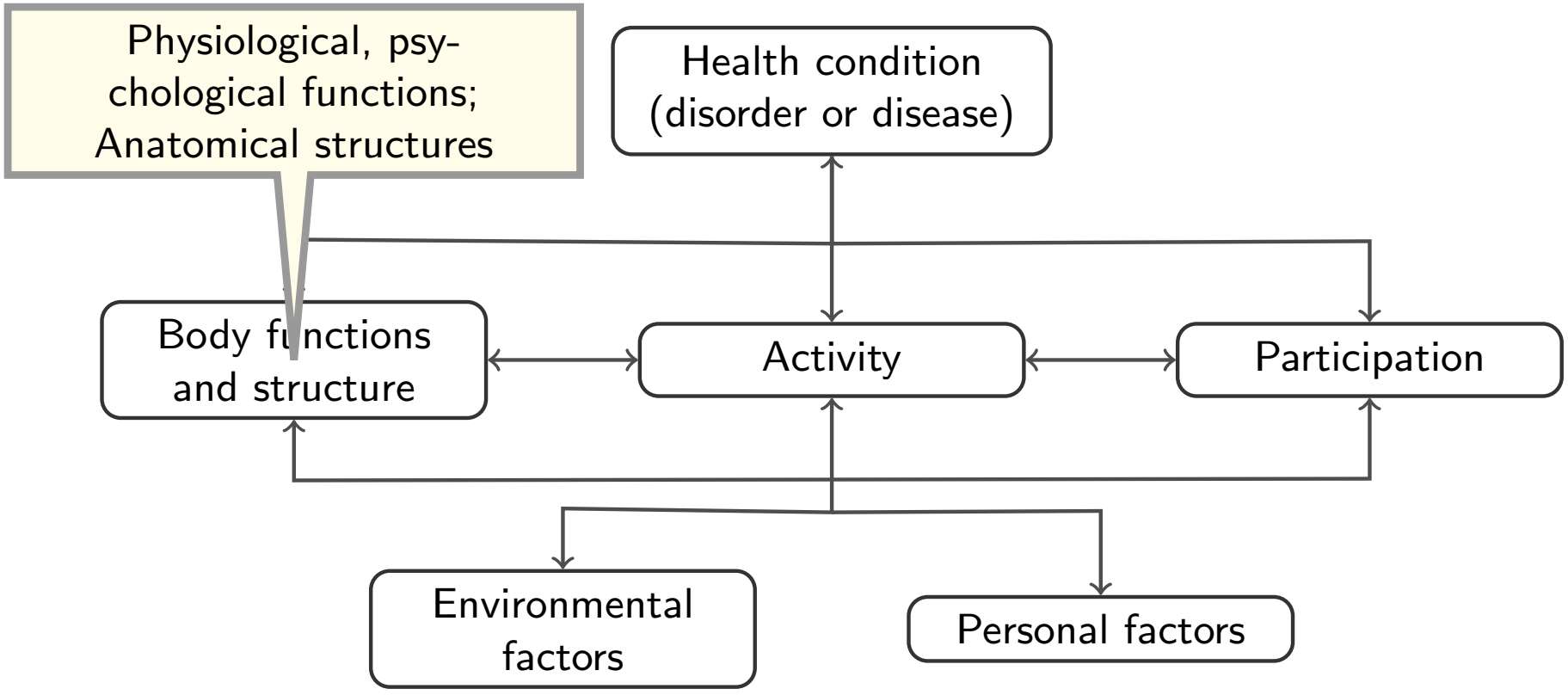
- Developed by World Health Organization (1980, 2002)
- Companion to ICD-*N*



The ICF Framework

International Classification of Functioning, Disability, and Health

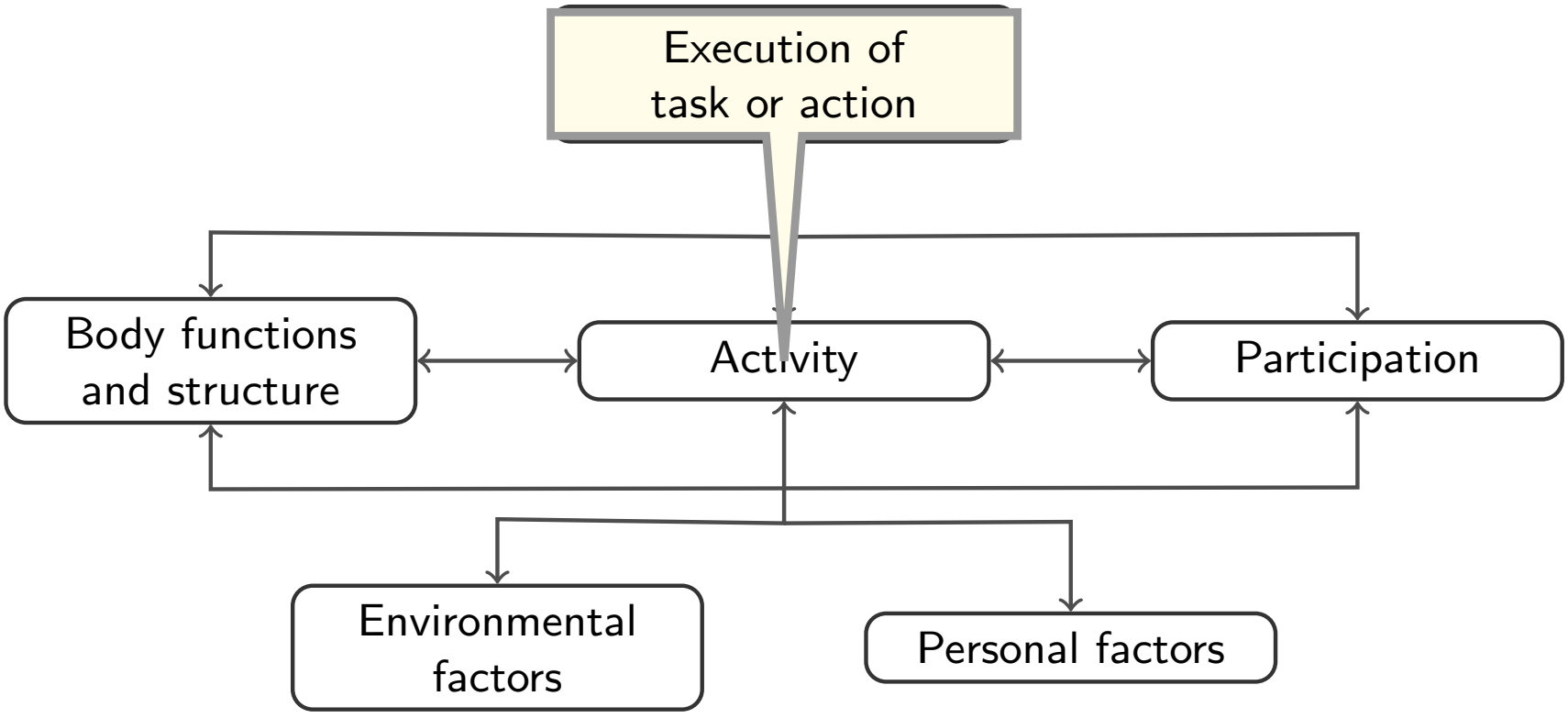
- Developed by World Health Organization (1980, 2002)
- Companion to ICD-*N*



