

Research and Applications

A dynamic reaction picklist for improving allergy reaction documentation in the electronic health record

Liqin Wang,^{1,2} Suzanne V. Blackley,³ Kimberly G. Blumenthal,^{2,4} Sharmitha Yerneni,¹ Foster R. Goss,⁵ Ying-Chih Lo,^{1,2,6,7} Sonam N. Shah ,^{1,2,8} Carlos A. Ortega ,¹ Zfania Tom Korach,^{1,2} Diane L. Seger,^{1,3} and Li Zhou^{1,2}

¹Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA, ²Harvard Medical School, Harvard University, Boston, Massachusetts, USA, ³Partners HealthCare, Boston, Massachusetts, USA, ⁴Division of Rheumatology, Allergy, and Immunology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA, ⁵Department of Emergency Medicine, School of Medicine, University of Colorado, Aurora, Colorado, USA, ⁶Division of Nephrology, Department of Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, ⁷Department of Data Science and Big Data Analytics, Providence University, Taichung, Taiwan, and ⁸Massachusetts College of Pharmacy and Health Sciences, Boston, Massachusetts, USA

Corresponding Author: Liqin Wang, PhD, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, 399 Revolution Drive, AR01-13-E58, Somerville, MA 02145, USA (lwang@bwh.harvard.edu)

Received 3 February 2020; Revised 3 March 2020; Editorial Decision 19 March 2020; Accepted 25 March 2020

ABSTRACT

Objective: Incomplete and static reaction picklists in the allergy module led to free-text and missing entries that inhibit the clinical decision support intended to prevent adverse drug reactions. We developed a novel, data-driven, “dynamic” reaction picklist to improve allergy documentation in the electronic health record (EHR).

Materials and Methods: We split 3 decades of allergy entries in the EHR of a large Massachusetts healthcare system into development and validation datasets. We consolidated duplicate allergens and those with the same ingredients or allergen groups. We created a reaction value set via expert review of a previously developed value set and then applied natural language processing to reconcile reactions from structured and free-text entries. Three association rule-mining measures were used to develop a comprehensive reaction picklist dynamically ranked by allergen. The dynamic picklist was assessed using recall at top k suggested reactions, comparing performance to the static picklist.

Results: The modified reaction value set contained 490 reaction concepts. Among 4 234 327 allergy entries collected, 7463 unique consolidated allergens and 469 unique reactions were identified. Of the 3 dynamic reaction picklists developed, the 1 with the optimal ranking achieved recalls of 0.632, 0.763, and 0.822 at the top 5, 10, and 15, respectively, significantly outperforming the static reaction picklist ranked by reaction frequency.

Conclusion: The dynamic reaction picklist developed using EHR data and a statistical measure was superior to the static picklist and suggested proper reactions for allergy documentation. Further studies might evaluate the usability and impact on allergy documentation in the EHR.

INTRODUCTION

Documentation of adverse reactions to drugs, foods, and healthcare product ingredients, such as adhesives and latex, in the electronic health record (EHR) is critical to patient safety. Most EHR systems

have an allergy module—or allergy list—to support this documentation. Despite being called an “allergy list,” diverse adverse reactions beyond allergies (eg, intolerances, contraindications) are documented in this list. Allergy list documentation is a routine compo-

