

AllergyMap: An Open Source Corpus of Allergy Mention Normalizations

Amy Y. Wang, MD, MBI¹, John D. Osborne, PhD¹,
Maria I. Danila, MD, MSc, MSPH¹, Andrew M. Naidech, MD, MSPH²,
David M. Liebovitz, MD²

¹University of Alabama at Birmingham Informatics Institute, Birmingham, Alabama

²Northwestern University, Chicago, Illinois

Abstract

Allergy mention normalization is challenging because of the wide range of possible allergens including medications, foods, plants, animals, and consumer products. This paper describes the process of mapping free-text allergy information from an electronic health record (EHR) system in a university hospital to standard terminologies and migration of those data into an enterprise EHR system. The review, mapping, and migration revealed interesting issues and challenges with the free-text allergy information and the mapping in preparation for implementation in the new EHR system. These findings provide insights that can form the basis of guidelines for future mapping and migration efforts involving free-text allergy data. As part of this process, we generate and make freely available AllergyMap, a mapping between free-text entered allergy medication to standard non-proprietary ontologies. To our knowledge, this is the first such mapping available and could serve as a public resource for allergy mention normalization and system evaluation.

Introduction

Impact of Allergy Documentation Challenges

Allergy mention normalization, in which entities mentioned in allergy text (“allergy mentions”) are identified and mapped to terminologies, is a key step in the process of abstracting meaning from free-text allergy mentions, such as with natural language processing (NLP). However, this normalization is challenging because of the wide range of possible allergens. Adverse drug reactions, including allergic reactions, are a major cause of morbidity, mortality, and increased health care costs, affecting up to 10-25% of patients, with allergic reactions causing 5-10% of these reactions^{1, 2, 3}. Up to 39% of patients have at least one allergy reported^{4, 5}, and patients described as having allergies may actually have drug intolerances that are not caused by allergic reactions⁴. Documentation of prior adverse reactions, such as drug allergies, is often inaccurate and outdated⁶⁻¹⁰. A major challenge to improved documentation is that allergy entries often include free text, which clinicians typically use when they cannot find a term for an allergy in the electronic health record (EHR) system¹¹ and can occur even when comprehensive vocabularies such as SNOMED CT are used, such as in the case of food allergies¹².

These issues with allergy documentation lead to suboptimal and even erroneous prescribing, resulting in allergic reactions and increased costs¹. Less obvious effects include reduced medication choices, such as with antibiotics, with increased antibiotic resistance and fewer available treatment options due to excessive use of broad-spectrum antibiotics^{1, 13-17}. Improving allergy documentation by mapping allergy text to controlled terminologies allows automated checking using computerized physician order entry (CPOE) and clinical decision support systems, improving quality and safety and decreasing adverse reactions¹⁸⁻²¹. However, inaccurate information limits the effectiveness of these systems^{19, 21}.

Data Extraction and Migration

Health systems are increasingly transitioning to new EHR systems for technological, business, and regulatory reasons^{22, 23}. A frequently cited reason for switching to a new EHR system is to obtain enhanced functionality, such as clinical decision support. Optimal communication and full allergy checking with these systems requires documentation of allergies using standard terminologies²⁴⁻²⁷, which requires that text allergy mentions be normalized to a standard terminology. This process is hampered by several challenges, including the lack of comprehensive and standard allergy terminologies and use of proprietary terminologies²⁶ which may have licensing and publishing restrictions. Additionally, allergy information is heterogeneous, representing not only drug preparations but also broad medication classes, drug ingredients, drug combinations, and non-medication allergens, such as foods, plants, animals, microorganisms, and metals²⁴. In addition to non-medication substances, free-text allergy entries can include the absence of specific allergies, which medications are tolerated (absence of an allergy), information about the type and severity of prior hypersensitivity reactions, and misplaced information, such as fluid restriction and

Table 1. Illustrative sample allergy entries.

Text	Status	Type
pollen grass mold	Active	NULL
ultrasound transmission gel	Active	Allergy
cleaning products	Active	NULL
opioid-like analgesic	Active	Allergy
purple grapes	Active	Allergy
plasma	Active	Allergy
chocolate	Active	NULL
general anesthetics	Active	Allergy
*****LATEX ALLERGY****	Active	Allergy
CONTRAST IODINE, LIPITOR	Active	Allergy
all oral hypoglycemics	Active	Allergy
Vicodin, ultram, neurontin	Active	Allergy
NSAIDS/ NORCO	Active	Unknown
alfasfa	Active	Allergy
PCN, ASA	Active	Allergy
gentamycin, clindamycin	Active	Other
beef, beef products, wheat, soy, cheese, preservatives, margarine, beans, corn, milk, sauce, peppers, gravy, paprika, turkey, breeding, oatmeal	Active	Allergy
many others; poor historian	Active	Idiosyncratic
adrenaline	Active	Allergy
eats eggs all right	Active	Allergy
Ibuprophen	Active	Allergy
ibuprophen	Active	Side Effect
ibuprophen	Active	Sensitivity
Cats Dogs Grass Trees Mold	Active	NULL
levaquin	Active	Allergy
Levaquin	Active	Intolerance
Levaquin	Active	NULL
levaquin	Active	Side Effect
levaquin	Active	Toxicity

