

# Evaluation of the *Restrict to MeSH* algorithm

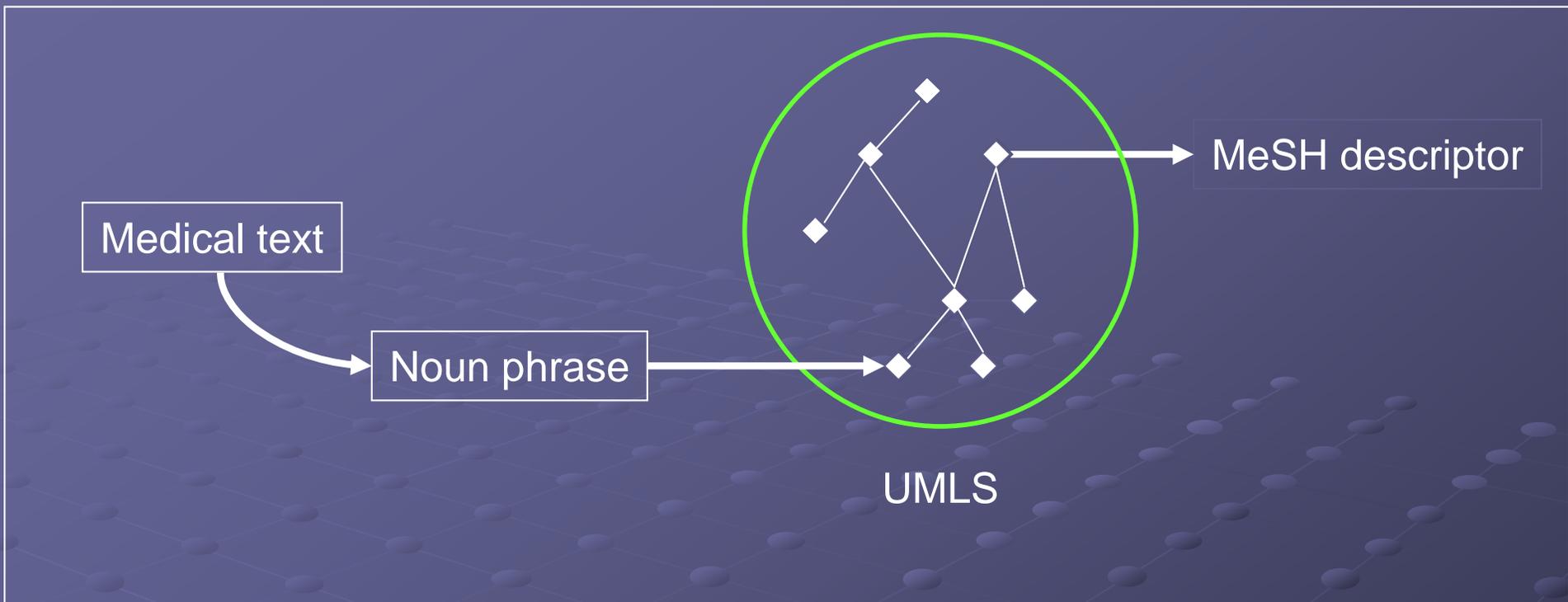
Student: Michael Bales

Mentor: Olivier Bodenreider

NLM Summer Research Rotation

# Background

- MEDLINE
- MeSH
- NLM Indexing Initiative
- UMLS Metathesaurus
  - Collection of terminological systems
  - 1.2 million concepts, each assigned a CUI



### Restrict to MeSH: Mapping CUIs to MeSH entities

UMLS CUI  
**C0202022: Primary malignant neoplasm of left lower lobe of lung**



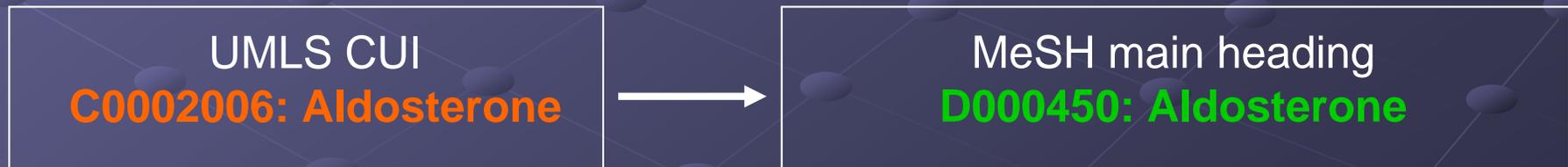
MeSH main headings  
**D001984: Bronchial neoplasms**  
**D008175: Lung neoplasms**

