Towards Semi-Automated Construction of Laboratory Test Result Comprehension Knowledgebase for a Patient-Facing Application

Zhe HE[a,1], Arslan ERDENGASILENG[b], Shubo TIAN[b], Karim Hanna[c], Yang GONG[d], Xiao LUO[e], Zhan ZHANG[f], Mia Liza A. LUSTRIA[a]
[a] *School of Information, Florida State University*
[b] *Department of Statistics, Florida State University*
[c] *Department of Family Medicine, Morsani College of Medicine, University of South Florida*
Morsani College of Medicine, University of South Florida
[d] *School of Biomedical Informatics, University of Texas Health Science Center at Houston*
[d] *Purdue School of Engineering & Technology, IUPUI*
[d] *Seidenberg School of Computer Science and Inforamtion Systems,. Pace University*

Abstract. Viewing lab test results is the most frequent activity patients do when accessing patient portals but lab results can be confusing for patients. There is a critical need to build tools to help patients better comprehend lab results and formulate questions to prepare for physician consults. In this project, we collected and annotated 251 textual records of lab results comprehension from AHealthyMe.com. To scale up the data collection and curation with automated named entity recognition, we evaluated different transformer-based language models including BioBERT, ClinicalBERT, RoBERTa_trial (RoBERTa model further pre-trained with clinical trial eligibility criteria corpus of ClinicalTrials.gov), and PubMedBERT. Results showed that RoBERTa_trial is the best performing model based on strict criteria and PubMedBERT model has the best performance based on lenient criteria. The models are expected to serve as an AI feature for the design and development of patient-facing technology called LabGenie in promoting lab results communication between providers and patients.

Keywords. Lab Test Results; Lab Portals; Natural Language Processing

1. Introduction

Wider adoption of EHRs and patient portals provides patients greater access to their lab test results. A better understanding of lab test results by patients and their caregivers can improve health behaviors and self-management of chronic conditions, increase medication adherence, and reduce utilization of high-cost healthcare services [1, 2].

[1] Corresponding Author: Zhe He, PhD; Email:zhe@fsu.edu

Greater comprehension of lab results can also lead to more efficient and effective doctor-patient communication, facilitate shared decision making, and consequently, enable patients to receive timely and patient-centered care.

Research shows that viewing lab test results is the most frequently reported activity of patients on patient portals. However, lab test results can be confusing and difficult to interpret. Lab results are often presented in table format with only standard reference points to aid patients. Moreover, while there are abundant online health information resources that are intended to help patients understand their lab test results, the organization of this information is not always convenient and easily accessible for patients and their caregivers. Additionally, these generic resources may not be very helpful or applicable to individual health contexts [4]. This can limit a patient's ability to properly estimate the potential impact of their own results, and thus, delay the receipt of critical medical attention when needed [3].

Therefore, it is urgent to explore how to help patients better comprehend lab results and improve their ability to make informed decisions. One popular approach involves the adoption of user-centered principles in the design of patient portals. Unfortunately, there has been a paucity of theory-driven, rigorous designs that can improve the communication of lab results to patients.

We are developing a patient-facing technology that holds the potential to impact patients' understanding of clinical communication by leveraging a robust knowledgebase that can organize and contextualize lab result-related information in a computable way. In this paper we demonstrate the feasibility of constructing such a knowledgebase by extracting lab test concepts and their relations contained in lab results comprehension records collected from a credible health source AHealthyMe.com. We also evaluated several state-of-the-art transformer-based named entity recognition (NER) models for the automated extraction of lab concepts and relations.

2. Methods

2.1. Data Source

AHealthyMe is a credible health website created and maintained by Blue Cross Blue Shield of Massachusetts. It provides a health library with articles, decision tools, and symptom checkers on a wide range of health and wellness topics. The materials are verified by medical professionals. We crawled the web pages of "Tests & Procedures" category of AHealthyMe.com and obtained 251 de-duplicated records related to lab test results from the sections "Does this test have other names?" and "What do my test results mean?". We focused on these two sections because they contain primary information that can aid in the comprehension of lab results. We refer these sections as "records" throughout this paper.

