FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
MEDIZINISCHE FAKULTÄT

Extraction and Prevalence of Structured Data Elements in Free-Text Clinical Trial Eligibility Criteria

Christian Gulden, Inge Landerer, Azadeh Nassirian, Fatma Betül Altun, Johanna Andrae

EFMI STC 2019

Background & Motivation

Background

- Recruitment into clinical trials remains challenging:
  - McDonald et al, 2006 [1]:
    "**Less than a third (31%)** of the trials achieved their original recruitment target"
  - Sully et al, 2008 [2]:
    "[…] **(55%)** of trials recruited their originally specified target sample size, with over three-quarters (78%) recruiting 80% of their target. […] **Nearly half (45%)** of trials received an extension of some kind."

[1] McDonald AM, Knight RC, Campbell MK, et al. *What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies.* Trials. 2006;7:9. Published 2006 Apr 7. doi:10.1186/1745-6215-7-9
[2] Sully BG, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. Trials. 2013;14:166. Published 2013 Jun 9. doi:10.1186/1745-6215-14-166

Background

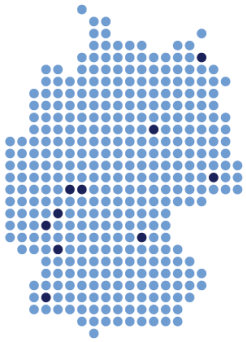
- (Manual) Recruitment fails to achieve its goals
- How about semi-automatic recruitment using computers?
- Electronic Health Records (EHRs) contain large quantities of patient data viable for recruitment
- Use this data to automatically recommend patients for clinical trials to practicioners

MII MIRACUM Use Case: IT-Support for patient recruitment

Medical Informatics Initiative

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
MEDIZINISCHE FAKULTÄT

- Leverage Digitilization in Medicine for Research and Care in Germany
- 4 consortia (MIRACUM, HIGHmed, SMITH, DIFUTURE)
- Build Data Integration Centers (DICs) containing harmonized data from various source systems
- Implement IT-Solutions (Use Cases) based on these DICs