institutions were not accessible at the time of submission of this publication). The authors reported 95% or better F-measure, precision, recall, and specificity (true negative rate) on the overall test data set. However, overall performance degraded on unique strings, with accuracy and recall dropping to just over 81% and 80% respectively, due mostly to rare singletons and spelling mistakes not handled by their algorithm.

More recent work by Goss et al.²⁹ evaluated allergy information extraction using the MTERMS NLP software.³¹ MTERMS was trained on 500 emergency department clinical notes, with another 400 annotated as a gold standard containing 217 allergen (or “no allergen”) mentions. They achieved over 98% annotator agreement and the system achieved a recall of 91%, precision of 84.4% and a F-measure of 87.6% when evaluating whether the system identified a positive allergen in clinical text. However, linking these allergic reactions to their specific allergen (normalization) yielded only a 69% F-measure. Neither the software nor dataset is publicly available.

This paper describes the mapping of free-text allergy information from a large university hospital to controlled terminologies as part of a data migration to an existing enterprise EHR system used in an academic health system. We describe the creation of AllergyMap, an allergy mention normalization corpus, and make it freely available.

Methods

Source and Annotation for Data Migration

The Northwestern University Institutional Review Board reviewed the study (STU00207067) and determined that it was not research involving human subjects.

pregnancy status²⁵. There is limited scientific literature on EHR transitions and data migrations in general²⁸, even less so for migration of free-text allergy information between EHR systems. Prior work has characterized the current state of allergy documentation, including the widespread use of free text, and described automated extraction and mapping methods, but not in the context of a data migration^{14, 25, 29}.

Ideally, it would be possible to use normalization software to map free-text allergies to standard terminologies. However, while medication vocabularies like RxNorm are kept reasonably up to date, it is much more difficult to track other allergens such as food preparations with many (often unknown) ingredients and an increasing variety of consumer products and chemicals. This problem is compounded by the lack of normalization corpora and widespread use of proprietary vocabularies, making it difficult to develop, let alone evaluate normalization software in this domain. Moreover, there has been limited work in developing NLP algorithms for allergy information extraction, and an evaluation of commercial NLP engines concluded that while automated extraction may facilitate a manual process, manual mapping and review are necessary for ensuring accuracy when creating automated medication lists³⁰.

Pioneering work by Epstein et al.²⁵ at Vanderbilt University used Transact-SQL to identify food and drug allergies in a perioperative information management system, using 24,599 entries from 9,445 records for training and 24,857 entries from 9,430 records for testing. However, the authors acknowledge that this software is likely not portable, as the software and dataset are not publicly available (the link to their lookup tables, which could support development of similar systems at other