2.2. Annotation

To identify key information from the records, we annotated the textual data using a web-based annotation tool INCEpTION [5], which provides a semantic annotation platform offering intelligent annotation assistance and knowledge management. It supports the seamless collaboration of multiple annotators, curators, and managers to

ensure the annotation quality. We followed a rigorous process consisting of five steps: (1) annotation guideline development, (2) a pilot test with training, inter-rater reliability assessment, and conflict resolution, (3) separate annotations, (4) curation, and (5) quality assurance. In **Step 1**, PI ZH and the doctoral student AE annotated 37 samples randomly selected from all the records and identified 11 entity types (e.g., "lab name", "normal range", "indication", "condition") and 7 relationships between these entities. PI ZH developed the annotation guideline that explains each entity type and relationship with concrete examples. The annotation guideline was further reviewed by AE. The details about the entity types and relationships with examples are provided in **Table 1**. The relationships between the entities are illustrated in **Figure 1**. In **Step 2**, four undergraduate pre-med students were recruited and trained to annotate the same set of 37 samples following the guideline during the pilot test phase. Interrater reliability was calculated between each student annotator and the PI. Conflicts and inconsistencies were resolved after discussion. In **Step 3**, four students separately annotated subsets of the 251 records. Each record was annotated by two annotators. In **Step 4**, ZH and AE performed the curation and consolidated the annotation. In **Step 5**, we performed quality assurance of the annotations with aggregate analysis of both entity types and relationships and made a few corrections to the annotations.

2.3. Named Entity Recognition with Transformers

To train and test the transformer-based models for automated entity recognition, we split the data into 70% training, 10% validation, and 20% testing. We tested a number of popular transformer-based NER models including BioBERT, ClinicalBERT, RoBERTa_trial (RoBERTa model further pre-trained with clinical trial eligibility criteria corpus of ClinicalTrials.gov in our prior work [6]), and PubMedBERT. We evaluated model performance using precision, recall, and F1 based on both the strict matching criterion (i.e., an exact match of both the entity type and entity with the gold-standard annotated entity) and the lenient matching criterion (i.e., only requires the predicted entity overlap with the gold-standard annotated entity).

Table 1. Entity types, relationship type, number of annotations (#), and examples

Entity Type (#)	Examples	Relationship Type (#)	Examples
lab_name (the name of the lab test) (N= 353)	**Creatinine** (Serum creatinine; blood creatinine)	has_alt_name (has alternative name) (N=616)	"Hemoglobin A1c" *has_alt_name* "HbA1c"
Indication (indication of a test result) (N=730)	A1C from 5.7% to 6.4%. **You may have prediabetes**. This means **you have a higher risk for diabetes in the future**.	has_range (has range or result) (N=714)	"Hemoglobin A1c" *has_range* "below 5.7%"
alt_lab_name (alternative name of a test) (N=620)	Creatinine (**Serum creatinine**; **blood creatinine**)	has_specimen_type (has specimen type) (N=112)	"Albumin" *has_speciment_type* "Urine"
abnormal_range (abnormal value range or result of a test) (N=406)	A1C of **6.5% or above** on 2 separate tests. You may have diabetes.	caused_by (abnormal result caused by) (N=94)	"lower-than-normal level of protein C" *caused_by* "Blood-thinning medicines, such as warfarin"
normal_range (normal range or result of a test) (N=350)	Creatinine (Serum creatinine; blood creatinine): **0.9 to 1.3 mg/dL** for adult males	has_action (has action) (N=67)	"positive for gonorrhea" *has_action* "on treatment to cure the disease"