# Restrict to MeSH

- Based on the principle of semantic locality
- Four techniques
  - Use synonymy
  - Use associated expressions (ATXs)
  - Explore the ancestors
  - Explore the other related concepts

# Use synonymy

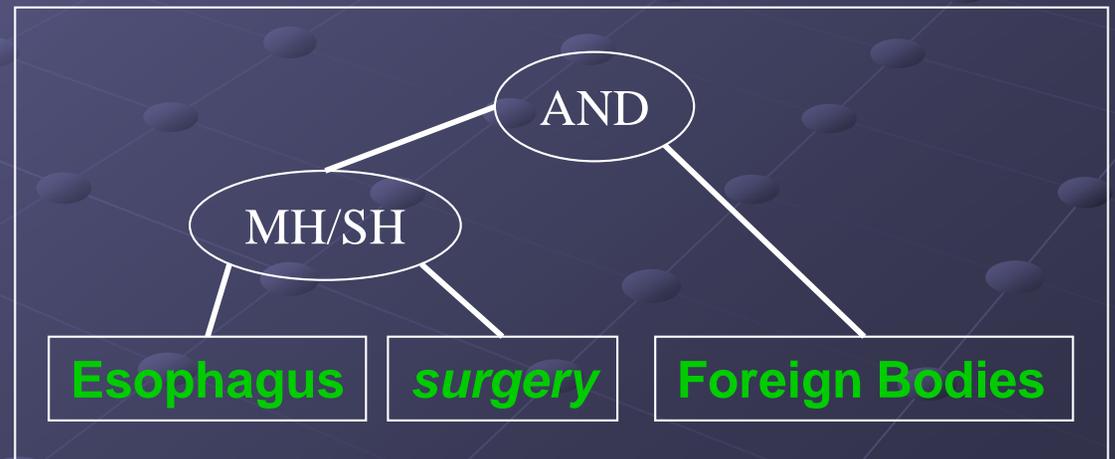
- Term mapped to source concept
- For this concept, is there a synonym term that comes from MeSH?



# Use associated expressions

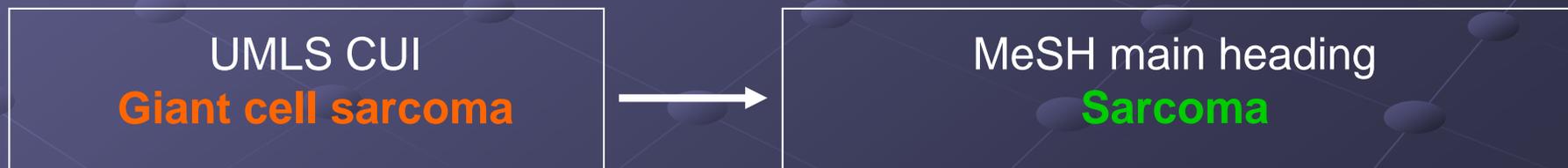
- Is there an associated expression (ATX) that describes this concept using a combination of MeSH main headings?
- ATXs correspond to ICD terms

Endoscopic removal  
of intraluminal  
foreign body from  
oesophagus without  
incision



# Use ancestors

- Build the graph of the ancestors of the concept
  - using parents and broader concepts
  - all the way to the top
  - exclude ancestors with incompatible semantic type
- From the graph, select the concepts that come from MeSH
- Remove those that are ancestors of another concept coming from MeSH
- Also try children or siblings as seed



# Use other related concepts

- Explore the other related concepts
- Exclude incompatible semantic types
- From those, select the concepts that come from MeSH