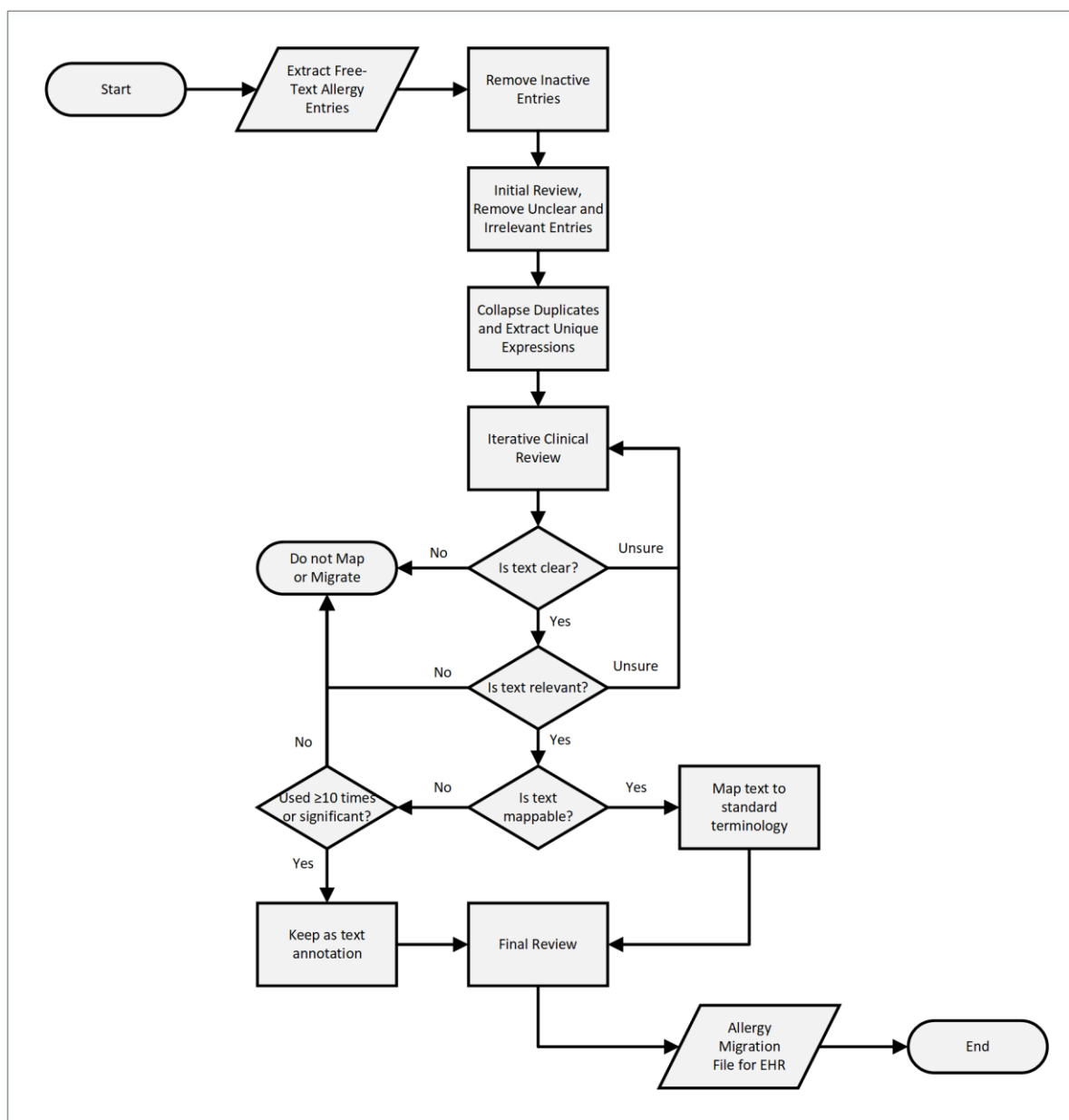


Figure 1. Workflow for mapping and migrating free-text allergy entries.

As part of migration to an enterprise EHR system (Figure 1), all active, free-text allergy entries first recorded between June 27, 2002 and November 25, 2015 were extracted from the existing inpatient EHR system at Northwestern Memorial Hospital in December 2015. All instances in which a text string was documented for a patient's allergy history were treated as separate instances and included duplicates if a string was recorded for multiple patients. Entries that were already recorded using standard terminologies and code sets were excluded. The extracted dataset did not include any protected health information or patient identifiers. Data elements accompanying each text entry included the status of the allergy entry (e.g., active, proposed, resolved, cancelled) and the nature of the adverse reaction, if known (e.g., allergy, side effect, intolerance, unknown). An interdisciplinary team of healthcare professionals, clinical informaticians, and health information technology professionals reviewed the data and discussed different priorities and approaches for the migration.

The review and mappings were performed between December 2015 and November 2016. The data analysis for this study occurred after the allergy mapping and data migration but before go-live of the new patient data in the new EHR system. For the initial extraction performed by the information services team, 14,685 text entries were

retrieved (Table 1), along with the status of the allergy (“Active” or “Cancelled”), and the type of reaction (Table 2). Each text entry is a separate instance in which that string was entered into the EHR system as an allergy annotation.

Review and Mapping

In the first round of processing, 1,678 entries were removed for being ambiguous, irrelevant, or of limited usefulness: health professionals were using the allergy field in the EHR to express other information besides drug sensitivities. Examples include patient weight, pregnancy status, gestational age, fetal weight, allowable medications, non-allergic sensitivities, lack of known allergies (e.g., “NKDA” or “NKA”), and fluid restrictions. Large classes of medications having different subclasses with varied structures, such as “unknown antibiotic,” “unspecified topicals” were deemed too vague to be useful to support allergy checking and were thus excluded from further consideration. After this initial processing, 8,153 entries remained, including duplicates. When duplicate entries were removed, 2,549 unique entries proceeded to the next phase of review.