The ICF Framework

International Classification of Functioning, Disability, and Health

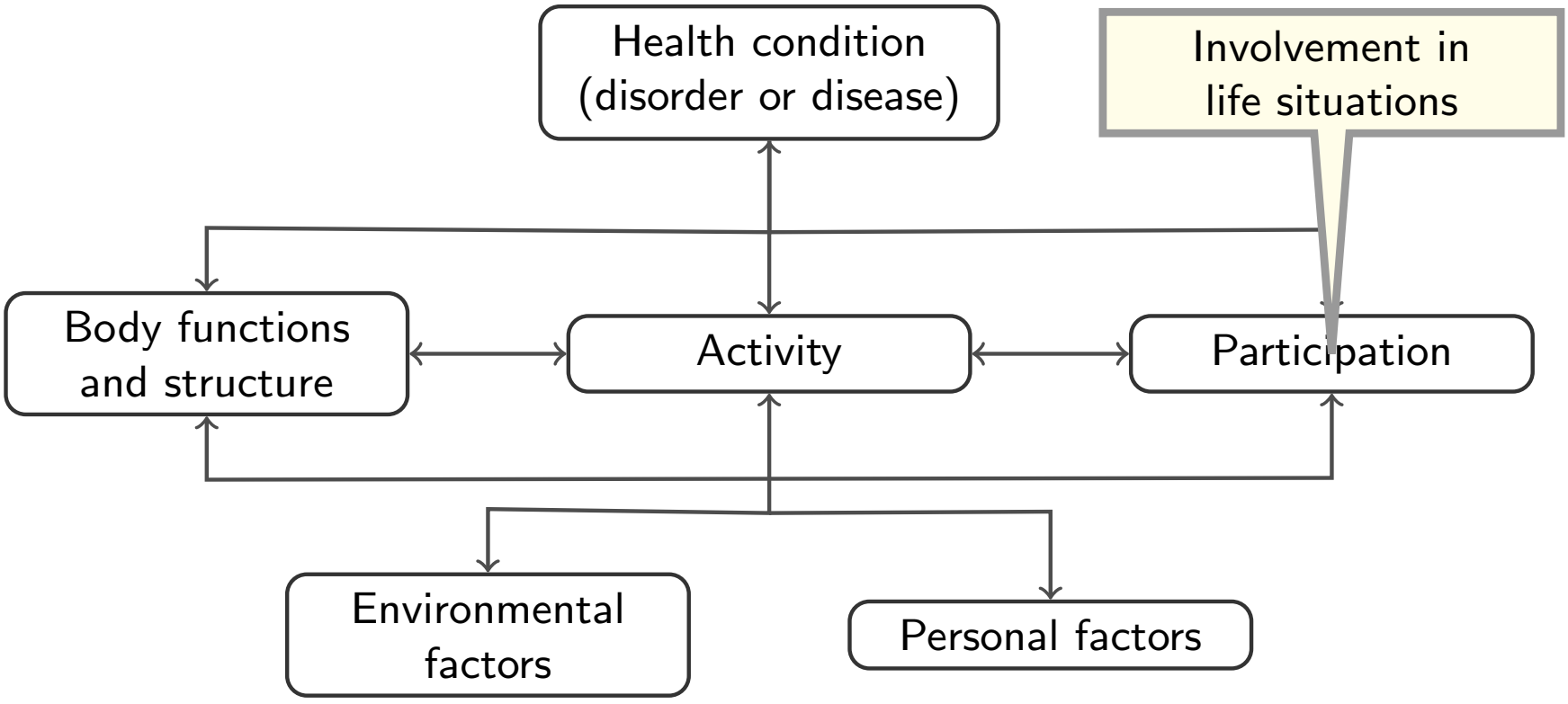
- Developed by World Health Organization (1980, 2002)
- Companion to ICD-*N*



The ICF Framework

International Classification of Functioning, Disability, and Health

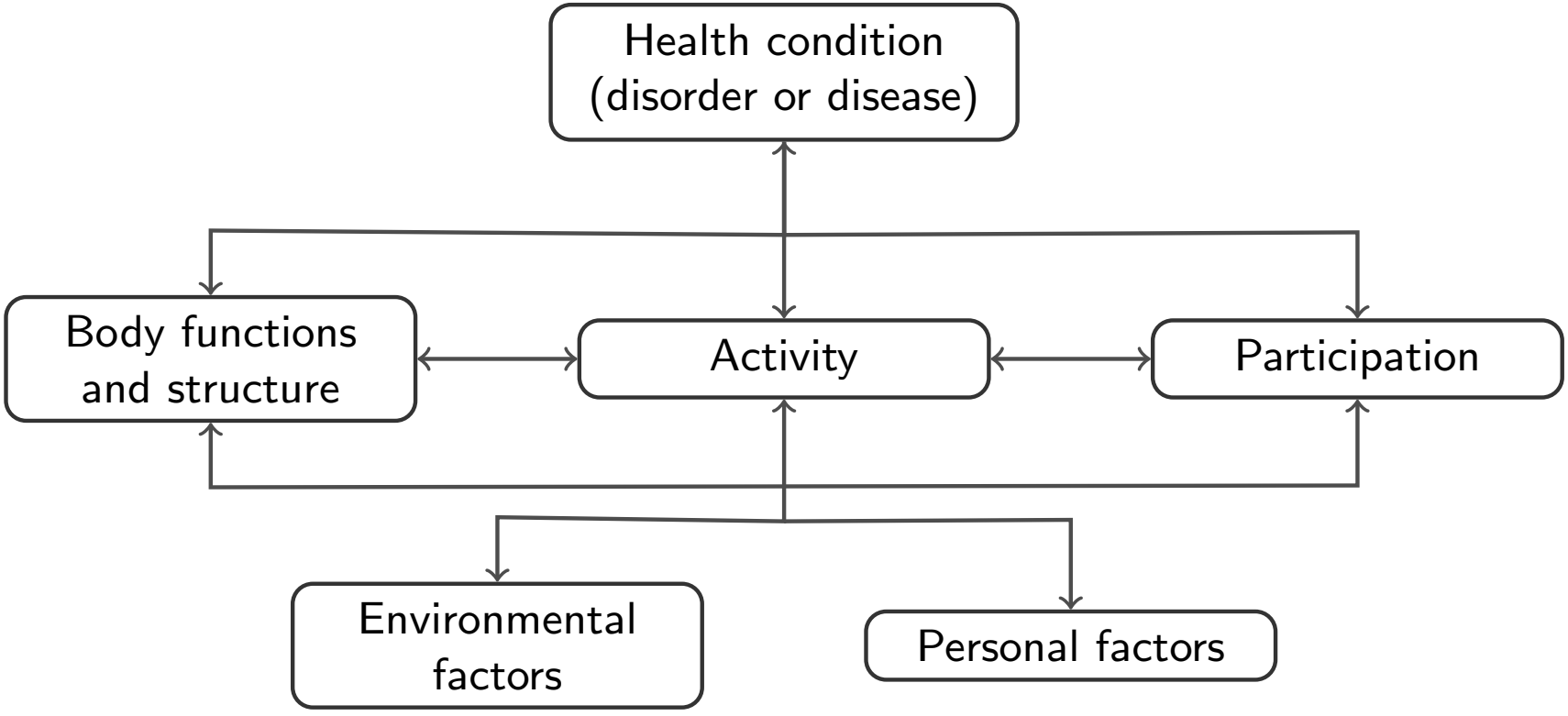
- Developed by World Health Organization (1980, 2002)
- Companion to ICD-*N*