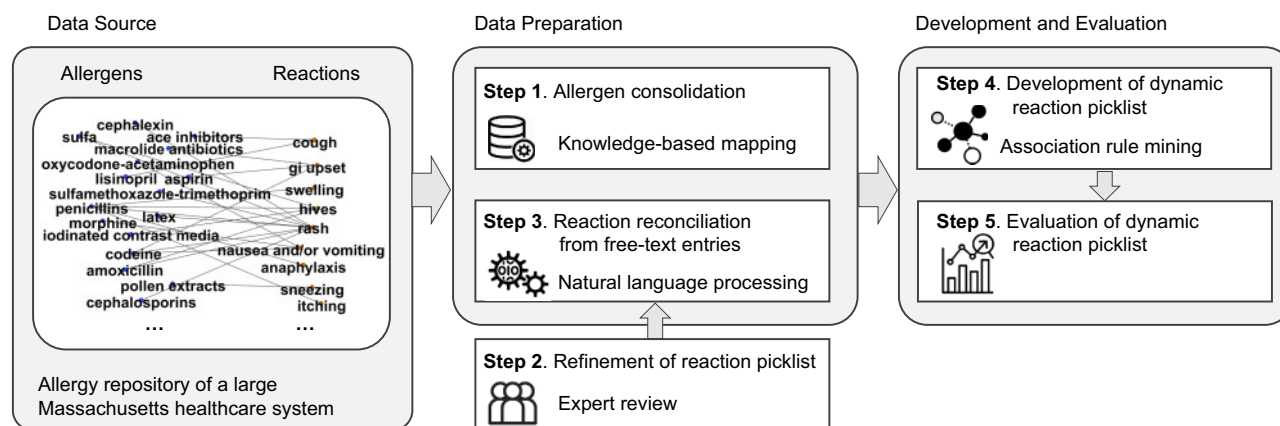


Figure 1. An overview of the workflow for the development and evaluation of the proposed dynamic reaction picklist. The source data were the allergy entries in an allergy repository of a large Massachusetts healthcare system containing allergens and reactions. The data preparation processes include (1) allergen consolidation conducted through knowledge-based mapping, (2) refinement of reaction picklist conducted through exclusive manual expert review, and (3) reaction reconciliation from free-text entries using natural language processing. The dynamic reaction picklist was then developed using association rule-mining measures and evaluated.

nent of ambulatory and hospitalized care¹; more than 35% of patients have at least 1 EHR-documented allergy, and 4% have 3 or more EHR-documented allergies.² While several allergy data elements can be captured in the allergy list (eg, reaction, severity, and type), there are only 2 routinely reported data elements: the culprit agent (also called allergen, such as a drug or food) and the resulting adverse reaction(s), such as rash, hives, or anaphylaxis. Complete and accurate reaction documentation is important to support clinical decision-making and inform safe prescribing in healthcare settings.

The current EHR allergy module has notable limitations that reduce its usefulness for clinical decision support.^{3–10} Most current EHR allergy modules rely on commercial or local dictionaries shown to the user as a single, static picklist and ordered alphabetically. Even among hospitals with the same commercial EHR vendor, reaction picklists vary by site; coded reaction options range from just 10 to more than 100 reactions. For example, a large Massachusetts healthcare system's reaction picklist contains 46 reactions (see [Supplementary Material Table 1](#)), while a Colorado healthcare system's picklist contains a different set of 156 reactions. Without a coded reaction lexicon comprehensive enough to encode the diverse reactions encountered in clinical practice, reaction entries are often incomplete and/or entered as free text.¹¹ Indeed, up to half of reaction fields are left blank, and at least one-sixth of reactions are entered as free text.¹²

The high prevalence of free-text and missing entries may be due to the inadequacy of reaction picklists and/or the inconvenience of searching for the most appropriate reaction using a static alphabetized picklist. To address this, we used natural language processing (NLP) to analyze a large volume of EHR allergen-reaction entries to develop a comprehensive reaction value set of 787 reactions.¹² Although the value set is a comprehensive reaction list, it would be too cumbersome for providers to search the entire list to enter the appropriate reaction(s). As such, in the present study, we developed a novel, data-driven, “dynamic” reaction picklist through a combination of manual review efforts and data-mining techniques. With the dynamic picklist, once an allergen is selected, the reaction picklist is populated with the most probable reactions, thereby allowing providers to easily select coded reaction(s).

MATERIALS AND METHODS

The study method was comprised of 5 steps: (1) allergen consolidation, (2) refinement of reaction value set, (3) reaction reconciliation from free-text entries, (4) development of dynamic reaction picklist, and (5) evaluation of the dynamic reaction picklist (see [Figure 1](#)).

Clinical setting and data sources

This study was conducted at a large Massachusetts healthcare system that includes Massachusetts General Hospital and Brigham and Women's Hospital (Boston). The healthcare system initially used an in-house developed EHR system but switched to a commercial EHR vendor (ie, Epic) beginning in 2015. We used EHR allergy data from all patients within our healthcare system between around March 28, 1980 and October 8, 2019. Data were entered by healthcare team members, and each patient may have 1 or multiple allergy entries, with each allergy entry corresponding to 1 allergen and any number of reactions. The allergy data were separated into development and validation datasets, with the development dataset including allergy entries from 1980 to October 8, 2018, and the validation dataset including entries from October 9, 2018 to October 8, 2019. Inactive allergy entries and entries consisting of “No Known Allergies” were excluded.