The clinical informaticians from this group oversaw the iterative mappings and review process. Entries in which the meaning was unclear or in which there were questions were escalated to clinicians for review. On clinical review, those that were found to be ambiguous, irrelevant, or of limited usefulness were not mapped. For entries with multiple drugs or substances, each component was reviewed and mapped separately. Unique strings were manually reviewed and mapped if possible. Preliminary mappings were created from the free-text entries to the standard terminologies used for allergy documentation in the enterprise EHR system. Mappers evaluated names, and if they referred to medications or other allergens, search the terminologies used for documenting allergies in the enterprise EHR system and attempted to assign mappings. Clinical drugs or drug ingredients were mapped to RxNorm. Allergens that could not be readily mapped to RxNorm, such as drug classes and non-drug allergens, were initially mapped to the National Drug File Reference Terminology (NDF-RT), which was used at the time in the target EHR system for these substances. As the goal was to populate patient allergies to allow allergy checking during medication prescribing, terms for food and other environmental allergens were mapped to corresponding drug products or drug ingredients whenever possible. For example, “egg” was mapped to egg as a drug ingredient rather than a food or substance.

For entries that did not match through simple text matching, medical students performed an initial set of manual mappings. Entries that could not be mapped were evaluated for relevance and frequency of use and migrated as text annotations if they were clinically significant or used to document allergy histories ten or more times. One approach would have been to add these entries to our existing or new ontologies in the enterprise EHR system so that they could also be deterministically referenced going forward. However, given the rapid timeline required for this mapping project, the ongoing larger EHR data migration, and migrations of data from multiple EHR systems, we chose to convert these entries into annotations. A follow-up project would be to review all annotations and add these as concepts to the enterprise EHR ontology so they can be used in the future.

Final Review

Physicians with expertise in clinical informatics performed final mapping review and adjudicated mapping decisions to clinical terminologies. Finally, all terms were given a final mapping to a UMLS 2018AB Concept Unique Identifier (CUI) to support mapping to other standard terminologies such as SNOMED CT. For terms already mapped to a concept in a standard terminology represented in UMLS, we used the corresponding CUIs for those concepts. All allergy free text mentions were manually screened to verify the absence of personal health information including all 18 identifiers outlined in the HIPPA privacy rule.

Inter-Annotator Agreement

A set of 100 free-text patient allergy terms was randomly selected and annotated by both AYW, a family physician with expertise in clinical informatics and clinical terminologies, and MID, a rheumatologist who has previous experience mapping clinical text. Annotation concordance was defined by F-measure, a more appropriate metric for computing agreement than traditional measures such as Cohen’s Kappa for this multi-token unrestricted text annotation task³² as F-measure approaches Cohen’s Kappa as the number of negative cases is high. Allergy mention boundaries and assignment of CUIs to mentions was performed using the latest release (1.3) of the BRAT software package (<https://github.com/nlplab/brat>). Within the software, the note field for each annotation was pre-populated with MetaMap based (<https://metamap.nlm.nih.gov/>) CUI information including concept name and semantic type

Table 2. Reaction types.

Reaction Type	Count
Allergy	11,646
Idiosyncratic	12
Intolerance	169
Null	2,033
Other	196
Secondary Effect	4
Sensitivity	26
Side Effect	501
Toxicity	11
Unknown	87

Table 3. Top UMLS mappings and use count.

Rank	UMLS Text	Use Count
1	Iodides	825
2	Dyes	697
3	Penicillin	637
4	Codeine	441
5	Latex	395
6	Aspirin	345
7	Contrast Media	310
8	Adhesive tape (device)	296
9	Morphine	242
10	Erythromycin	240
11	Shellfish	225
12	Intravenous pyelogram	209
13	Anti-Inflammatory Agents, Non-Steroidal	204
14	Amoxicillin	186
15	Filamentous fungus	185
16	Angiotensin-Converting Enzyme Inhibitors	182
17	Ibuprofen	172
18	Fish - dietary	150
19	Meperidine Hydrochloride	144
20	Dust	131
21	Hydromorphone Hydrochloride	125
22	Acetaminophen	118
23	Vancomycin	117
24	Ampicillin	100
25	Clindamycin	98
26	Tetracycline Antibiotics	92
27	Poaceae	89
28	Cephalexin	85
29	Seafood	85
30	Povidone-Iodine	80
31	Cow's milk	80
32	Adhesives	79
33	Tetracycline	78
34	Compazine	78
35	Dog family	77
36	Felis catus	75
37	Bupivacaine Hydrochloride	73
38	Promethazine Hydrochloride	72
39	Nuts	72
40	Family Felidae	71
41	Hay fever	70
42	Pollen	62
43	Trees (plant)	61
44	Morphine Sulfate	53
45	Metronidazole	52
46	Bee sting	51
47	Beef (dietary)	50
48	Sulfonamide Anti-Infective Agents	50
49	Mushroom - dietary	49
50	tetanus toxoid vaccine, inactivated	48

although these were in practice rarely used unless the annotator was already familiar with the UMLS concept. Links to the UMLS Browser were provided for concept lookup and annotators were required to determine the allergy entity boundary, the UMLS concept and to remove the pre-existing MetaMap populated CUI information from the Note field. F-measure was computed by a modified version of the brat-utils software package (<https://github.com/savkov/BratUtils>) which required that true positive values match not only the entity boundary, but also have identical note fields (CUIs). All entities were annotated with the most appropriate single concept and can be post-coordinated with an allergy concept. Therefore, food and drug allergies are represented only by the appropriate food or drug CUI as UMLS lacks a complete cross product of food and drugs concepts with the allergy concept.