The ICF Framework

International Classification of Functioning, Disability, and Health

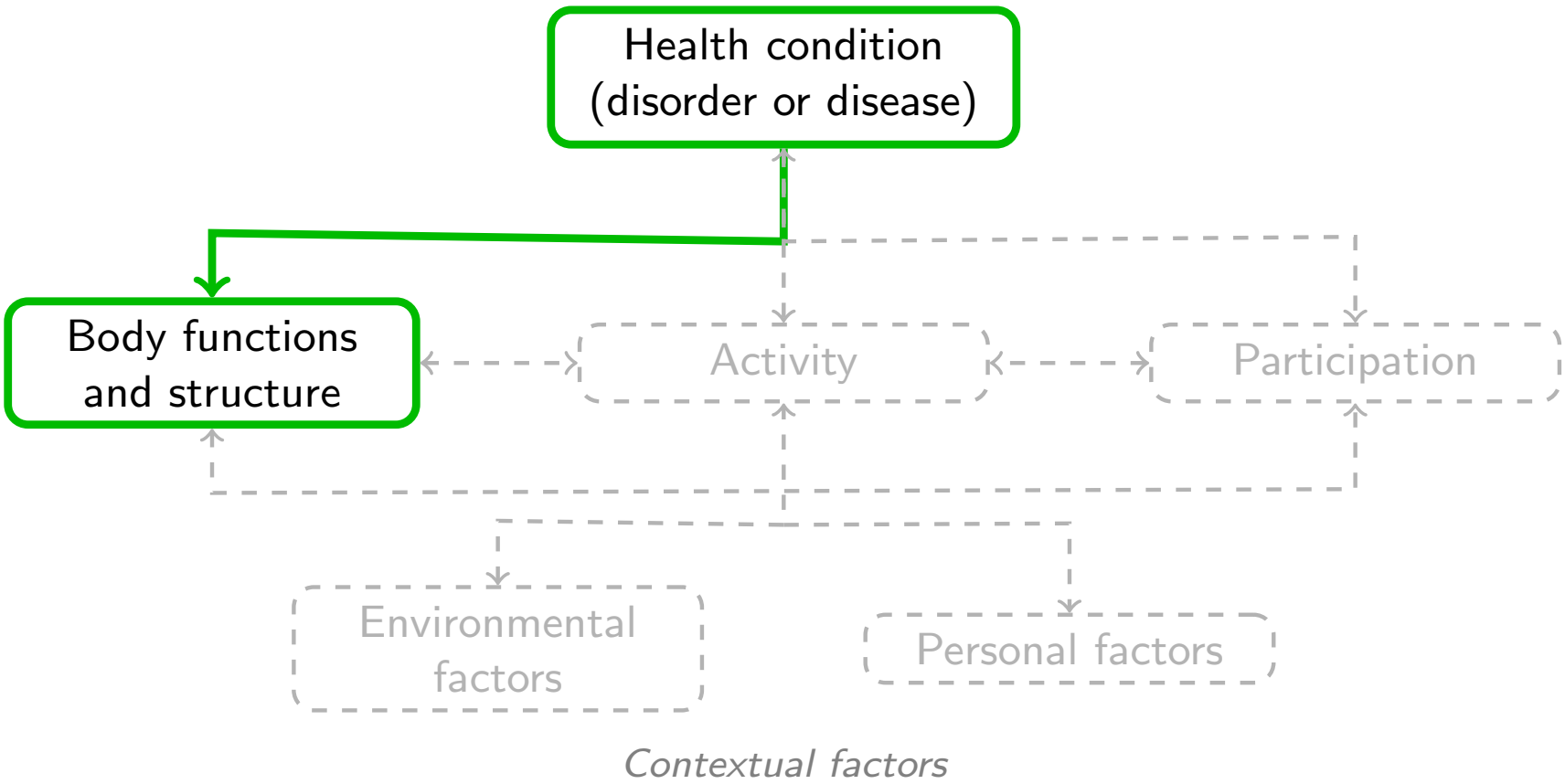
- Developed by World Health Organization (1980, 2002)
- Companion to ICD-*N*



