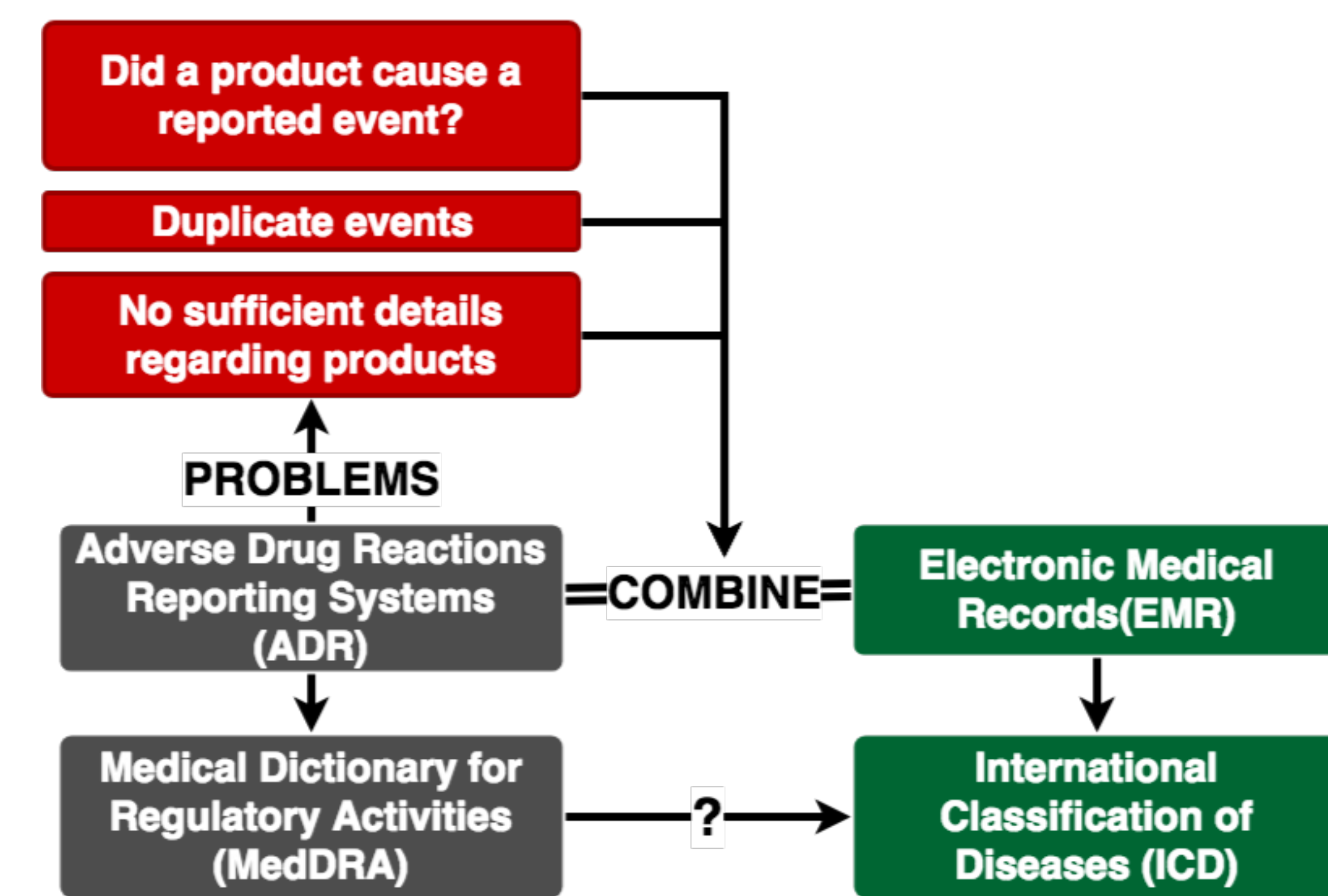


INTRODUCTION



OBJECTIVE

- Examine current MedDRA to ICD mappings
- Suggest candidates from ICD to be mapped with MedDRA terms using an automatic mapping approach

METHODS

CURRENT MAPPING

Analyze current mapping statistics using existing terminology services.

- UMLS:** Terms similar in meaning share the same CUI (Concept Unique Identifier). Corresponding terms in MedDRA and ICD are extracted using CUI as their common feature.
- OHDSI:** Using concept relationships, MedDRA is first mapped to SNOMED, an intermediary data source, then to ICD.

AUTOMATIC MAPPING

TF-IDF (term frequency-inverse document frequency) encoding is a method often used in search engines.