Results

Two medical students performed preliminary mappings, and then two physician informaticians (AYW and AMN), including one with expertise in standardized health terminologies (AYW), performed iterative review and mapping. Some entries contained as many as 20 different entities, such as names of medications or other substances. In these cases, each medication or substance was mapped individually. At the end of the mapping process, 2,237 entries were mapped to standard terminologies, and 312 were deemed unmappable. Reasons for not mapping include lack of clarity, ambiguity, an overly broad concept, irrelevance, or nonexistence of an appropriate term in the target terminologies. For unmappable terms, there was further review to determine if the original text would be used to provide a text annotation in the enterprise EHR system. Of the 312 unmappable terms, the clinicians recommended that 170 text entries be used to populate a text field in the allergy history section of the enterprise EHR system and that 142 entries not be migrated. The mapped UMLS concepts of the most common allergens in the dataset and their use counts are shown in Table 3.

For inter-annotator agreement, the F-measure was 0.83 using AYW annotations as the gold standard. which is excellent given the complexity of this task.

At the end of the manual review and mapping process, the mappings and recommended text annotations were imported into the enterprise EHR system and used to populate allergy histories in the destination EHR system. The information technology team at Northwestern Memorial Healthcare (NMH) checked the migrated

allergy data for data quality issues and migrated it into the new EHR system. Physicians performed manual spot-checks on mapped allergy data. The hospital went live on the enterprise EHR system March 3, 2018. Our allergy corpus is freely available at <https://github.com/amywangmd/AllergyMap>.

Discussion

We have created a high-quality allergy mention normalization corpus and have made it freely available. Our work differs from prior work²⁵ in that all entries were manually processed and cleaned to determine if they contained active, legitimate, and nonduplicate entries. This significantly reduced our free-text entry count from 14,685 to 2,549 but was required to generate a high quality corpus because free-text allergies may contain non-allergy medication information such as “Coumadin/atrial fibrillation”²⁹. Our mappings also include environmental allergens

Table 4. Comparison with existing allergy normalization resources.

	Vanderbilt ²⁵	Brigham and Women's Hospital ²⁹	Northwestern University (this work)
Document source	Perioperative text	Emergency notes	Inpatient EHR allergy text
Full note	No	Yes	Yes
Annotated documents	25158 training 25164 testing	400	8153
Unique mentions	3275 training 3662 testing	NA (<=217)	2237
Reactions, environmental allergens	None	Yes	Yes
Mentions not mapped	NA training 723 testing	NA	312

like that of Goss et al.²⁹ but not the sensitivity reactions. A more complete comparison of our work to pre-existing allergy normalization corpora is shown in Table 4.

Mapping Challenges

There were a number of challenges encountered during the mapping (Table 5). Many Text expressions that escaped the initial culling process were later determined to be vague, unclear, or irrelevant, while others had issues that did not affect the meaning (e.g., extraneous characters or punctuation). Broad classes of medications or substances cannot be categorized easily, such as “unspecified topicals,” and “oil.” The free-text fields also contained drug reactions, medications that the patients tolerate, and results of allergy testing. It appears that certain clinicians and even entire departments were using fields to alert health care professionals to important information about the patient, such as fluid restriction status and pregnancy and breastfeeding information. Fields also contained up to 20 or more elements as well as combinations of different types of elements. An important result of this work indicated by our manual review process is that a surprisingly large number of free-text allergy entries described the absence of allergy information or other communication as shown in Table 5. This makes simple allergen look-up approaches prone to error and suggests that even small free-text allergy will require more advanced NLP similar to the larger, allergy containing notes annotated by Goss et al²⁹.

Mapping Approaches

Although free-text allergy information is complex, a systematic process can be used to map it to terminologies in preparation for a data migration to an EHR system. Our experience is consistent with the literature reporting that data migration is the most challenging aspect of an EHR transition. Given the broad implications for all members of the healthcare team that interact with allergies and medications, the project required an interdisciplinary team of health professionals and was iterative.

While automated methods such as NLP would be preferable, to our knowledge, there is no allergy-specific normalization software available, and more general solutions such as MetaMap³³ and cTAKES³⁴ are unlikely to have the performance required for an EHR transition²⁹ where failing to note a patient allergy could be harmful or fatal. In any case, since there are no freely available corpora for allergy mention normalization, we needed to create one before evaluating any NLP software. AllergyMap will help facilitate the creation of such software.