The ICF Framework

International Classification of Functioning, Disability, and Health

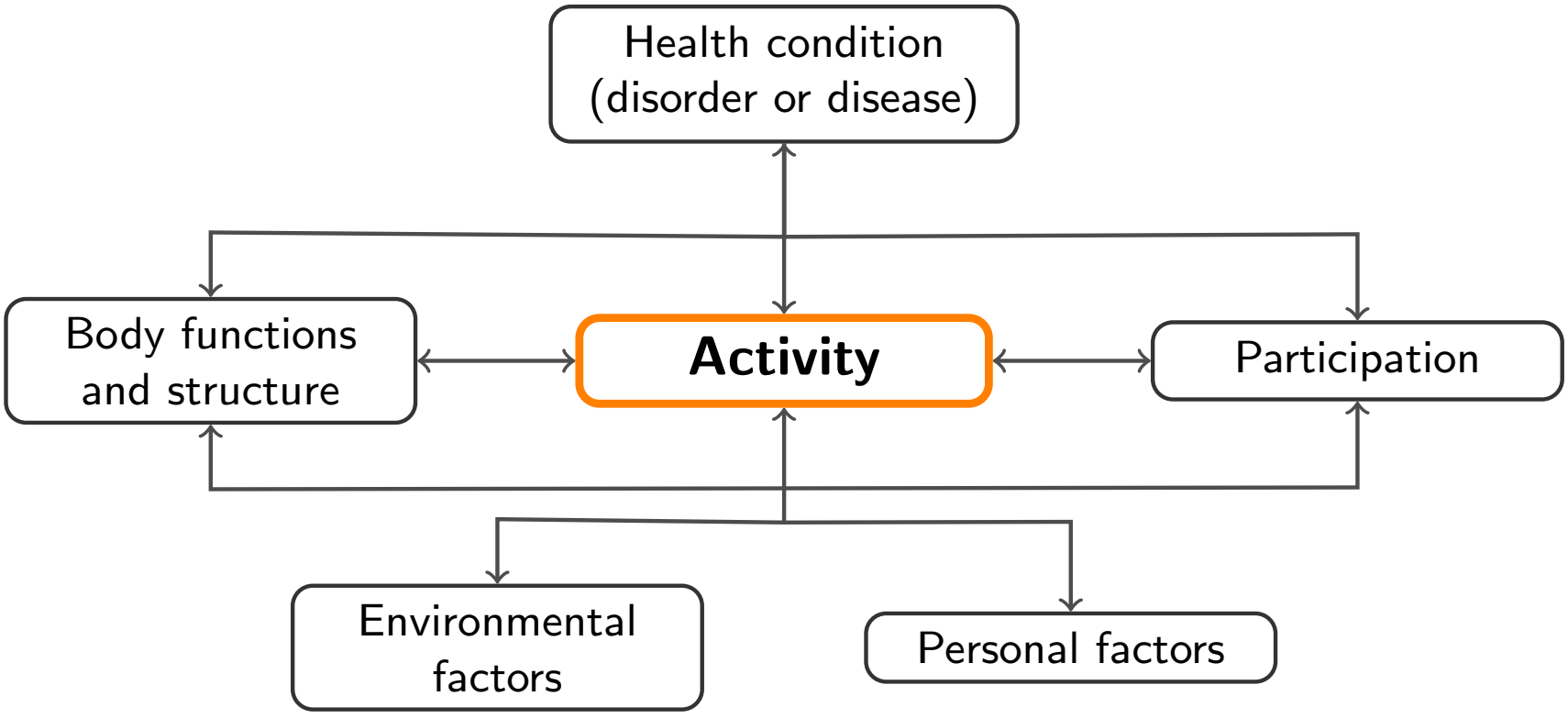
- Developed by World Health Organization (1980, 2002)
- Companion to ICD-*N*



The ICF Framework

International Classification of Functioning, Disability, and Health

- Developed by World Health Organization (1980, 2002)
- Companion to ICD-*N*



Corpus analysis of functioning language

Goal: describe linguistic patterns in documentation of functioning

Primary resource: **BTRIS**

- 155,000 documents from throughout NIH Clinical Center
- 68,000 from Rehabilitation Medicine Department
- 85,000 from various other CC departments
- Automatically deidentified before use

Auxiliary resources

- 450,000 documents from Ohio State Wexner Medical Center
- 2 million documents from MIMIC-III

Classification schema

Two levels of data classification: *Domain* and *Discipline*.

Domain

- **Functioning** - concerned primarily with patient functioning
- **Diagnostic** - concerned primarily with diagnosis/treatment of health conditions

Classification schema

Two levels of data classification: *Domain* and *Discipline*.

Discipline

- **Therapy** - documents related to therapeutic encounters (phys ther, occ ther, pulm ther, etc)
- **Medical** - non-therapeutic medical documents (majority of records)
- **Ancillary** - ancillary care encounters, including psychological evaluation and social work
- **Other** - primarily administrative documents

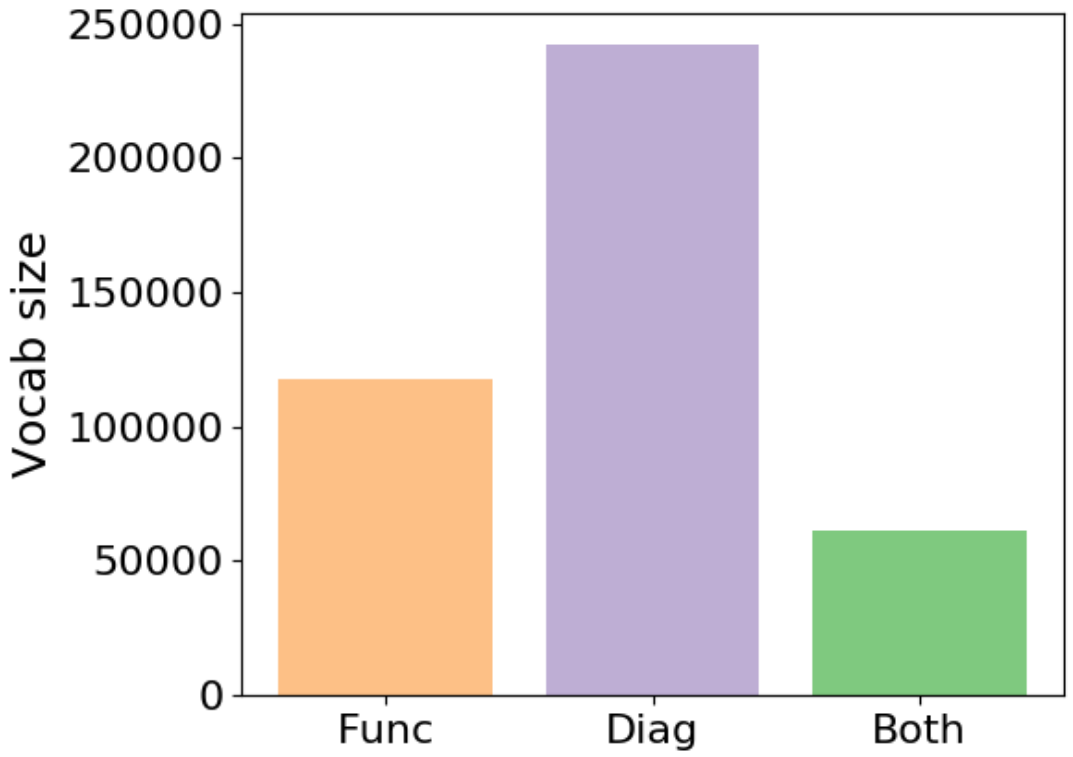
Document classes within corpora

		BTRIS	MIMIC-III
Domain	Diagnostic	68,501	2,075,079
	Function	59,532	8,101
Discipline	Medical	71,799	2,074,112
	Therapy	49,055	5,431
	Ancillary	6,496	2,670
	Other	683	967