Allergen consolidation

The reference allergen table contained 42 027 allergens among which are many synonyms, duplicates, brand names, or other variations. We consolidated allergens by developing a mapping among allergens when they had (1) a versioning history in the database, (2) the same name, or (3) the same ingredients and allergen groups. First, for allergens that were indicated as expired or hidden but still associated with active allergy entries in our database, we mapped them to active allergens using the recorded mapping history. Second, we consolidated allergens with the same names by mapping the less frequently entered allergen(s) to the 1 entered most frequently. Furthermore, we consolidated allergens that had the same ingredients and allergen groups. To identify those allergens, we used the mappings between allergens and allergen groups, between allergens and First Databank (FDB) medications, and between FDB medications

and RxNorm ingredients.¹³ We excluded allergens at the drug class level (eg, penicillins) from the consolidation process, as we tried to avoid merging drug class-level allergens with specific drug allergens under that class. Among allergens with the same ingredients and allergen groups, we chose the allergen with the most entries in the database as the preferred allergen term and considered the rest synonyms. After the allergen consolidation, we updated the allergen entries in our database using the preferred allergen names and generated an allergen and reaction co-occurrence table.

Refinement of reaction value set

We refined the reaction value set through expert panel review, using the following 3-part process: (1) merging 2 reaction lists, (2) identifying clinician preferred terms, and (3) consolidating reaction terms that were overly granular, similar, or less-frequently used.

First, we merged the reaction picklist used in our EHR allergy module with our reaction value set. There were notable differences between the reaction terms included in the 2 datasets. The current EHR reaction picklist contained 46 reactions, including commonly reported reactions (eg, rash, hives, itching) and serious reactions (eg, anaphylaxis, Stevens-Johnson syndrome). It did not include many other common reactions, such as rhinitis, syncope, or bleeding. In a prior study, we identified a comprehensive reaction value set of 787 reactions,¹² but the concepts in this value set were much more granular compared to the EHR picklist. To address this, we mapped the terms in the reaction value set to those in the picklist, using the latter as preferred terms. For example, we mapped “dyspnea” in the reaction value set to “shortness of breath” in the picklist, and similarly mapped “nausea” and “nausea and vomiting” to “nausea and/or vomiting.”

Second, as our goal was to develop a reaction picklist that could be used by clinicians for reaction documentation in the EHR, we updated the preferred terms of the reaction value set based on clinicians’ usage preference. This was accomplished by calculating the occurrence of the SNOMED CT¹⁴ preferred terms and their synonyms in the allergy database. We then identified a subset of reaction concepts whose synonyms were more frequently entered by clinicians in the allergy database than the SNOMED CT preferred terms. These reaction concepts and the most frequent synonyms were manually reviewed by at least 2 clinicians (FG, SNS, ZTK, KGB) to determine preferred reaction terms for data entry. For example, “blood in urine” was the preferred term for a SNOMED CT concept (ID: 34436003), but its synonym “hematuria” was much more frequently mentioned in the free-text comments in our allergy database and was therefore chosen as the preferred term in the refined reaction value set. Similarly, while “dyssomnia” was the preferred term in SNOMED CT, “sleep disturbance” occurred more frequently in our database and, as such, was chosen as the preferred term in the picklist.

Third, an expert panel with allergy subject matter expertise (FG, KGB, SY, LZ, YL, LW) manually reviewed the merged reaction list to further evaluate and consolidate the included reactions. Specifically for the final reaction picklist, we consolidated terms with similar meanings (eg, mapping “feeling nervous” to “anxiety”) and terms with parent-child relationships (eg, mapping “finger swelling” and “hand swelling” to “swelling of extremities”). We also identified and removed general or vague terms (eg, “toxicity,” “addiction,” “allergic reaction”) and condition terms (eg, “diabetes,” “bronchitis”) from the list. All decisions to merge or remove reactions were agreed upon by at least 2 of the experts.

Reaction reconciliation from free-text entries

We used an NLP tool suite called Medical Text Extraction, Reasoning, and Mapping System (MTERMS)^{15,16} to process the free-text comments field in our database to extract any reaction terms mentioned. For this task, we used a knowledge-based approach within the MTERMS allergy module. Using a lexicon,¹² it extracted the reaction terms used by our clinicians and mapped them to the corresponding preferred terms in the refined reaction value set. It also handled contextual information, such as negations. For each allergy entry, we excluded any reactions found in free text that were also among the coded reactions for that entry. This resulted in an updated table of allergy entries where each allergy entry contained all unique reactions in the refined reaction value set.