Manual and Automated Methods

Our high inter-annotator agreement suggests that this mapping process is reliable and reproducible. However, an important concern is the heavy reliance on manual review, mapping, and annotation. While manual clinical review serves as a gold standard, it requires significant time, effort, and specialized expertise.³⁵ Thus, our objective is to continue to explore methods for increasing and optimizing automation while improving the workflow and usability of manual mapping.

Clinical Implications

This mapping effort has broader implications than simply migrating and preserving allergy histories faithfully to ensure quality and safety. The availability of an allergy normalization dataset can help with the use of automated

Table 4. Categories of mapping challenges.

Issue	Examples
Vagueness or overly broad categories or not useful (do not map)	"Doesn't remember the drug" "Begins with a C" "all mycins" "almost all antibiotics" "unspecified topicals " "oil"
Not related to allergy (do not map)	"current wt. = 1355g 11/20/2003" "*****PT IS 10 WEEKS PREGNANT 6/15*****" "BREAST FEEDING" "concentrate all fluids" "AVOID LIPID LOWERING AGENTS (PATIENT ON A STUDY)"
Absence of allergy (do not map, can keep as text annotation)	"TESTED PCN NEGATIVE" "oral PCN likely OK per patient" "Diamox is OK" "peanuts ok"
Combination Terms (map allergy only)	"pt. states some abx/unknown name/has had pcn without reaction"
Reaction without allergen (do not map)	"anaphylaxis" "hives" "dystonic reactions"
Concatenation of multiple values (review and map each separately)	"Amoxicillin, erythromycin, PCN, sulfa, cephazil, altor, zithromax, cipro, flagyl, tetracycline, demorol, aspirin, darvocet, bextra, viox, neomycin, toradol, morphine, tylenol, tylenol II, motrin, advil, vicodin."
Extraneous characters (ignore extra characters)	*****L A T E X*****

methods to analyze and migrate free-text allergy information. The optimal outcome would be to migrate all free text to standards for structured entry, better interoperability, and automated checking. As mapping all terms was not possible, we sought to preserve potentially useful information by using the free text to populate text annotations in the destination EHR system. Given the known inaccuracies present in allergy data, showing this information to clinicians and patients provides opportunities to discuss and confirm with patients and improve accuracy during subsequent encounters. A more accurate allergy history improves allergy checking, quality, safety, medication options, and antibiotic stewardship, and reduces adverse effects, unnecessary costs, unnecessary use of broad-spectrum antibiotics, resistance, and complications. Clinicians may trust the information more and be less likely to override automated alerts during electronic prescribing.

Limitations

There a number of limitations of this project. Our allergy data were from a single university hospital in a large metropolitan area within an academic health system. The EHR system that was the source of the allergy data was used only in an inpatient setting. Only the text strings, active/inactive status, and type of reaction were available for the mapping, without additional context such as type of hypersensitivity, severity of allergic reaction, or coded patient diagnoses. However, there was no information available about anatomic locations of reactions. This additional information may have improved the understanding of the text and quality of mappings. Non-drug allergens were mapped to NDF-RT, which was required by the target EHR system but would not be considered ideal or in accordance with current interoperability standards.³⁶

The mapping and review were performed mainly by physicians with expertise in clinical informatics. It would be interesting to have other health professionals who use allergy information perform the same exercise and compare the results. Previous work by Goss et al²⁹ reported an annotator agreement of 98%, which we find reasonable for this task. The primary challenge in allergy mention normalization is the large number of allergens spanning multiple (often proprietary) vocabularies, not difficulty in interpreting text as is often the case with disease mention normalization and other clinical normalization tasks.

Conclusion

Mapping free-text allergy entries to standards and migrating those data to a different EHR system is challenging but presents important opportunities for improving safety, quality, and interoperability. While we have developed a mapping and migration process for preserving intent in free-text allergy information, more significantly, we have developed a freely available corpus that can be used to develop and evaluate NLP allergy mention normalization algorithms. To our knowledge, it is the first freely available corpus of its kind. Future work involves using this corpus to train and develop allergy normalization software that can preserve allergy information and thus improve quality and safety. We also welcome other stakeholder to comment on and contribute to this corpus to make it more robust, current, and applicable to multiple use cases.

Acknowledgments

This work was supported by research funding from the Informatics Institute and the Center for Clinical and Translational Science of the University of Alabama at Birmingham, funded under grant 1TL1TR001418-01 from the National Center for the Advancement of Translational Science (NCATS). This project received additional support from NIH U01 NS1107792, R01 NS110779, and K18 HS023437. The authors also thank Thomas Moran, Northwestern University Feinberg School of Medicine, and Northwestern Memorial Healthcare (NMH) for their contributions to this project. This work would not have been possible without the many people who contributed to the project, including the NMH EHR migration team members who initially extracted and processed the data and Alexander Pyeden and Victor Saunders, the medical students who performed the initial manual review and mappings. The authors also thank James Cimino for his editorial assistance.