- Labels assigned heuristically, based on document type in source EHR
- OSUWMC shows similar splits to MIMIC-III

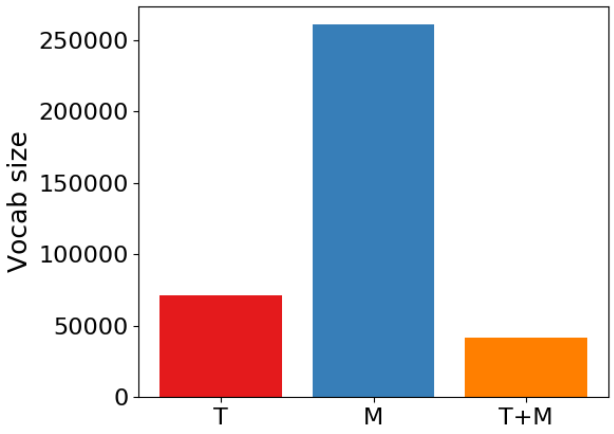
Vocabularies are highly distinctive

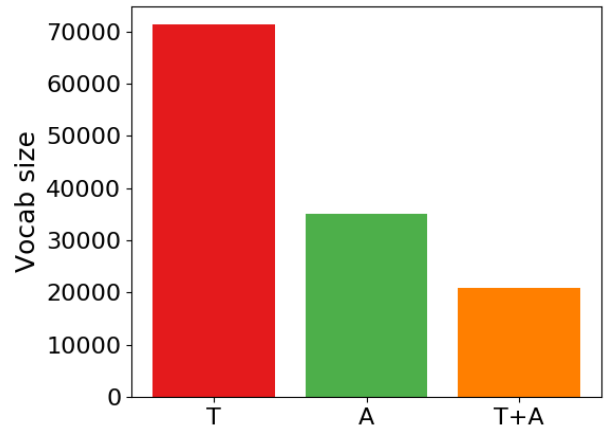
Approx 50% overlap between Functional and Diagnostic vocabularies (BTRIS)

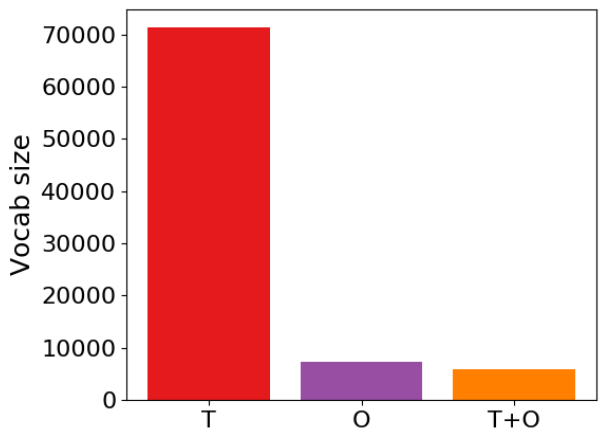


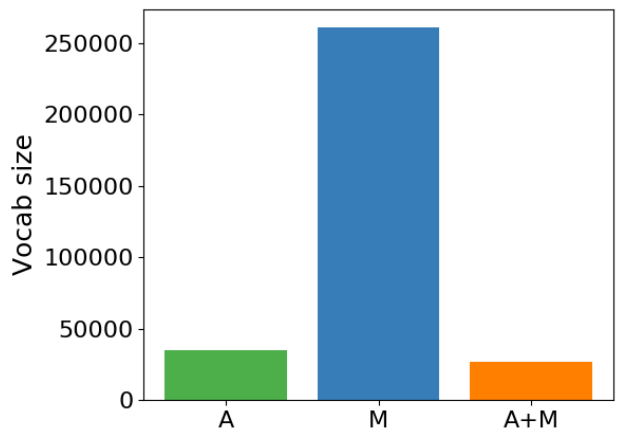
Vocabularies are highly distinctive

Similar small overlaps at Discipline level









Vocabularies are highly distinctive

Example subset-specific words:

Domain

- Functional: `amusing`, `care/cleaning`, `antipsychotic-induced`
- Diagnostic: `hernia/cyst`, `prebronchodilator`, `ovulated`

Discipline

- Therapy: `activities/interventions`, `self-advocacy`
- Ancillary: `youngster`, `downplaying`
- Medical: `accentuated`, `leukoencephalopathy`

Can distinguish documents with very high accuracy by keyword frequency

Manual review findings

Stratified random sample of 75 documents (sampled from each class)