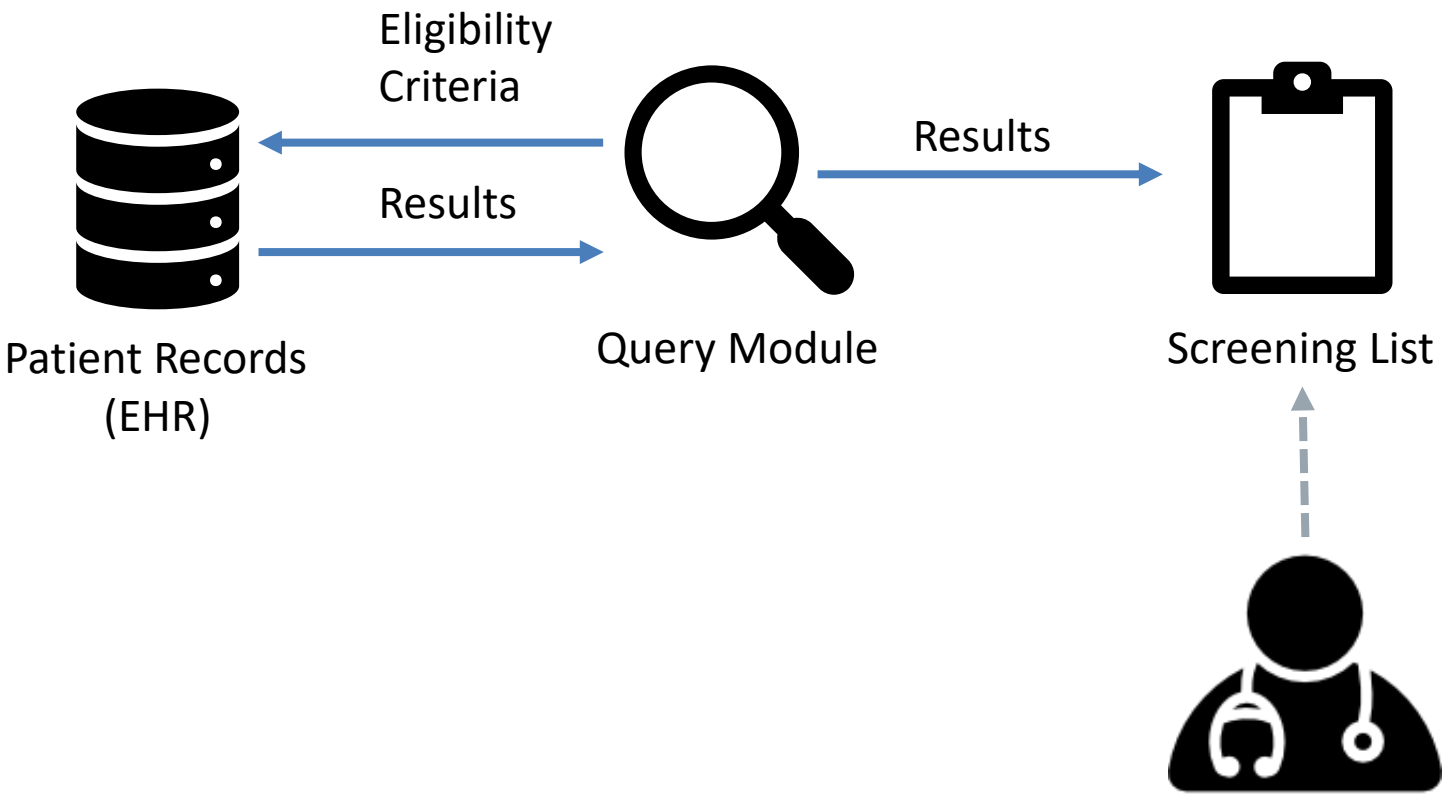
miracum

Medical Informatics in Research and Care in University Medicine

Use Case: IT Support for Patient Recruitment

MIRACUM: IT Support for Patient Recruitment

- Simplified



Research Question

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
MEDIZINISCHE FAKULTÄT

What structured data elements are the most relevant for clinical trial eligibility screening?

Methods

Methods

1. Select 50 clinical trials
2. Extract the free-text eligibility criteria
3. Distribute the criteria among the 10 MIRACUM sites
4. Analyze the criteria and extract structured data elements
5. Validate and harmonize the extracted data elements
6. Generate summary statistics over the results

Selection of 50 Trials

- Source: clinicaltrials.gov
- Actively recruiting
- At least one MIRACUM site is participating in the trial
- Represent a broad range of clinical disciplines
  - Oncology, neurology, …

Extraction of Criteria

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
MEDIZINISCHE FAKULTÄT

NIH U.S. National Library of Medicine
ClinicalTrials.gov

```xml
<clinical_study>
    <required_header>
        <download_date>
ClinicalTrials.gov processed this data on March 11, 2019
        </download_date>
        <link_text>Link to the current ClinicalTrials.gov record.</link_text>
        <url>https://clinicaltrials.gov/show/NCT03870100</url>
    </required_header>
    <id_info>
        <org_study_id>SHR-1222-101</org_study_id>
        <nct_id>NCT03870100</nct_id>
    </id_info>
    <brief_title>
The Safety, Tolerability and PK/PD Study of A Single Subcutaneous Injection of SHR-1222 in He
    </brief_title>
    <official_title>
Use the Protocol Title. The Safety, Tolerability, Pharmacokinetics and Pharmacodynamics Study
    </official_title>
    <sponsors>
        <lead_sponsor>
            <agency>Jiangsu HengRui Medicine Co., Ltd.</agency>
            <agency_class>Industry</agency_class>
        </lead_sponsor>
    </sponsors>
    <source>Jiangsu HengRui Medicine Co., Ltd.</source>
    <oversight_info>
        <has_dmc>No</has_dmc>
        <is_fda_regulated_drug>No</is_fda_regulated_drug>
        <is_fda_regulated_device>No</is_fda_regulated_device>
    </oversight_info>
    <brief_summary>
        <textblock>
This is a Single Center, Randomized, Double-Blind, Dose Escalation, Placebo Parallel Control
        </textblock>
    </brief_summary>
    <detailed_description>
        <textblock>
50 adult healthy subjects with 5 dose groups will be enrolled in the study, including six sub
        </textblock>
    </detailed_description>
    <overall_status>Not yet recruiting</overall_status>
    <start_date type="Anticipated">March 20, 2019</start_date>
    <completion_date type="Anticipated">November 30, 2019</completion_date>
    <primary_completion_date type="Anticipated">November 30, 2019</primary_completion_date>
    <phase>Phase 1</phase>
```