UMLS CUI

**Nicotinic Acid 0.15 MG /  
Riboflavin 0.02 MG /  
Thiamine 0.06 MG  
Oral Tablet**



MeSH main headings

**Niacin  
Riboflavin  
Thiamine  
Tablets**

# Methods

## Quantitative evaluation (all CUIs)

- From three perspectives:
  - CUIs
  - MeSH main headings
  - Mapping method

## Data used:

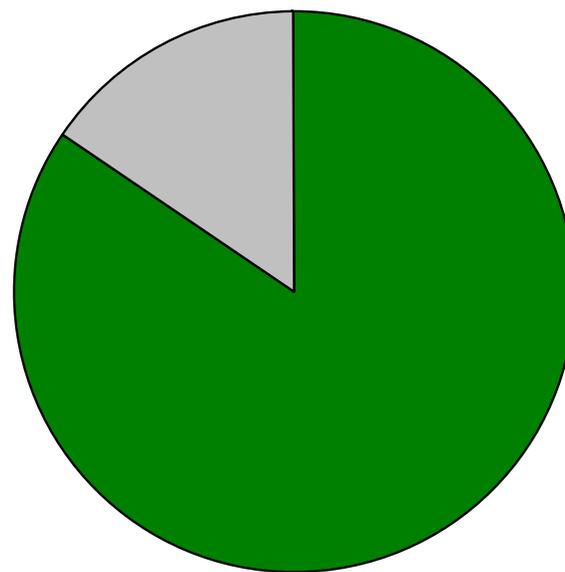
- UMLS Metathesaurus 2006AA
- RtM-suggested mappings

# Qualitative evaluation

- Assess performance on individual mappings
- Random sample of 50 CUIs
- Answer a set of questions for each CUI and mapping
- Detailed output of RtM