Evaluated cTAKES performance on SBD, POS tagging, NER

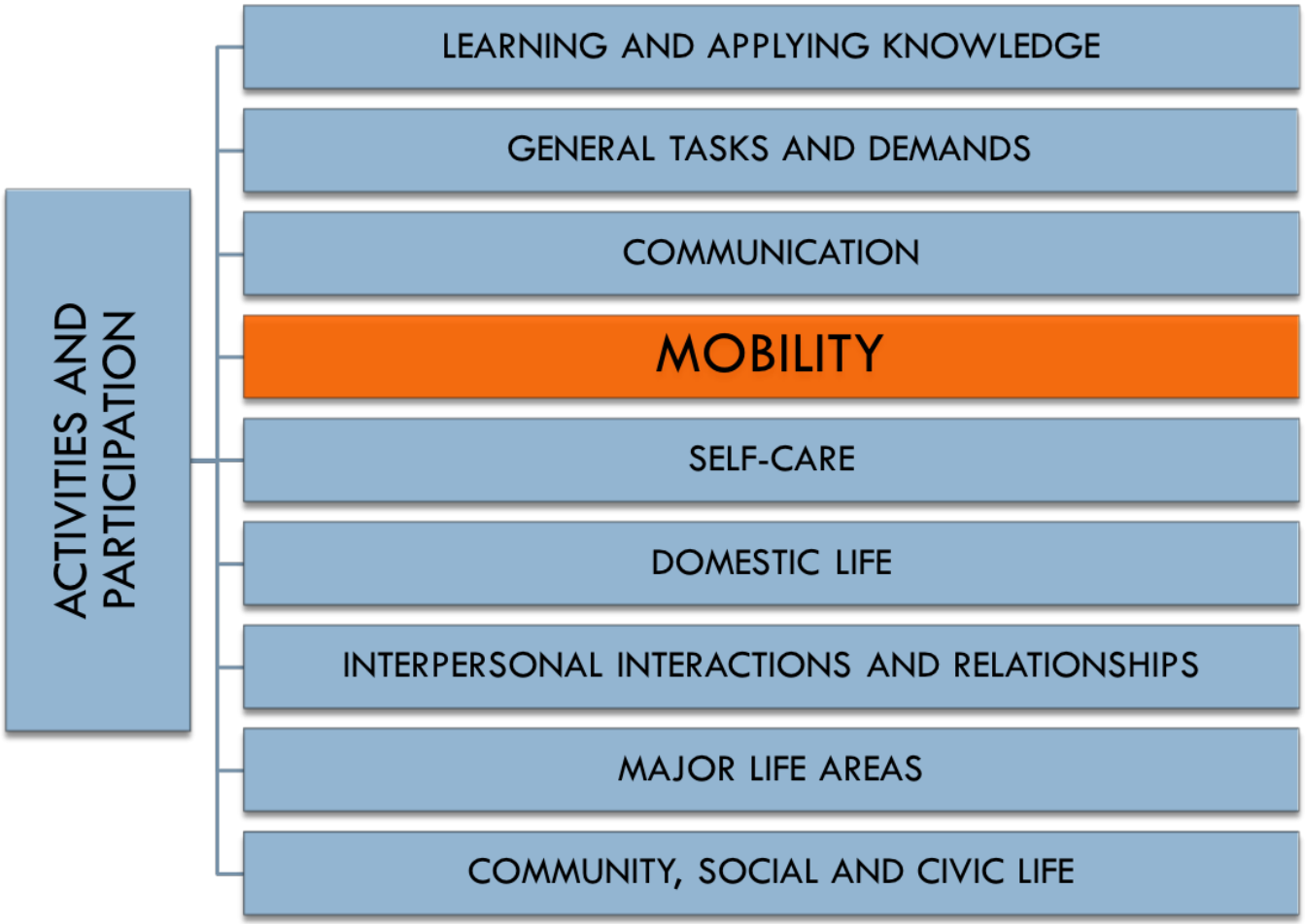
- Used to represent go-to clinical NLP tools in practice

Findings:

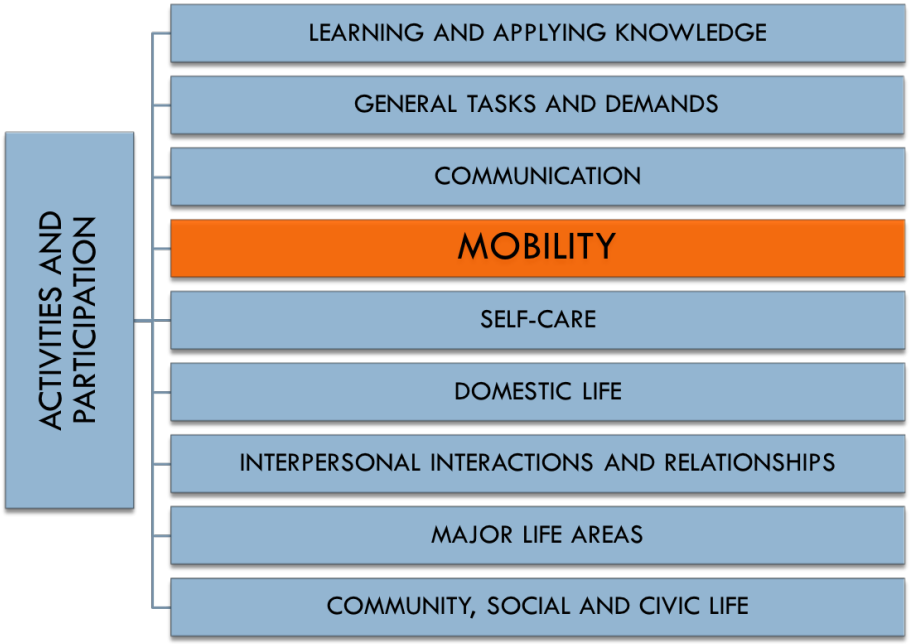
- Sentence splitting overeager; separates some information
- NER and normalization consistently poor on functioning information
  - Many critical concepts missed entirely
  - bed $\rightarrow$ "Bornholm eye disease"
- Reliable performance on Diagnostic/Medical data

Outline

Pilot study: Mobility



Pilot study: Mobility



Mobility has several advantages as an information domain:

- Self-contained; does not rely heavily on environmental factors
- Manageable in size: we use 13 3-digit ICF codes
  - Capture health outcome evaluations
- Correlated to work disability