specimen_type (the specimen type of a lab test) (N=118)	Albumin (**Urine**)	has_condition (has condition) (N=225)	"Creatinine normal range 0.9 to 1.3 mg/dL" *has_condition* "adult males"
Cause (cause of a certain abnormal range/result) (N=91)	A lower-than-normal level of protein C may be caused by: **Blood-thinning medicines, such as warfarin**, **Kidney problems**, **Deficiency in vitamin K, inherited protein C deficiency**.	has_indication (a certain result has a certain indication) (N=728)	"Hemoglobin A1c from 5.7% to 6.4%" *has_indication* "You may have prediabetes. This means you have a higher risk for diabetes in the future**"**
Condition (a certain condition on which a certain result is dependent) (N=95)	Creatinine: 0.9 to 1.3 mg/dL for **adult males**, 0.6 to 1.1 mg/dL for **adult females**.		
age_group (age group) (N=76)	0.5 to 1.0 mg/dL for **children ages 3 to 18** years. 0.3 to 0.7 mg/dL for **children younger than age 3**.		
action (action to take to achieve a goal or given a certain result) (N=71)	To lower the calcium level in your urine, your healthcare provider might suggest that **you eat more vegetables and fruits and less animal products, like red meat and eggs**		
gender (male or female) (N=54)	Creatinine: 0.9 to 1.3 mg/dL for adult **males**		

3. Results

Of all the 251 lab test-related records, our systematic annotation discovered 2,964 annotated entities of 11 entity types and 2,556 relationships of 7 relation types. The numbers of annotations of the entities and relationships are reported in **Table 1**. The numbers of relationships are reported in **Figure 2**. For NER, the best model for strict matching in terms of F1 is PubMedBERT (0.60), followed by RoBERTa_trial (0.59), ClinicalBERT (0.56), BioBERT (0.56). Best F1 score of lenient matching is PubMedBERT (0.82), followed by RoBERTa_trial (0.81), ClinicalBERT (0.81), BioBERT (0.81). As the best performing model,

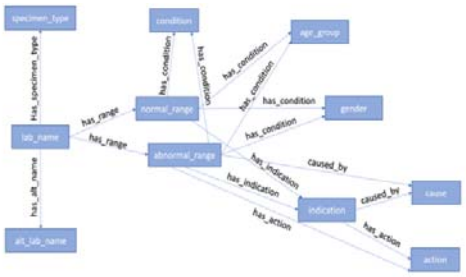


Figure 1. Relationships between entities

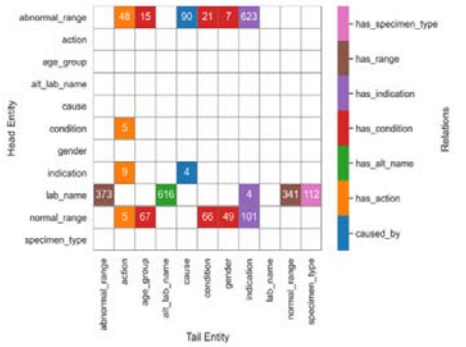


Figure 2. Number of annotated instances of relationships between entities.

PubMedBERT achieves the F1 scores in terms of strict matching for different entity types as follows: alt_lab_name (0.89), specimen_type (0.86), age_group (0.82), gender (0.80), lab_name (0.70), normal_range (0.64), indication (0.41), abnormal_range (0.41), condition (0.28), cause (0.15), action (0.05). For detailed results, see **Table 2**.

Table 2. Performance of transformer models for named entity recognition

	Strict			Lenient		
	Precision	Recall	F1	Precision	Recall	F1
BioBERT	0.5262	0.5974	0.5596	0.7844	0.8406	0.8115
ClinicalBERT	0.5281	0.6055	0.5641	0.7879	0.8374	0.8119
RoBERTa_Trial	0.5461	0.6393	0.5890	0.7799	0.8519	0.8143
PubMedBERT	**0.5670**	**0.6409**	**0.6017**	**0.7949**	**0.8551**	**0.8239**