References

1. Kuperman GJ, Bobb A, Payne TH, et al. Medication-related clinical decision support in computerized provider order entry systems: a review. *J Am Med Inform Assoc* 2007;14:29–40.
2. Lesar TS, Briceland L, Stein DS. Factors related to errors in medication prescribing. *JAMA* 1997;277:312–17.
3. Gandhi TK, Burstin HR, Cook EF, et al. Drug complications in outpatients. *J Gen Intern Med* 2000;15:149–54.
4. Zhou L, Dhopeswarkar N, Blumenthal K, et al. Drug allergies documented in electronic health records of a large healthcare system. *Allergy*. 2016 Sep;71(9):1305-13. doi: 10.1111/all.12881. Epub 2016 Apr 6.
5. Jones TA Como JA. Assessment of medication errors that involved drug allergies at a university hospital. *Pharmacotherapy*. 2003 Jul;23(7):855-60.
6. Gay KJ, Hill C, Bell T. Accuracy of drug-allergy recording in a district general hospital. *Int J Pharm Pract*. 2009 Aug;17(4):253-5.
7. Lyons N, Rankin S, Sarangarm P, Washington C 3rd, Weiss SJ, Ernst AA. Disparity in patients' self-reported and charted medication allergy information. *South Med J*. 2015 Jun;108(6):332-6. doi: 10.14423/SMJ.0000000000000301.
8. Khalil H, Leversha A, Khalil V. Drug allergy documentation--time for a change? *Int J Clin Pharm*. 2011 Aug;33(4):610-3. doi: 10.1007/s11096-011-9525-y. Epub 2011 May 26.
9. Bowrey DJ, Morris-Stiff GJ. Drug allergy: fact or fiction? *Int J Clin Pract*. 1998;52(1):20-1.
10. Pilzer JD, Burke TG, Mutnick AH. Drug allergy assessment at a university hospital and clinic. *Am J Health Syst Pharm*. 1996 Dec 15;53(24):2970-5.
11. Zimmerman CR, Chaffee BW, Lazarou J, et al. Maintaining the enterprisewide continuity and interoperability of patient allergy data. *Am J Health Syst Pharm*. 2009;66:671–9
12. Plasek, Joseph M., Foster R. Goss, Kenneth H. Lai, Jason J. Lau, Diane L. Seger, Kimberly G. Blumenthal, Paige G. Wickner et al. "Food entries in a large allergy data repository." *J Am Med Inf Assoc*. 2015;23(1): e79-e87.
13. Blumenthal KG, Acker WW, Li Y, Holtzman NS, Zhou L. Allergy entry and deletion in the electronic health record. *Ann Allergy Asthma Immunol*. 2017 Mar;118(3):380-381. doi: 10.1016/j.anai.2016.12.020. Epub 2017 Jan 23.
14. MacLaughlin EJ, Saseen JJ, Malone DC. Costs of beta-lactam allergies: selection and costs of antibiotics for patients with a reported beta-lactam allergy. *Arch Fam Med*. 2000 Aug;9(8):722-6.
15. Satta G, Hill V, Lanzman M, Balakrishnan I. β -lactam allergy: clinical implications and costs. *Clin Mol Allergy*. 2013 Nov 27;11(1):2. doi: 10.1186/1476-7961-11-2.
16. Lutomski DM, Lafollette JA, Biaglow MA, Haglund LA. Antibiotic allergies in the medical record: effect on drug selection and assessment of validity. *Pharmacotherapy*. 2008 Nov;28(11):1348-53. doi: 10.1592/phco.28.11.1348.