Confluence

Perioperative Chemotherapy Compared To Neoadjuvant Chemoradiation in Patients With Adenocarcinoma of the Esophagus (ESOPEC)

https://clinicaltrials.gov/ct2/show/NCT02509286

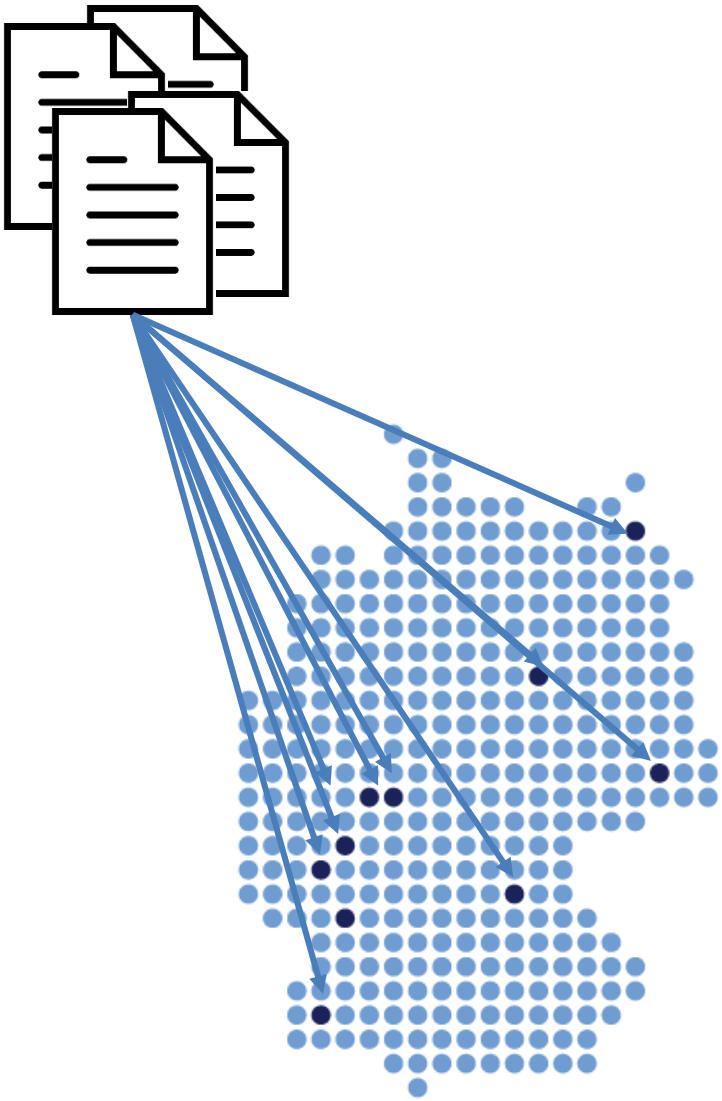
Inclusion-Criteria

Free-text Criteria
Ages Eligible for Study: 18 Years and older (Adult, Older Adult)
Sexes Eligible for Study: All
Histologically verified adenocarcinoma of the esophagus according to the UICC definition (TNM7)
Pre-treatment stage cT1N+, M0 or cT2-4a, N0/+, M0
Age ≥18 years
No prior abdominal or thoracic radiotherapy
ECOG Performance status 0-2
Adequate cardiac function (Patients with a cardiac history (e.g. myocardial infarction, heart failure, coronary artery