- Extract top 5 scoring results from ICD, for each MedDRA term query.
- Evaluate encoder results using current mappings as gold standard
- Suggest ICD candidates for MedDRA terms that aren't currently mapped

RESULTS

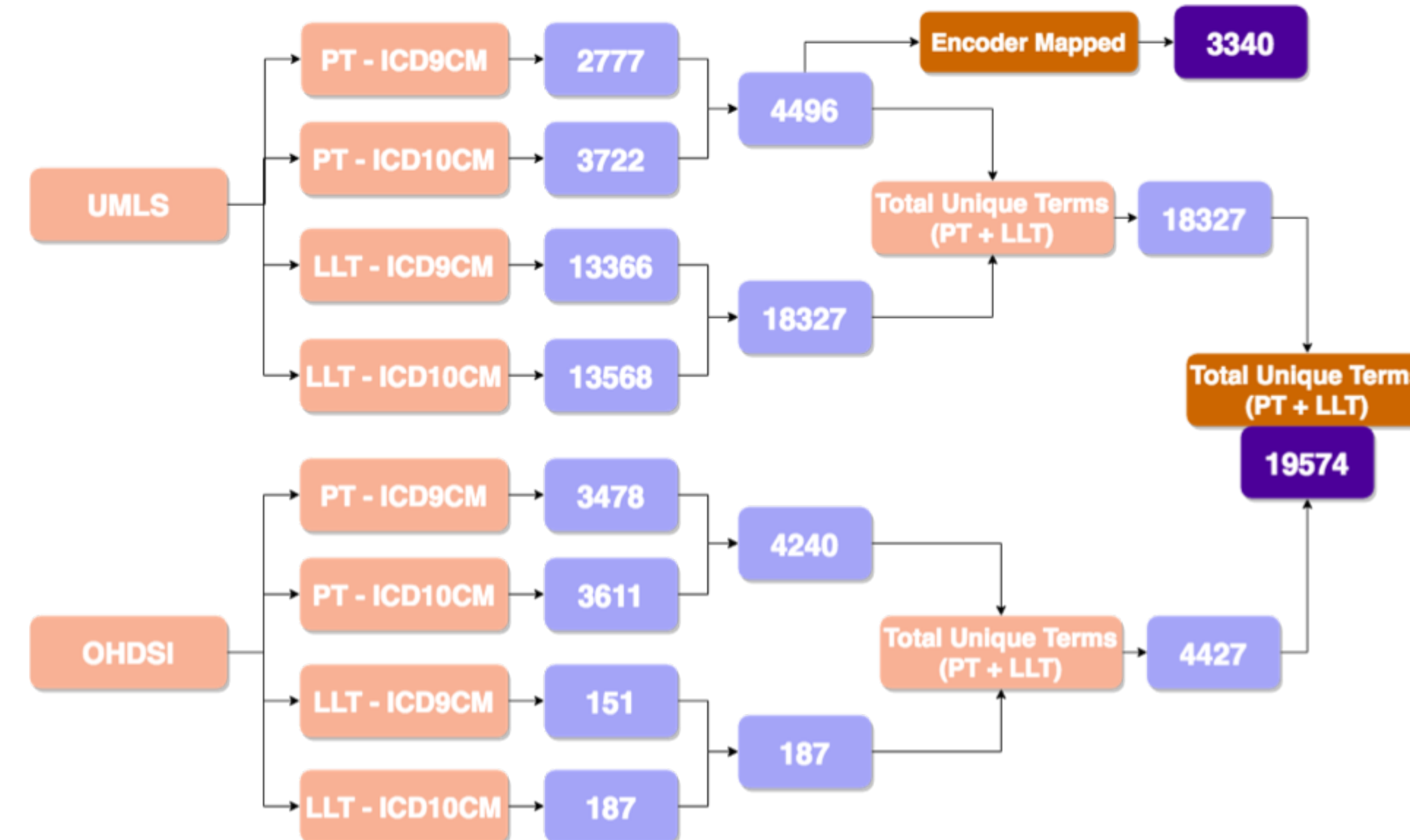


Figure 1. Flowchart of current mapping from MedDRA to ICD

CURRENT MAPPING RESULTS

- Total of 19574 unique terms mapped (see **Figure 1** above).
- All these terms are reclassified and combined using official MedDRA data. Overall 27.309% of MedDRA terms at PT level are mapped. (see **Table 1** above).
- Distribution of the 6305 PT terms mapped among various categories are calculated (see **Figure 2** below).
 - Categories such as Investigations, Social Circumstances, Surgical and Medical Procedures have relatively low mapping coverage.

ENCODER EVALUATION

- Using the 4496 PT terms mapped in UMLS as a golden standard, 74.29% of the terms are correctly mapped by the encoder (see **Figure 1** above).
- The incorrect mapping results are evaluated by a medical professional. Some encoder results can also be mapped to MedDRA in addition to existing UMLS mappings, .