17. Shah NS, Ridgway JP, Pettit N, Fahrenbach J, Robicsek A. Documenting penicillin allergy: the impact of inconsistency. *PLoS One*. 2016 Mar 16;11(3):e0150514. doi: 10.1371/journal.pone.0150514. eCollection 2016.
18. Bouwmeester MC, Laberge N, Bussi res JF, Lebel D, Bailey B, Harel F. Program to remove incorrect allergy documentation in pediatrics medical records. *Am J Health Syst Pharm*. 2001 Sep 15;58(18):1722-7.
19. Edwards M, Moczygemba J. Reducing medical errors through better documentation. *Health Care Manag (Frederick)*. 2004 Oct-Dec;23(4):329-33.
20. Topaz M, Seger DL, Slight SP, Goss F, Lai K, Wickner PG, Blumenthal K, Dhopeswarkar N, Chang F, Bates DW, Zhou L. Rising drug allergy alert overrides in electronic health records: an observational retrospective study of a decade of experience. *J Am Med Inform Assoc*. 2016 May;23(3):601-8. doi: 10.1093/jamia/ocv143. Epub 2015 Nov 17.
21. Shen W, Wong B, Chin JY, Lee M, Coulter C, Braund R. Comparison of documentation of patient reported adverse drug reactions on both paper-based medication charts and electronic medication charts at a New Zealand hospital. *N Z Med J*. 2016 Oct 28;129(1444):90-96.
22. West S. Need versus cost: understanding EHR data migration options. *J Med Pract Manage*. 2013;29(3):181-3.
23. Schreiber R, Koppel R, McGreevey JD, Craven CK, editors. What could go wrong? Migrating from one EHR to another. Panel presentation at: AMIA Fall Symposium. 2015 Nov 14-18; San Francisco.
24. Sampalli T, Shepherd M, Duffy J. Clinical vocabulary as a boundary object in multidisciplinary care management of multiple chemical sensitivity, a complex and chronic condition. *J Multidiscip Healthc*. 2011 Apr 14;4:91-102. doi: 10.2147/JMDH.S17564.
25. Epstein RH, St Jacques P, Stockin M, Rothman B, Ehrenfeld JM, Denny JC. Automated identification of drug and food allergies entered using non-standard terminology. *J Am Med Inform Assoc*. 2013 Sep-Oct;20(5):9628. doi: 10.1136/amiajnl-2013-001756. Epub 2013 Jun 7.
26. Goss, Foster R., Li Zhou, Joseph M. Plasek, Carol Broverman, George Robinson, Blackford Middleton, and Roberto A. Rocha. "Evaluating standard terminologies for encoding allergy information." *J Am Med Inform Assoc*. 2013;20(5): 969-979.
27. Topaz M, Seger DL, Goss F, Lai K, Slight SP, Lau JJ, Nandigam H, Zhou L. Standard information models for representing adverse sensitivity information in clinical documents. *Methods Inf Med*. 2016;55(2):151-7. doi: 10.3414/ME15-01-0081. Epub 2016 Feb 24.
28. Zandieh SO, Abramson EL, Pfoh ER, Yoon-Flannery K, Edwards A, Kaushal R. Transitioning between ambulatory EHRs: a study of practitioners' perspectives. *J Am Med Inform Assoc*. 2012;19(3):401-6.
29. Goss FR, Plasek JM, Lau JJ, Seger DL, Chang FY, Zhou L. An evaluation of a natural language processing tool for identifying and encoding allergy information in emergency department clinical notes. *AMIA Annu Symp Proc*. 2014 Nov 14;2014:580-8. eCollection 2014.
30. Jagannathan, Vasudevan, Charles J. Mullett, James G. Arbogast, Kevin A. Halbritter, Deepthi Yellapragada, Sushmitha Regulapati, and Pavani Bandaru. "Assessment of commercial NLP engines for medication information extraction from dictated clinical notes." *Int J Med Inform*. 2009;78(4): 284-291.
31. Zhou L, Plasek JM, Mahoney LM, Karipineni N, Chang F, Yan X, Chang F, Dimaggio D, Goldman DS, Rocha RA. Using Medical Text Extraction, Reasoning and Mapping System (MTERMS) to process medication information in outpatient clinical notes. *AMIA Annu Symp Proc*. 2011;2011:1639-48. Epub 2011 Oct 22.
32. Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. *J Am Med Inform Assoc*. 2005 May-Jun;12(3):296-8. doi: 10.1197/jamia.M1733. Epub 2005 Jan 31. PMID: 15684123; PMCID: PMC1090460.
33. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp*. 2001:17-21. PMID: 11825149; PMCID: PMC2243666.
34. Savova, Guergana K., James J. Masanz, Philip V. Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C. Kipper-Schuler, and Christopher G. Chute. "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications." *J Am Med Inform Assoc*. 2010;17(5): 507-513.
35. Saitwal H, Qing D, Jones S, Bernstam EV, Chute CG, Johnson TR. Cross-terminology mapping challenges: a demonstration using medication terminological systems. *J Biomed Inform*. 2012 Aug;45(4):613-25. doi: 10.1016/j.jbi.2012.06.005. Epub 2012 Jun 28. Erratum in: *J Biomed Inform*. 2012 Dec;45(6):1217. PMID: 22750536; PMCID: PMC4398308.
36. U.S. Office of the National Coordinator for Health IT. 2020 Interoperability Standards Advisory Reference Edition [Internet]. Washington, DC; U.S. Office of the National Coordinator for Health IT; Dec 2019 [cited 2021 Jul 20]. 171 p. Available from: <https://www.healthit.gov/isa/sites/isa/files/inline-files/2020-ISA-Reference-Edition.pdf>