XML

HTML

Distribute Work



Confluence

50 Trials

5 Trials / site

Extraction of Structured Data Elements

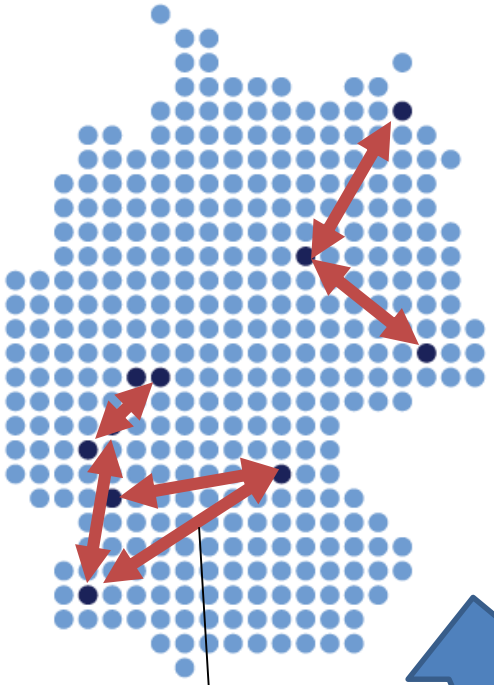
Criteria	Simplified	Objective	Group	Element
Age ≥ 18 years				
ECOG Performance status 0-2				
Subject has a clinically relevant ECG abnormality, in the opinion of the investigator				
Hepatorenal syndrome (type I or II) or screening serum creatinine >2 mg/dL (178 µmol/L)				
…				

Extraction of Structured Data Elements

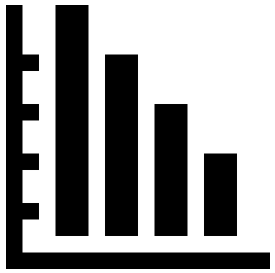
Criteria	Simplified	Objective	Group [1]	Element [1]
Age ≥ 18 years		✓	Demographics	Date of Birth
ECOG Performance status 0-2		✓	Scores & Classifications	ECOG Performance Status
Subject has a clinically relevant ECG abnormality, in the opinion of the investigator		✗		
Hepatorenal syndrome (type I or II) or screening serum creatinine >2 mg/dL (178 µmol/L)	Hepatorenal syndrome (type I or II)	✓	Diagnosis	Diagnosis code
	Screening serum creatinine >2 mg/dL (178 µmol/L	✓	Laboratory Findings	Creatinine in serum
…				

[1] Doods, Justin et al. “A European inventory of data elements for patient recruitment.” Studies in health technology and informatics 210 (2015): 506-10 .

Harmonize Results



tally and merge tables



(preliminary) results

inter-site peer-review

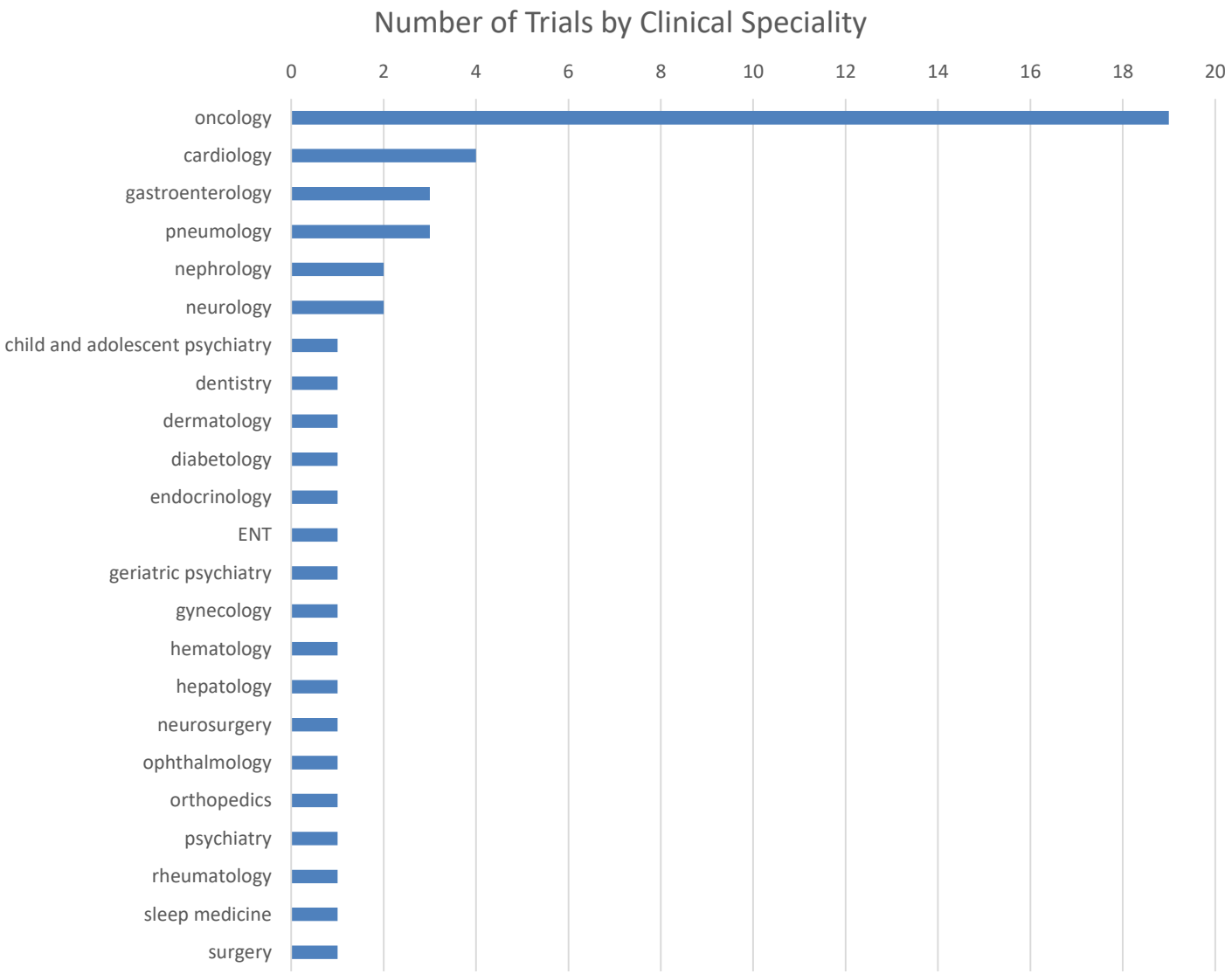
revision loop
(e.g. eliminating duplicates, merging synonyms)