# Quantitative evaluation results from the perspective of **CUIs**

**Percent of CUIs assigned at least one  
mapping to MeSH**

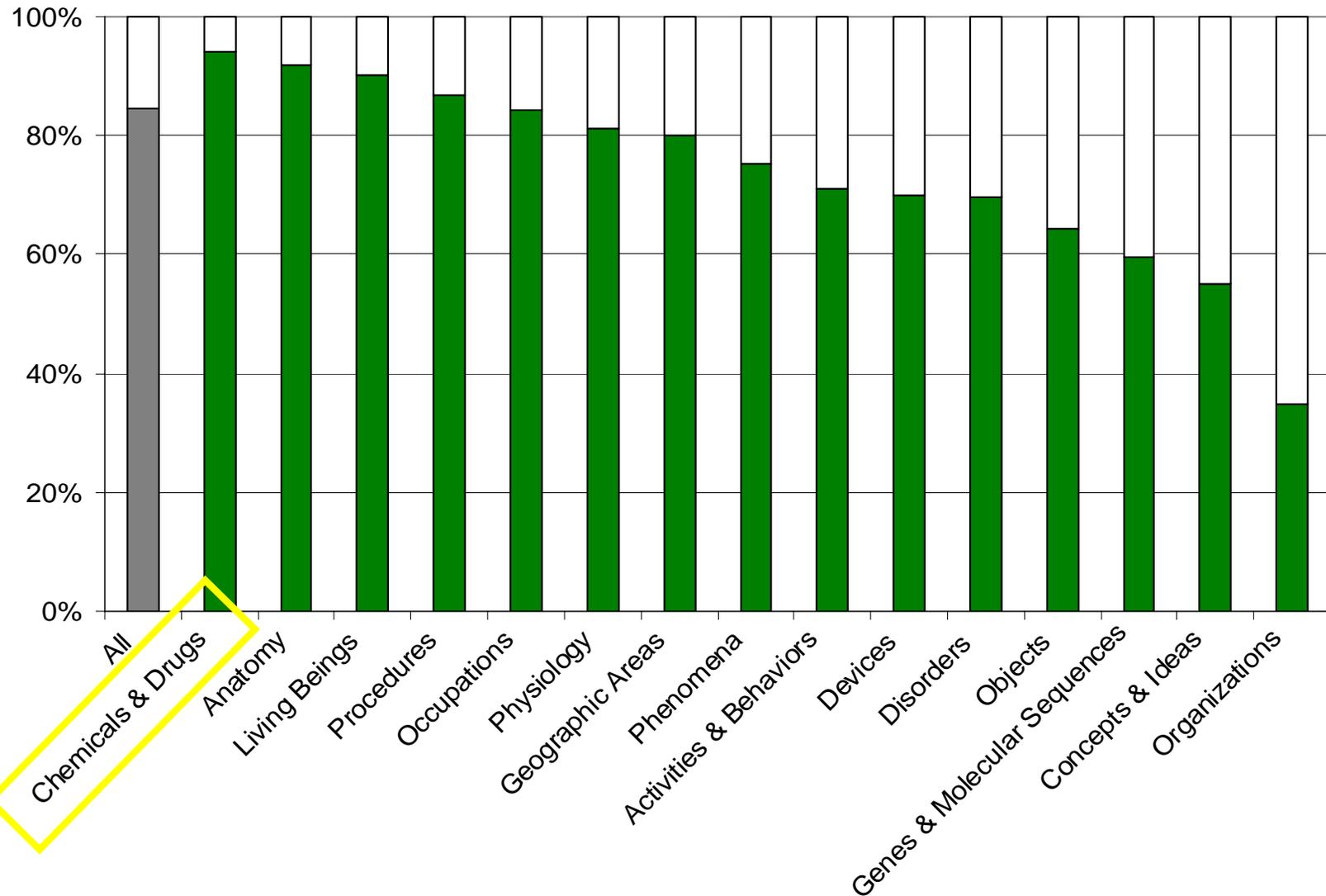


■ Mapped  
■ Not mapped