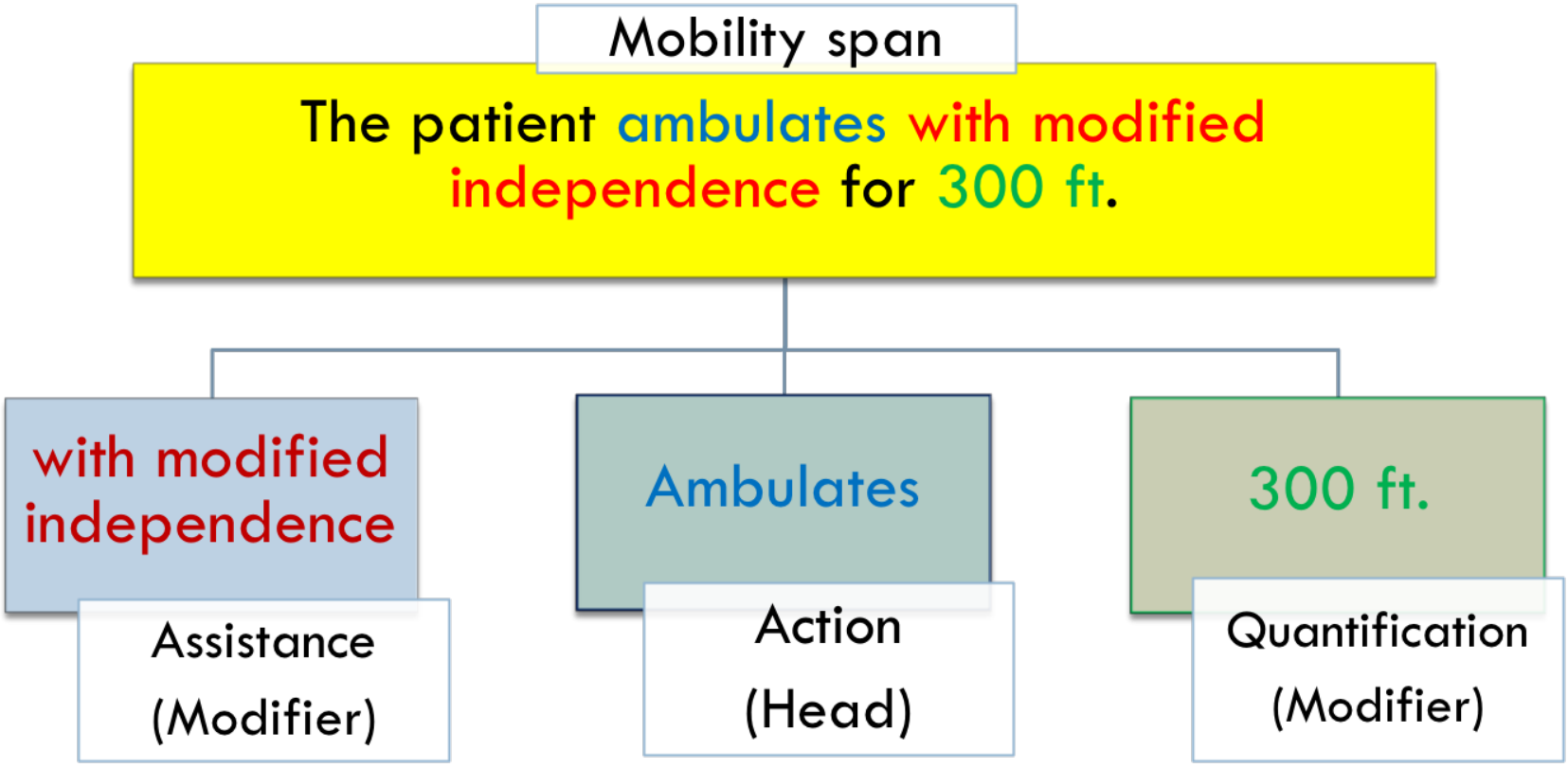
Mobility coding

Code	Name
d410	Changing basic body position
d415	Maintaining a body position
d420	Transferring oneself
d430	Lifting and carrying objects
d435	Moving objects with lower extremities
d440	Fine hand use
d445	Hand and arm use
d450	Walking
d455	Moving around
d460	Moving around in different locations
d470	Using transportation
d475	Driving
d480	Riding animals for transportation

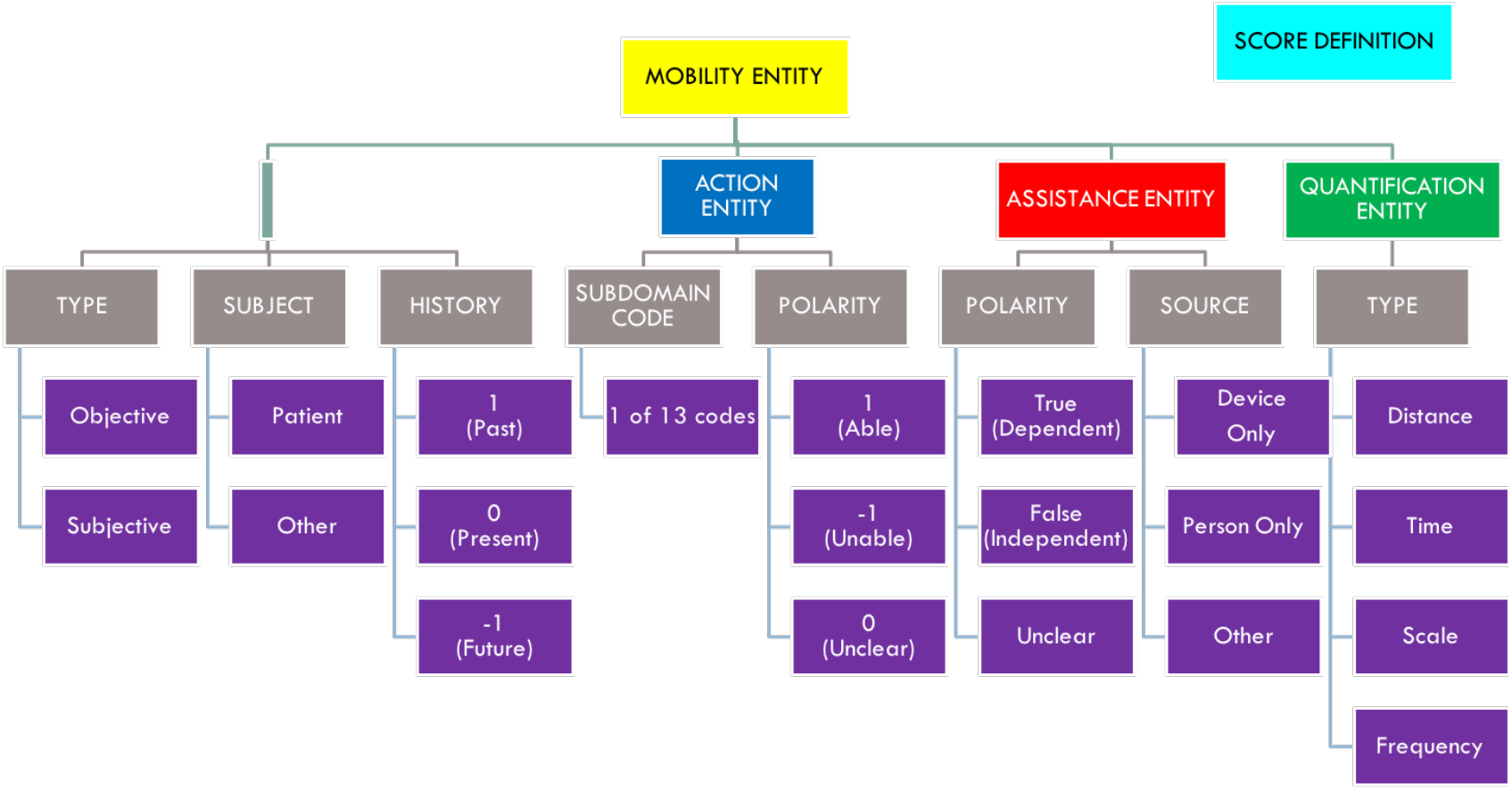
Example usage:

He is independent with bed mobility supine to sit and sit to supine **[d410]**. He can move with min A from sit on raised surface to stand w/o a.d. **[d410]** He experienced frequent LOB and able to self-correct 50% of the time. He was provided with r.w. and was able to walk in the room with it **[d450]**.