Development of the dynamic reaction picklist

Based on the updated allergen-reaction entries, we developed the dynamic reaction picklist. We explored 3 approaches to ranking reaction concepts with respect to a given allergen substance: *support*, modified lift (*lift'*), and derived term frequency inverse document frequency (ie, *tf-idf'*).^{17,18} The formulations of the 3 measures that we used to develop the dynamic picklist are defined in the following formulas. $P(x)$, $P(y)$, and $P(x, y)$ are the proportions of allergy entries containing allergen x and reaction y and co-occurring allergen x and reaction y , respectively. $f_{x,y}$ is the frequency of allergy entries in the database in which x and y co-occurred. f_x is the frequency of allergy entries containing allergen x , N is the total number of allergens, and n_y is the number of allergens that co-occurred with reaction y . Note that the original formula for *lift* tends to raise the association scores for extremely rare reactions, such as those with only 1 occurrence. We added the square root function to the $P(y)$ to reduce the effect of the lift of rare reactions, thus the modified *lift* is denoted as *lift'*.

$$\text{support}(x \rightarrow y) = \frac{P(x, y)}{P(x, y)} \quad (1)$$

$$\text{lift}'(x \rightarrow y) = \frac{P(x, y)}{P(x) \sqrt{P(y)}} \quad (2)$$

$$\text{tfidf}'(x \rightarrow y) = \frac{f_{x,y}}{1 + f_x} * \log\left(\frac{N}{1 + n_y}\right) \quad (3)$$

The above formulas were applied only to those reactions that co-occurred with an allergen. If an allergen did not co-occur with all of the reactions in the reaction value set, it was impossible to form a dynamic reaction picklist for that allergen that includes the full value set. Thus, for each allergen, we ranked the reactions that never co-occurred with that allergen by the frequency of each reaction’s occurrence in the development dataset and concatenated that list to the list of reactions that did co-occur with that allergen. This ensures that each allergen will have a fully ranked reaction picklist.

Evaluation of the dynamic reaction picklist

We used retrospective allergy entries documented between October 09, 2018 and October 08, 2019 to evaluate the relevance, effectiveness, and usefulness of the proposed dynamic picklist. Allergy entries with new allergens that never occurred in the development dataset or without any associated reactions were excluded from the validation dataset.

To choose proper evaluation metrics, we considered this ideal scenario: When a clinician documents an allergy in the EHR, after choosing the allergen, the dynamic reaction list should suggest the relevant reactions among the top- k returned list. The smaller the k , the more efficiently the clinician can select the correct reaction. If the user cannot find the desired reaction(s) among the top k , they might search among the entire list, in which case the dynamic reaction list becomes less use-

Table 1. Characteristics of the development and validation datasets from the EHR allergy database

Characteristics	Allergy Data		
	Development	Validation	Total
Patients, n	1 683 245	272 108	1 850 444
Allergy entries, n	3 743 553	490 774	4 234 327
Active allergy entries, n (%)	3 280 743 (87.6)	475 583 (96.9)	3 756 326 (88.7)
Allergy entries with free-text comments, n (%)	1 535 657 (41.0)	156 193 (31.8)	1 691 850 (40.0)
Allergy entries with at least 1 reaction, n (%)	2 171 548 (58.0)	276 348 (56.3)	2 448 538 (57.8)
Allergens before consolidation, n	10 265	5968	11 181
Allergens after consolidation, n	6720	4818	7463
Allergens after consolidation and have at least 1 associated reaction, n	5553	3419	6127
Unique reactions used after the refinement and reaction list reconciliation, n	468	459	469