84.5% of UMLS CUIs were mapped to at least one MeSH entity

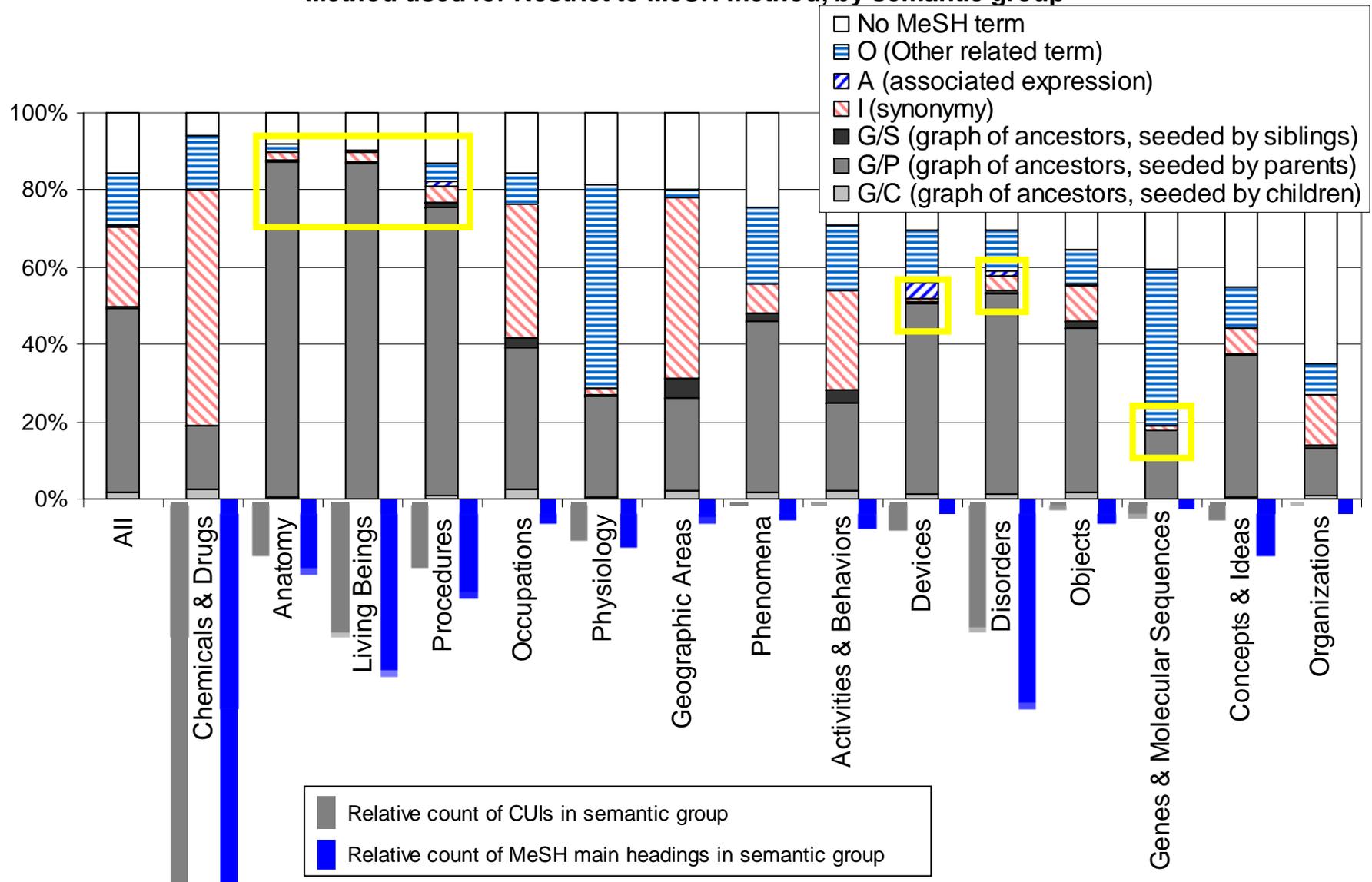
# From the perspective of CUIs

Mappings by Restrict to MeSH method, by semantic group