4. Discussion and Conclusions

In future work, we will perform relationship extraction, store the extracted data in a Neo.5j knowledge graph, and develop a tool based on the knowledge graph to automatically generate tailored question prompt lists to help prepare patients for consultations with their providers. For example, given the lab test result “albumin: 7.3 g/dL”, we will indicate that it is higher than the normal range (3.4 to 5.4 g/dL) based on the annotated interpretation of albumin. Based on the annotation, high albumin level may be caused by “acute infection, burns, and stress from surgery or a heart attack”. We can search for such information in the patient’s EHR. In case the patient had surgery, we can generate questions relevant to their health context that they can ask their physicians (i.e., “Is my low albumin due to stress from surgery?”). We will also investigate other entity and relationship extraction models for complex relation extraction involving conditional relations, which will facilitate automated question generations.

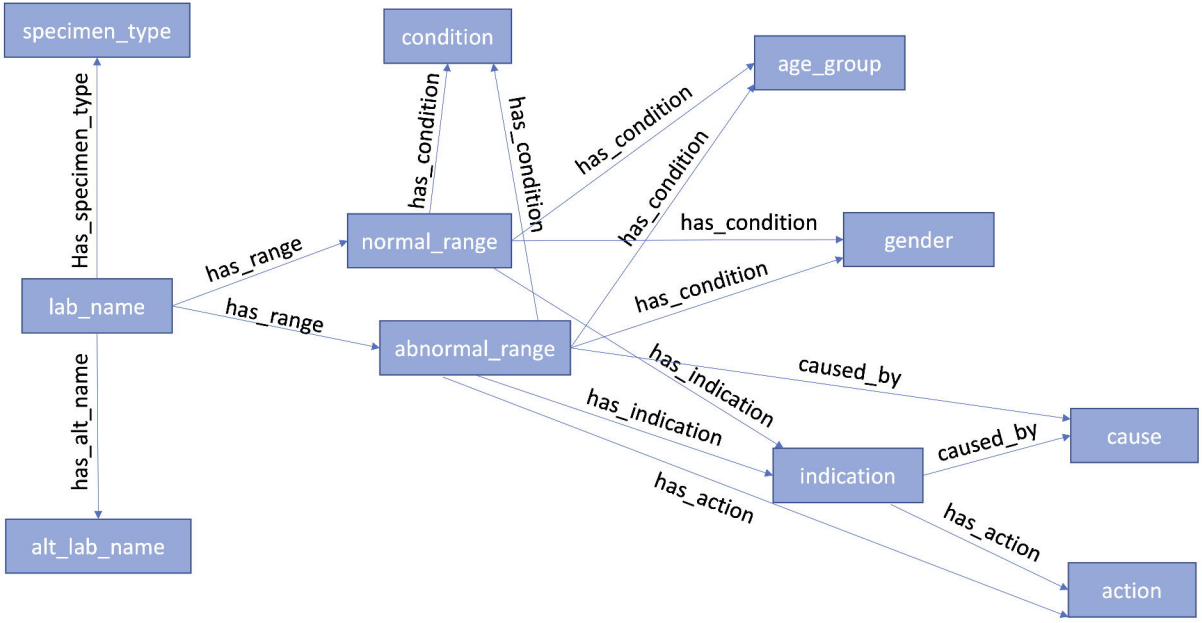
Acknowledgments

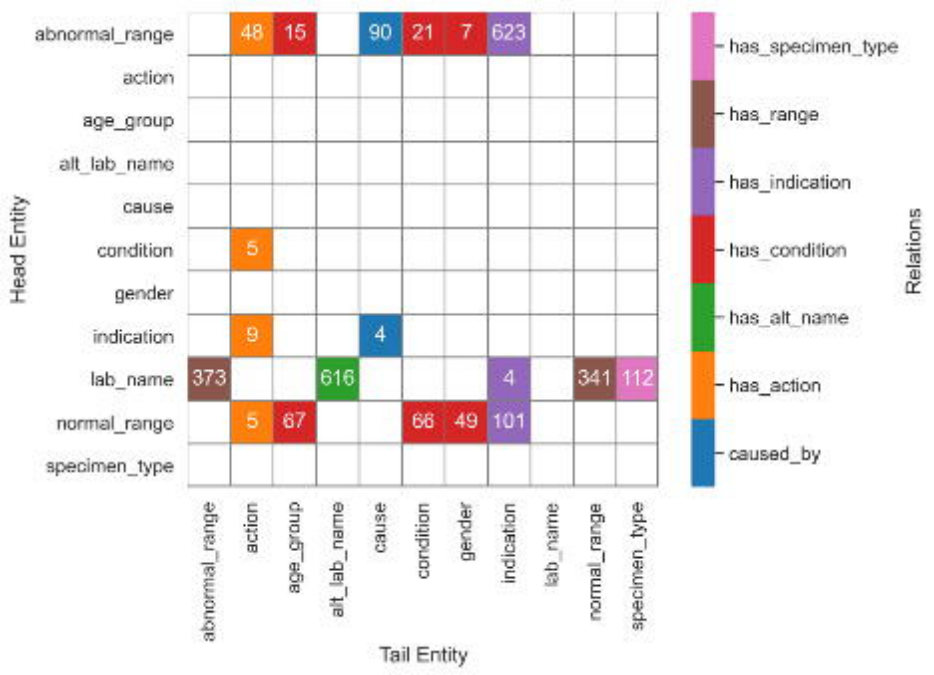
We would like to thank Jessica Valyou, Shawntah Thomas, Maddy Dupuis, and David Garner for annotating the data collected from AHealthyMe.com. This work was supported by the planning grant of FSU Institute for Successful Longevity and UF-FSU Clinical and Translational Science Award under award number UL1TR001427.

References

1. Zikmund-Fisher BJ, Scherer AM, Witteman HO, et al. Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results. *J Am Med Inform Assoc JAMIA*. 2017;24(3):520-528.
2. Zhang Z, Citardi D, Xing A, Luo X, Lu Y, He Z. Patient challenges and needs in comprehending laboratory test results: mixed methods study. *J Med Internet Res*. 2020;22(12):e18725.