Results

Selection of 50 Trials

- Total number of trials registered on clinicaltrials.gov: 277.228
- Currently recruiting and MIRACUM site is participating: 416
- Manual filtering yielded final 50, e.g.
  - Removing post-market and registry studies
  - achievieving broad representation of disciplines

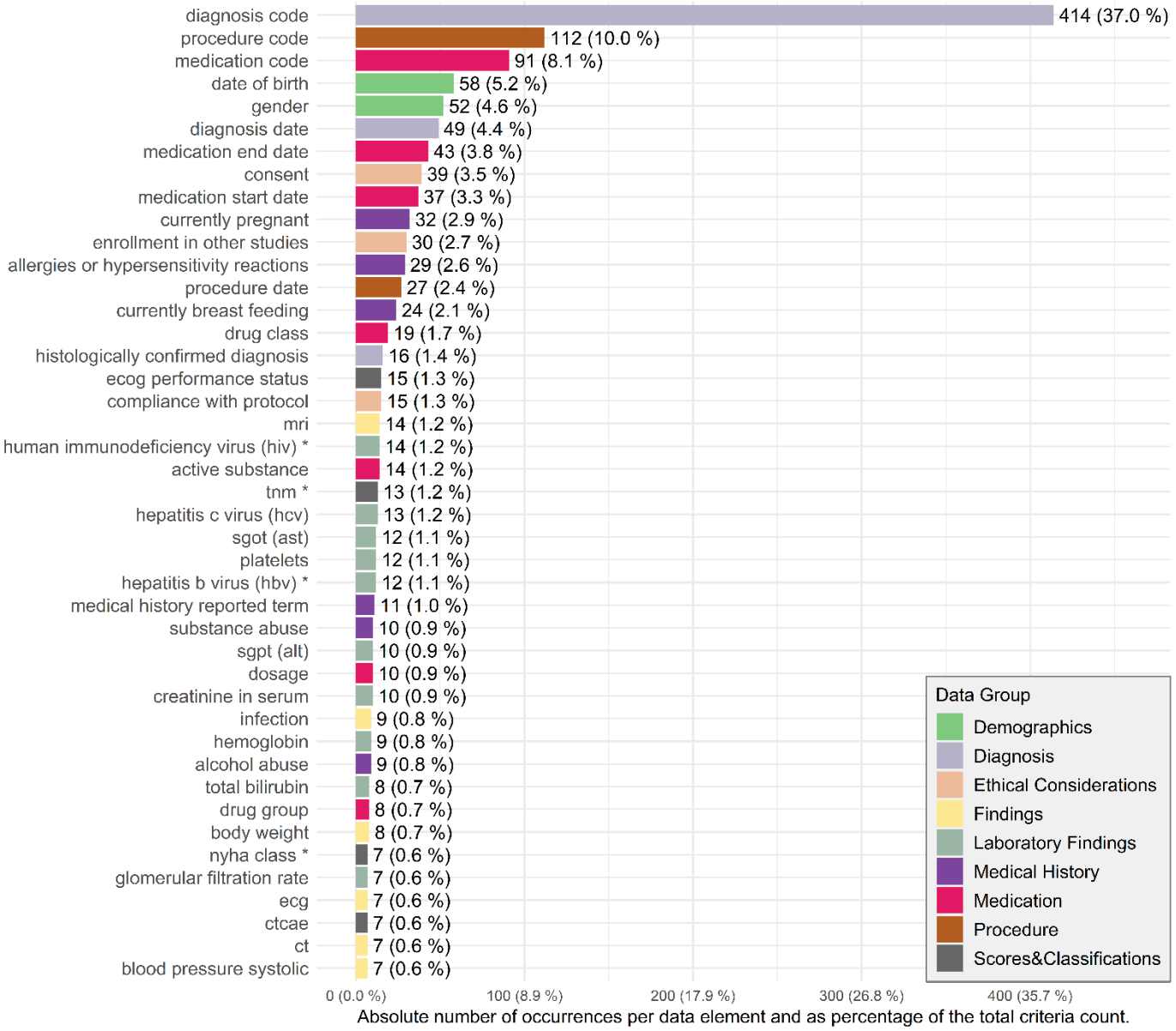
Selected Trials



Extraction of Data Elements

- Total of **1.120** free-text criteria (inclusion + exclusion)
- Mean number of criteria per study: **22** (IQR=18, Range=6-49)
- **130** (12 %) were identified as non-objective
- Total of **1.625** data elements extracted
- **204** unique items remained
  (after harmonization and aggregating duplicates)

Distribution



Discussion & Outlook

Discussion

- Prioritized list of data elements required to be present in the EHR to facilitate semi-automated patient recruitment
- Manual approach is quality-assured, but exceptionally labor-intensive

Outlook

- Automated methods to extract data elements:
  - See work by Weng et al. [1]
  - Automatically assess how well a study may be supported by semi-automated patient recruitment systems.
  - Automatically formalize criteria. See Criteria2Query [2], EligIE [3]
- Future work:
  - Analyzing the data quality of EHRs for the purpose of patient recruitment based on the data in this work