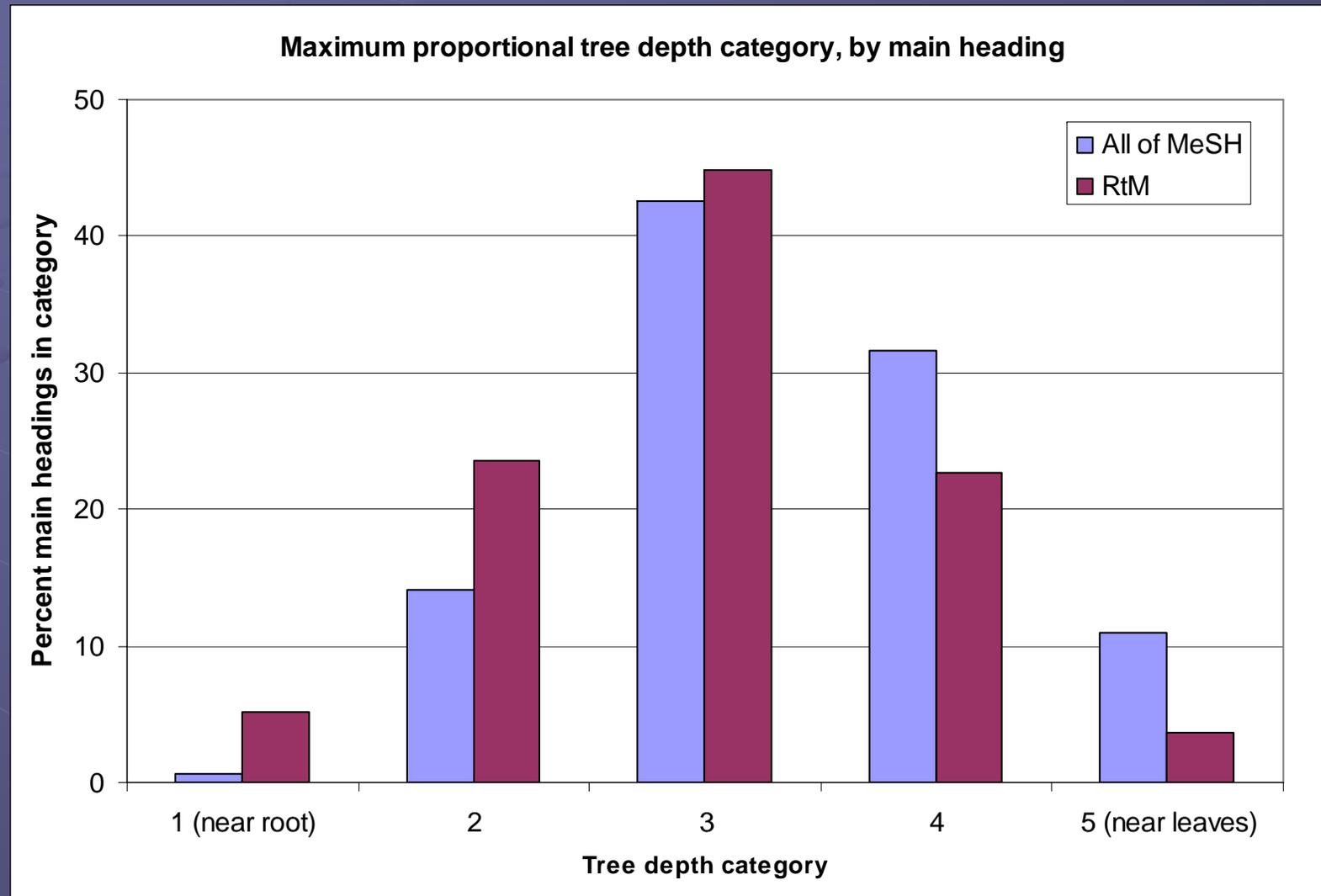


# From the perspective of **mapping method**

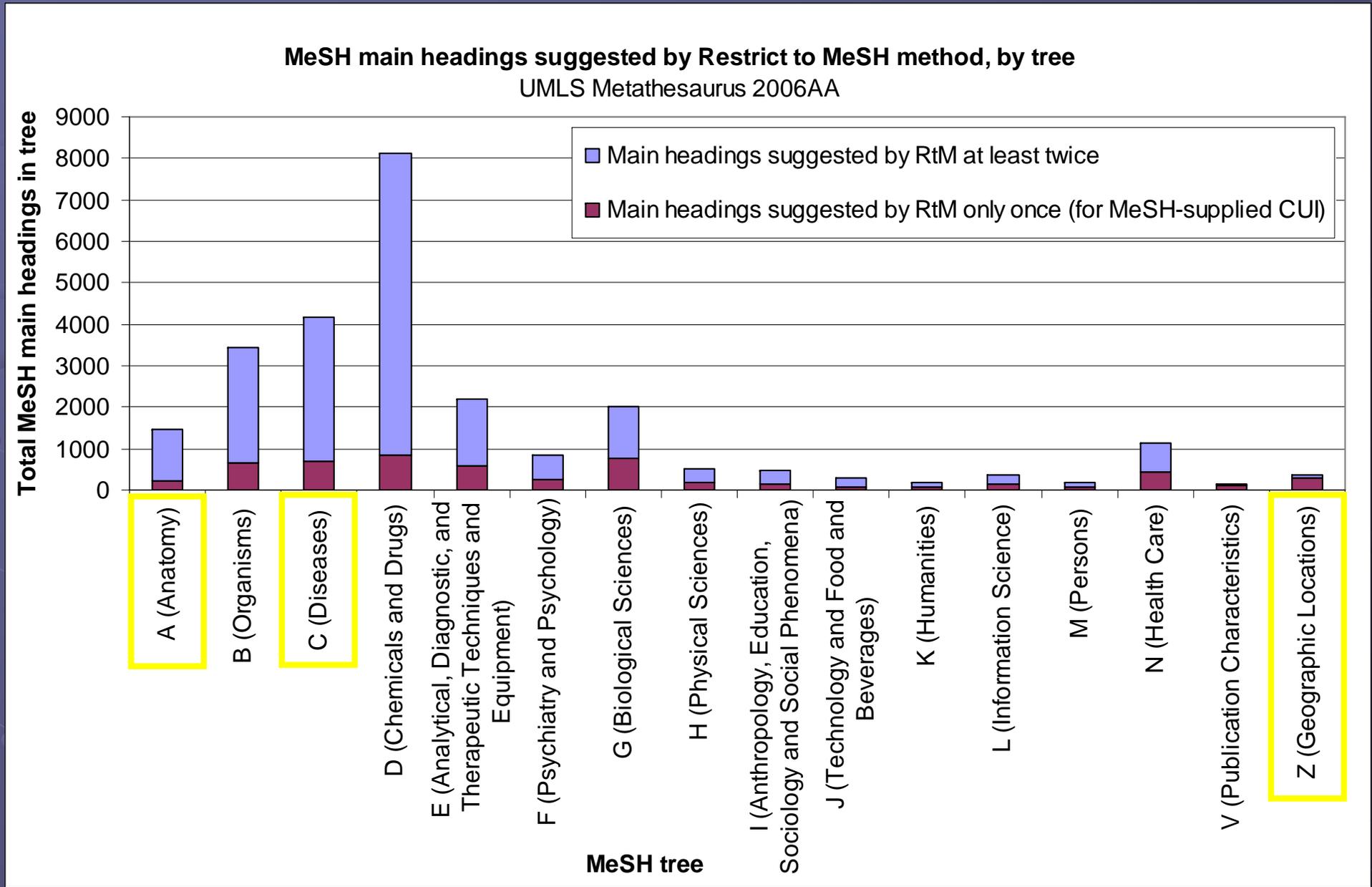
Method used for Restrict to MeSH method, by semantic group