3. Zikmund-Fisher BJ, Scherer AM, Witteman HO, Solomon JB, Exe NL, Fagerlin A. Effect of Harm Anchors in Visual Displays of Test Results on Patient Perceptions of Urgency About Near-Normal Values: Experimental Study. *J Med Internet Res*. 2018;20(3):e98. doi:10.2196/jmir.8889
4. Leyland R, Freedman DB. A survey of Lab Tests Online-UK users: a key resource for patients to empower and help them understand their laboratory test results. *Ann Clin Biochem*. 2016;53(6):669-679.
5. Klie JC, Bugert M, Boullosa B, Eckart de Castilho R, Gurevych I. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In: *Proceedings of the 27th International Conference on Computational Linguistics*; 2018:5-9.
6. Tian S, Erdengasileng A, Yang X, et al. Transformer-Based Named Entity Recognition for Parsing

Clinical Trial Eligibility Criteria. *ACM-BCB ACM Conf Bioinforma Comput Biol Biomed ACM Conf Bioinforma Comput Biol Biomed*. 2021;2021:49.


specimen_type
condition
age_group
Has_specimen_type
has_condition
has_condition
has_condition
has_condition
normal_range
has_condition
gender
has_range
lab_name
has_condition
has_range
abnormal_range
has_indication
caused_by
has_indication
cause
caused_by
indication
has_action
has_action
has_alt_name
alt_lab_name
action


Head Entity
abnormal_range
action
age_group
alt_lab_name
cause
condition
gender
indication
lab_name
normal_range
specimen_type
48
15
90
21
7
623
5
9
4
373
616
4
341
112
5
67
66
49
101
Tail Entity
abnormal_range
action
age_group
alt_lab_name
cause
condition
gender
indication
lab_name
normal_range
specimen_type
Relations
has_specimen_type
has_range
has_indication
has_condition
has_alt_name
has_action
caused_by