[1] Butler A, Wei W, Yuan C, Kang T, Si Y, Weng C. The Data Gap in the EHR for Clinical Research Eligibility Screening. AMIA Jt Summits Transl Sci Proc. 2018;2017:320–329. Published 2018 May 18.
[2] Yuan C, Ryan PB, Ta C, et al. Criteria2Query: a natural language interface to clinical databases for cohort definition. *J Am Med Inform Assoc*. 2019;26(4):294–305. doi:10.1093/jamia/ocy178
[3] Kang T, Zhang S, Tang Y, et al. EligIE: An open-source information extraction system for clinical trial eligibility criteria. J Am Med Inform Assoc. 2017;24(6):1062–1071. doi:10.1093/jamia/ocx019

Cheers to the MIRACUM UC1 Team!

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
MEDIZINISCHE FAKULTÄT

Erlangen
Christian Gulden, Hans-Ulrich Prokosch
Dresden
Martin Sedlmayr, Christian Kamann, Azadeh Nassirian
Frankfurt
Kristina Götze, Betül Altun
Freiburg
Martin Boeker, Johanna Andrae, Saskia Kiefer, Adrian Tassoni, Inge Landerer
Gießen
Romina Blasini, Florian Eberhardt, Achim Michel-Backofen
Greifswald
Thomas Bahls, Torsten Leddig, Kai Fitzer, Martin Bialke
Mainz
Tony Hauptmann, Marcel Stark, Gunther Höning
Mannheim
Oliver Opitz, Thomas Ganslandt,
Marburg
Jens-Peter Reese, Michael Wittenberg, Thomas Wündisch, Martin Middeke
Magdeburg
Johannes Bernarding, Tim Herrmann, Anke Lux, Philipp Bornfleth

Thank you for your attention!