# From the perspective of MeSH main headings



# From the perspective of MeSH main headings



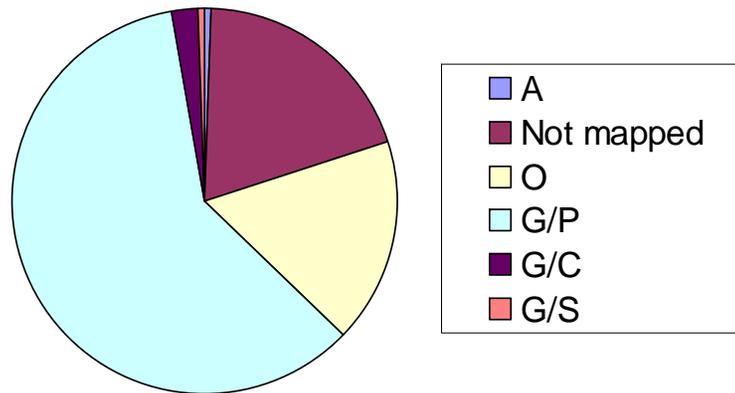
# Qualitative evaluation

Assess the quality of mappings on an individual level

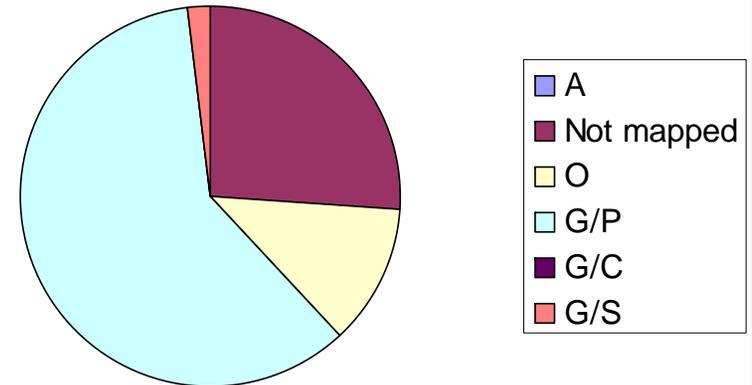
- What mapping method is used?
- If there is no mapping to MeSH, why not?
- Is the mapping in the same semantic neighborhood as the CUI?
  - If not, how did this lateral semantic drift occur?
- Is the mapping at the same level of specification as the CUI?
  - If not, how did this semantic drift occur?
- If the graph of ancestors was used, how many tree levels were climbed before selecting the suggested mapping?
- For CUIs that mapped to several MeSH entities, what proportion were appropriate mappings (none, some, or all)?

# Mapping methods in sample were similar to those used for all CUIs

Mapping methods used for all of MeSH



Methods used in random sample



# Qualitative evaluation – results

- 13 of 50 CUIs were not mapped to MeSH
  - Orphans (n=6)
  - Mapping via ancestors not possible (n=6)
  - Crossing semantic boundary (n=1)
- 37 CUIs were mapped
  - 33 of 37 (89%) were in same semantic neighborhood as CUI
  - All 37 were more general than the CUI
    - Amiodarone overdose → Overdose
    - Giant cell sarcoma → Sarcoma

**Siblings**

(none)

[direct children and narrower concepts of direct parents and broader concepts]

Referral, allied health prof

**Other Related Concepts**

(none)

**Co-occurring Concepts**

(none)

bc1 Referral, allied health prof LEGEND

Start again Apply new parameters

Restrict to vocabulary: Show all

Highlight vocabulary: Nothing

UMLS data: UMLS\_2006AA

Type of hierarchical rel.:  All  Parent/Child only  Broader/Narrower only

Transitive reduction:  yes  no

**Similar Concepts**

(none)

**Allegedly Synonyms**

(none)

**Closest MeSH Terms**

(none)

# Conclusion

- RtM already achieves good performance
- Minor enhancements will improve method
- Rapid growth in biomedical literature
- Effective manual and automated indexing methods increasingly needed

# Acknowledgements

- Olivier Bodenreider
- May Cheh
- Rob Logan
- Tom Rindflesch
- NLM Training Grant