Structure of Mobility information



Structure of Mobility information



Initial annotations

Subset of Physical Therapy documents selected for annotation

250 documents fully annotated

- Mobility mentions: 2,978
- Actions: 2,867
- Assistance: 1,671
- Quantification: 1,227
- Score definition: 157

- Thieu, T et al. *Inductive identification of functional status information and establishing a gold standard corpus*. BIBM, 2017.

Examples (synthesized)

[He has resumed playing basketball, [**jogging** *d455*], [**doing squats** *d410*]]

Pt's husband states [she is [**mostly independent** *ASST*] for [**ambulation** *d450*]]

and [has about [[**10 steps** *QUANT*] **up to each level, which she is able to climb** *d455*]], *but*

- Mentions may be overlapping (see third example)
- Mobility mentions may not have Action *or* Assistance *or* Quantification

Setting a baseline on Mobility mention recognition

Two baseline methods:

- Stanford NER (CRF-based model)
- LSTM-CRF (Lample et al., 2016)

Predict unique combinations of tags:

```
she-O is-O mostly-M/ASST independent-M/ASST
for-M ambulation-M/ACT
```

Evaluation settings:

- 3-fold cross-validation
- Evaluating based on partial match and exact match

Results so far

Both baseline models doing quite well on Mobility NER

- 80-90% F-measure on exact matches
- $>$90% F-measure on partial matches
- Score Definition and Quantification are easiest, Mobility is hardest

Results so far

Both baseline models doing quite well on Mobility NER

- 80-90% F-measure on exact matches
- >90% F-measure on partial matches
- Score Definition and Quantification are easiest, Mobility is hardest

Some caveats

- Pretty high linguistic regularity in our data (common training for all providers)
- Expect to see decreased performance on more diverse data

Current successes and failures

Significant portion of perfect matches (even complex cases)

[He evals as [ind w/ [walking ACT] and self-care w/o the use of assistive devices ASST]]

[Pt should continue with core strengthening exercises and her [jogging ACT]/[hiking ACT]]

Some complete misses

[currently 3/10 while [lying supine ACT] on bed]

[Pt reported she led w/ left LE during both [ascent and descent ACT]]

Current successes and failures

Some evidence of lexical memorization

- Gold: `[w/o the need for gait aids ASST]`
- Predicted: `[[w/o the need ASST] for [gait ACT] aids]`

Generally, CRF model does better on multi-label cases

- More mistakes on Mobility-only tokens
- Overextends some sub-entity bounds

Seems to perform better on Mobility mentions at start of sentence than mid-sentence

Outline

Technical challenges

The big one: lack of standardized terminology / ontology for functioning information

- Many clinical NLP successes rely on SNOMED, UMLS, etc
- ICF is a conceptual framework, not a robust terminology
  - *d450* is "walking", not "ambulation"

Approaches

- Develop a terminology! *(big ask)*
- Investigating methods for learning representations of concepts from text + seed set of terms

Technical challenges

Another big one: lack of annotated data

- Many successful NLP methods require huge amounts of labeled data
- Haven't yet established common standards for what data structures/relationships at application level should look like

Approaches

- Drawing on existing data from other domains
  - Pre-training text representations on large web corpus and tuning on BTRIS improves NER performance
- Starting to look at unsupervised/semi-supervised methods to transfer data from related tasks; help increase training data

Linguistic challenges

Document formatting

Form fields/options in templates $\rightarrow$ concepts that aren't actually observed

- `Difficulty walking X`
  `Difficulty sitting`
  `Difficulty xfer chair to bed`

Therapy notes often include patient goals

- Functional data, but hypothetical
- `Walk 300' (2 weeks)`

Linguistic challenges

Document content

Self-reported functional observations

- Often not practical to evaluate in clinical setting
- Highly different linguistic structure
- `pt spontaneously weight shifting 'my bottom hurts'`

Relevant descriptions of current environment/occupation

- May often imply functional requirements
- `Plans to go to son's house; 6 steps to enter one level apartment`

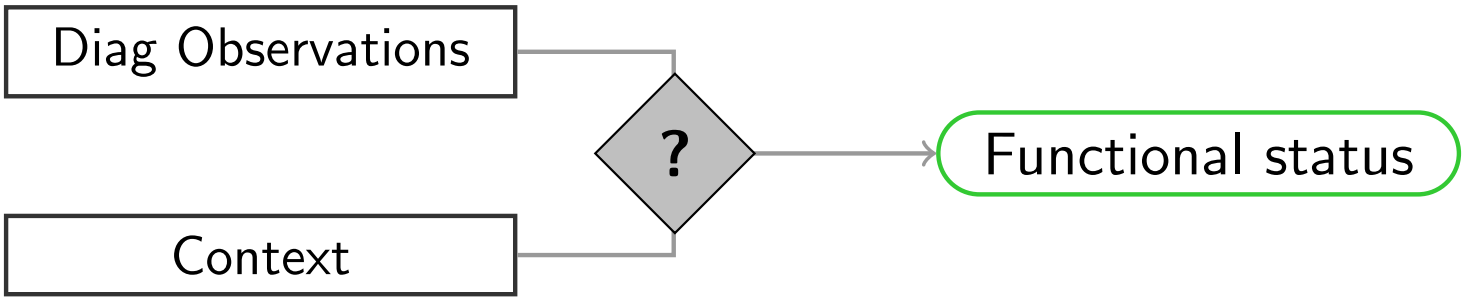
Document structure varies highly between institutions

The moonshot question

Many patients have minimal access to rehabilitative care

- Means most medical evidence will be entirely diagnostic

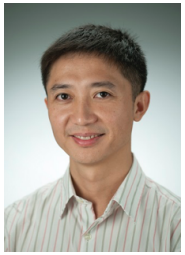
How can we develop methods to infer functional status from diagnostic observations?



Acknowledgments

Elizabeth Rasch

Leighton Chan



Chunxiao Zhou

Diane Brandt

Julia Porcino

Thanh Thieu

Pei-Shu Ho

Jonathan Camacho

Min Ding

Albert Lai

Suzanne Tamang

Eric Fosler-Lussier

And!

Ayah Zirikly
Jamil Hashmi
Alex Marr
Lisa Nelson

Ongoing research support comes from the Intramural Research Program at the National Institutes of Health, Clinical Research Center and through an Inter-Agency Agreement with the US Social Security Administration.

NIH National Institutes of Health
Clinical Center

SOCIAL SECURITY ADMINISTRATION USA

Thank you!

Contact me at:
denis.griffis@nih.gov
drgriffis.github.io

National Institutes of Health
Clinical Center

The Ohio State University

Washington University in St. Louis

Social Security USA Administration