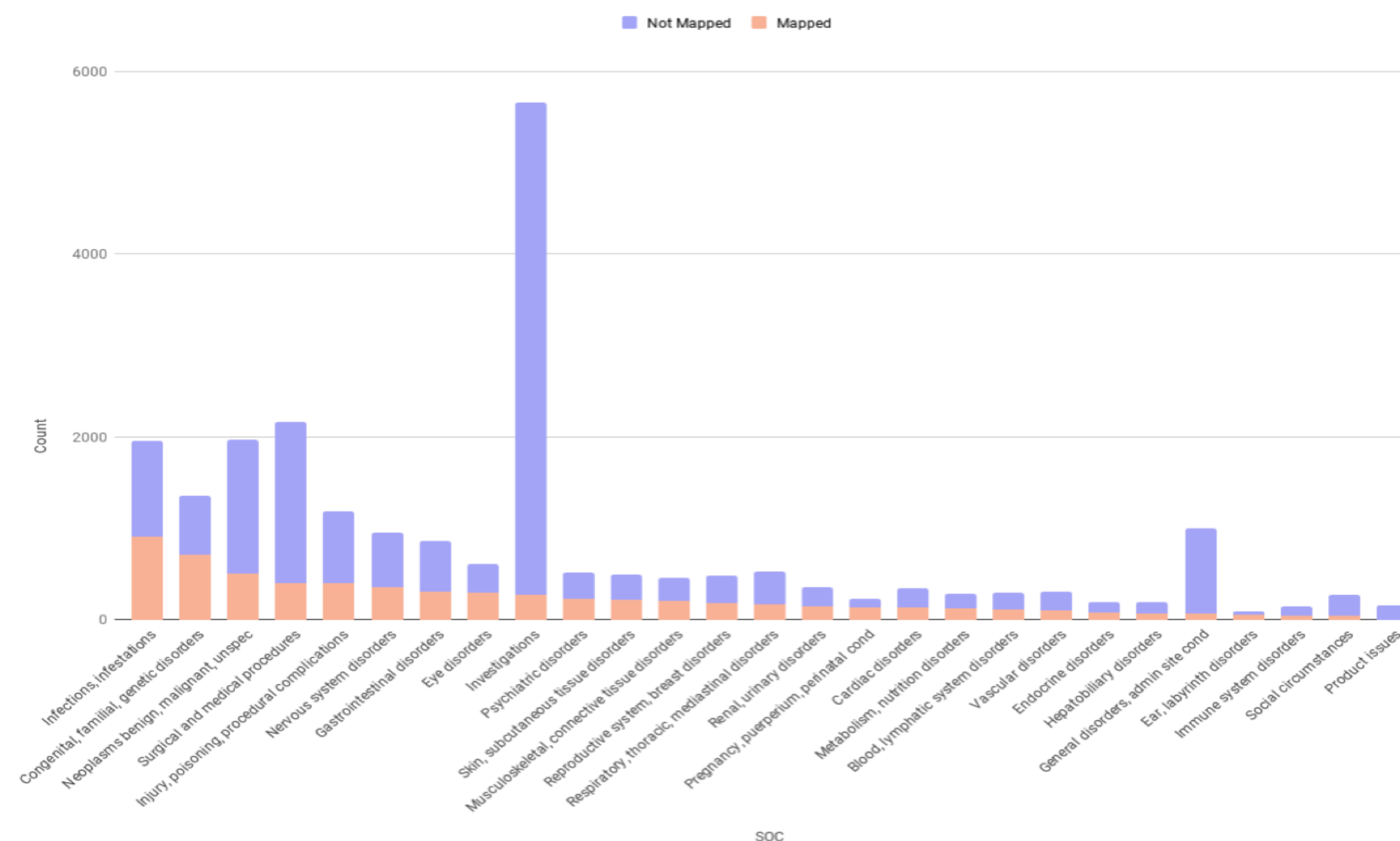


Figure 2. Mapping Distribution Among MedDRA System Organ Class

DISCUSSION

TF-IDF encoder works well in filtering out stop words and providing lexically similar results, but has limitations in certain areas. Below are some potential steps in improving the automatic mapping so that it is better fitted to work with medical terminologies:

LIMITATION	FUTURE STEPS
Providing candidates for long single-word terms for disorders	Parse for stem word, prefix, suffix
Specifying hierarchical relationships between terms	Expand mapping relations complexity to consider hierarchical information (e.g. OHDSI concept relationships)
Identifying syndrome/diseases versus diagnosis/treatment	Increase term multidimensionality by extracting keywords using MeSH

INNOVATION

FRAMESHIFTING: Using automatic tools often used in search engines to analyze similarity between MedDRA and ICD

OBSERVATION: Detected errors in current mapping tools; analyzed why terms are not being mapping by observing

DISSECTING PROBLEM: Analyzed mapping on various MedDRA term levels

REFERENCE

- Center for Drug Evaluation and Research. "Questions and Answers on FDA's Adverse Event Reporting System (FAERS)." U S Food and Drug Administration Home Page. June 04, 2018. Accessed July 30, 2018. <https://www.fda.gov/drugs/guidancecomplianceregulatoryinformation/surveillance/adversedrugeffects/>.

ACKNOWLEDGEMENTS

UTHealth Innovation for Cancer Prevention Research Summer Undergraduate Fellowship [Cancer Prevention and Research Institute of Texas (CPRIT) grant # RP160015] National Library of Medicine of the National Institutes of Health under Award Number R01LM011829. Jingqi Wang, MS, provided resources for the automated mapping process. Alokandanda Ghosh, MD, MS provided evaluation for the encoder mapping results and helped with discussion. The findings and conclusions are those of the author(s) and do not necessarily represent the official position of the funding agency.