ful. Therefore, in our evaluation, we investigated the performance of the dynamic reaction picklist with the top-k suggested reactions at $k \leq 15$ under the assumption that a list of more than 15 terms on a single screen might be less acceptable to clinicians. Precision and recall at top k are 2 classic evaluation metrics for recommendation systems, where precision at k is the proportion of recommended items in the top-k set that are relevant, while recall at top k is the proportion of relevant items found in the top-k recommendations. We argue that, in our application scenarios, it is more important to have all true positive reaction terms among the suggested top k ($k \leq 15$) reactions (recall), while it is less important to measure the proportion of positive reactions to the entire suggested list (precision) as the number of true positive reactions was often small, ranging from 1 to a few. Therefore, we measured the performance of the dynamic reaction list in terms of the recall at top k ($k \leq 15$), which is the proportion of true positive reactions among the top k suggested reactions by the dynamic reaction picklist to the total number of true reactions in the validation dataset. In addition, we compared the performance of 3 dynamic reaction picklists ranked by *support*, *lift*, and *tfidf* and the static reaction picklist, where the static reaction picklist was a list of reactions ranked by their total occurrence frequency in the development dataset, irrespective of allergen chosen. We used the Friedman test to verify the significance of the difference, followed by pairwise comparisons using the Nemenyi posthoc test.¹⁹

RESULTS

As of October 8, 2019, the allergy database contained 4 234 327 allergy entries of 1 850 444 unique patients among which 3 756 326 were active allergy reaction entries (excluding “No Known Allergies” entries and allergy entries marked as inactive) with 2 289 511 active allergy entries having 1 or more reactions. The development dataset contained 3 743 628 allergy entries of 1 683 678 patients (see Table 1). Among these, 3 280 743 (87.6%) were active allergy entries, 1 535 657 (41.0%) were free-text reaction comments, and 2 171 548 (58.0%) allergy entries had at least 1 reaction in the refined reaction value set. The validation dataset contained a total of 490 774 allergy entries corresponding to 272 108 patients. Among these, 475 583 (96.9%) were active allergy entries, 156 193 (31.8%) were free-text reaction comments, and 276 348 (56.3%) allergy entries had at least 1 reaction.

The allergen reference table in our EHR database contained a total of 42 639 allergens, but only 26.2% ($n = 11\,181$) of these allergens had allergy entries in patient records, with 24.1% ($n = 10\,265$) of allergens in the development dataset, and 14.0% ($n = 5968$) of

allergens in the validation dataset. Through the allergen consolidation process, a total of 13 177 (30.9%) allergens were mapped to other allergens in the allergen reference table. First, a total of 5668 allergens which were indicated as “deleted” or “hidden” were simply mapped to their current allergen versions. Second, 3438 allergens were mapped to other allergens by name. Furthermore, a total of 4071 allergens were mapped to other allergens with which they shared allergen groups and RxNorm ingredients. After consolidation, the number of allergens was reduced from 42 639 to 29 462. Among the consolidated allergen list, 7463 allergens were present in the entire allergy dataset, with 6720 (90.0%) in the development dataset and 4818 (64.5%) in the validation dataset. However, counting only those allergens that had at least 1 recorded resulting reaction, 6127 allergens were present in the entire allergy dataset, with 5553 in the development dataset, 3419 in the validation dataset, and 3265 in both datasets.

During the manual refinement of the reaction lexicon, the original reaction value set of 787 reaction terms was consolidated to 490 reactions. The refined reaction value set was submitted to the National Library of Medicine (NLM) Value Set Authority Center (VSAC). [Supplementary Material Table 2](#) shows the full refined reaction value set with mappings to SNOMED CT concepts. The 50 most frequent reactions are listed in [Table 2](#).

We used the development dataset, or the 2 171 548 allergy entries, to create the dynamic reaction picklists for the allergens occurring in the development dataset. [Table 3](#) shows the 15 most common reactions for amoxicillin and ibuprofen ranked by 3 different association measures. We validated the performance of 3 dynamic reaction picklists and the static reaction picklist using the recall of the top-k suggested reactions. [Figure 2](#) shows the recalls at top k ($k \leq 15$) of the dynamic reaction picklists ranked by *support*, *lift*, and *tfidf* and the static picklist ranked by the reaction frequency in the database. For example, the recall at top 15 for *support* is 0.822 indicating that, if we suggest the top 15 reactions ranked by *support*, 82.2% of reactions in the validation dataset were among the top 15 suggested reactions. The dynamic reaction picklist ranked by *support* achieved recalls of 0.632, 0.763, and 0.822 at top 5, 10, 15, respectively, while *tfidf* achieved recalls of 0.593, 0.738, and 0.804, and *lift* achieved recalls of 0.533, 0.674, and 0.749 at top 5, 10, 15, respectively. The static reaction picklist ranked by frequency achieved recalls of 0.504, 0.655, and 0.734 at top 5, 10, and 15, respectively. Using the Friedman test, the 4 picklists showed significant differences in their recalls at top k ($k \leq 15$), with P value $< .001$. With the posthoc test, the best dynamic picklist ranked by *support* significantly outperformed the static picklist in terms of the recall at

Table 2. Top 50 most prevalent adverse reactions in the reaction picklist

Ranking	Reaction	Allergy entries (N = 3 743 553) n (%)	Ranking	Reaction	Allergy entries (N = 3 743 553) n (%)
1	Rash	557 188 (14.9)	26	Hypotension	12 090 (0.3)
2	Hives	339 952 (9.1)	27	Rhinitis	11 199 (0.3)
3	GI upset	197 981 (5.3)	28	Musculoskeletal pain	11 108 (0.3)
4	Nausea and/or vomiting	171 474 (4.6)	29	Pain	9891 (0.3)
5	Itching	139 414 (3.7)	30	Dystonia	9758 (0.3)
6	Anaphylaxis	118 274 (3.2)	31	Syncope	9163 (0.2)
7	Swelling	88 620 (2.4)	32	Facial swelling	8612 (0.2)
8	Cough	59 154 (1.6)	33	Throat tightness	7878 (0.2)
9	Mental status change	53 961 (1.4)	34	Edema	7870 (0.2)
10	Shortness of breath	46 713 (1.2)	35	Throat swelling	7691 (0.2)
11	Angioedema	46 651 (1.2)	36	Tongue swelling	7573 (0.2)
12	Diarrhea	37 997 (1.0)	37	Seizures	7467 (0.2)
13	Headaches	35 765 (0.1)	38	Swelling of extremities	6996 (0.2)
14	Dizziness	31 974 (0.9)	39	Lip swelling	6893 (0.2)
15	Myalgia	27 511 (0.7)	40	Agitation	6555 (0.2)
16	Palpitations	24 948 (0.7)	41	Bleeding	6521 (0.2)
17	Bronchospasm	24 041 (0.6)	42	Dermatitis	6481 (0.2)
18	Sneezing	23 911 (0.6)	43	Nasal congestion	6378 (0.2)
19	Wheezing	18 364 (0.5)	44	Confusion	6221 (0.2)
20	Anxiety	17 075 (0.5)	45	Abdominal pain	6090 (0.2)
21	Erythema	16 128 (0.4)	46	Chest pain	5887 (0.2)
22	Flushing	15 105 (0.4)	47	Renal toxicity	5753 (0.2)
23	Fever	14 751 (0.4)	48	Burning sensation	5601 (0.1)
24	Fatigue	13 135 (0.4)	49	Blisters	5587 (0.1)
25	Hallucinations	12 732 (0.3)	50	Gastrointestinal hemorrhage	5584 (0.1)

Table 3. Top 15 most common reactions for Amoxicillin and Ibuprofen using 3 association measures

Rank	Top 15 most common reactions for Amoxicillin			Top 15 most common reactions for Ibuprofen		
	<i>Support</i>	<i>Lift'</i>	<i>Tfidf'</i>	<i>Support</i>	<i>Lift'</i>	<i>Tfidf'</i>
1	Rash	Rash	Rash	GI upset	GI upset	GI upset
2	Hives	Hives	Hives	Hives	Hives	Hives
3	GI upset	Diarrhea	GI upset	Rash	Swelling	Swelling
4	N/V	GI Upset	Diarrhea	Swelling	GI Hemorrhage	Rash
5	Itching	Itching	N/V	N/V	Bleeding	N/V
6	Diarrhea	N/V	Itching	Anaphylaxis	Facial swelling	GI Hemorrhage
7	Swelling	Maculopapular rash	Swelling	Itching	Rash	Angioedema
8	Anaphylaxis	Vaginitis	Anaphylaxis	Angioedema	Angioedema	Anaphylaxis
9	SOB	Swelling	Vaginitis	SOB	Gastritis	Bleeding
10	Angioedema	Erythema multiforme	Maculopapular rash	GI Hemorrhage	Ulcer	Itching
11	Vaginitis	Anaphylaxis	Angioedema	Bronchospasm	Peptic ulcer	Facial swelling
12	Fever	Rash w/ joint pains ^a	SOB	Facial swelling	Eye swelling	SOB
13	Maculopapular rash	SOB	Fever	Bleeding	Renal toxicity	Renal toxicity
14	Erythema	Morbilliform rash	Rash w/ joint pains ^a	Wheezing	N/V	Bronchospasm
15	Facial swelling	Angioedema	Erythema multiforme	Lip swelling	Lip swelling	Wheezing

Abbreviations: GI hemorrhage, gastrointestinal hemorrhage; N/V, nausea and/or vomiting; SOB, shortness of breath.

^aRash w/joint pains (serum sickness).

top k ($k \leq 15$), P value $< .001$. [Supplementary Material Table 3](#) shows the top 15 relevant reactions associated with certain common drug allergens including penicillins, sulfonamide antibiotics, amoxicillin, and codeine, ranked by *support*.

DISCUSSION

In this study, we identified an allergy reaction value set consisting of 490 reaction concepts for EHR allergy documentation and dissemi-

nated this refined value set to the NLM VSAC. We created a dynamic reaction picklist that ranked reactions dynamically based on specific allergen by leveraging association rule-mining algorithms and more than 3 decades of allergy entries of a large Massachusetts healthcare system. In validating the dynamic reaction picklist using 1 year of retrospective allergy entries, we found that among the 3 association-mining measures (ie, *support*, *lift'*, and *tfidf'*) for generating the dynamic reaction picklist, *support* had the best recall for the top-ranked reactions outperforming the other 2 ranking algo-

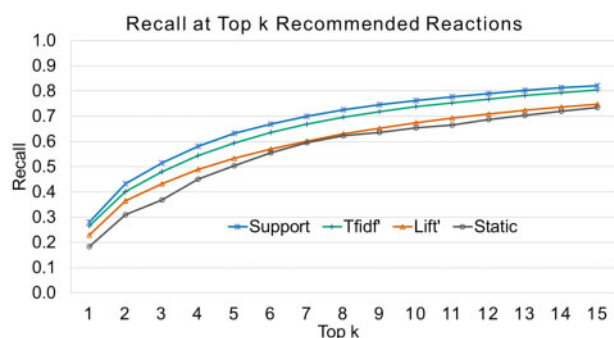


Figure 2. The recalls at top k of the dynamic reaction picklists ranked by *support*, *tfidf*, and *lift* and the static reaction picklist ranked by reaction frequency.

algorithms. The dynamic reaction picklist ranked by *support* significantly outperformed the static reaction picklist ranked by frequency, indicating that it was able to suggest more probable reactions for allergy documentation. This study contributes to the standardization of a reaction value set for EHR allergy documentation as the value set was developed based on a standard terminology (ie, SNOMED CT) and can be easily adopted or adapted across healthcare systems. The study also contributes to a novel conceptual implementation of dynamic reaction picklists that will support the development of a prototype allergy entry module that has the potential to vastly improve EHR allergy reaction documentation. The novel idea of a dynamic picklist is generalizable and can be applied to other domains to facilitate efficient and accurate EHR documentation.

During the course of the reaction value set development, we identified multiple challenges related to using proprietary knowledge bases or standard terminologies for such tasks. Most current commercial EHR systems allow users to configure the clinical content of different modules. Users often purchase proprietary knowledge bases which they supplement with a certain level of customization. However, for adverse reactions, no standard or commercial products are available; therefore, the number and type of reactions included in the reaction picklist often varies across EHR implementations. In a prior study, we attempted to develop a reaction value set using SNOMED CT, a large and comprehensive standard terminology, and free-text allergy reaction entries.^{12,14} While SNOMED CT might be used to generate a value set for the reaction picklist, certain challenges remain. First, there are some synonymous concepts not listed as synonyms, such as “throat tightness” and “tightness in throat.” Second, the concepts in SNOMED CT are highly granular, and clinicians may not find it clinically meaningful to differentiate between similar but distinct concepts (eg, “itchy rash” and “red rash” versus “rash”). In addition, the free-text entries in the allergy database contained concepts related to patient diagnoses (eg, “pneumonia”) and unspecific language (“changes in skin texture”, “toxic effect”), introducing noise to the reaction value set. These issues limited our ability to directly use SNOMED CT to develop the reaction value set and dynamic picklists. Therefore, extensive manual curation combined with data-mining was necessary in order to develop a high-quality reaction value set.

The refinement of the reaction value set was conducted by a team of diverse backgrounds using a systematic approach. The advising expert panel included an allergist, an emergency physician/clinical informatician, a nephrologist/clinical informatician, a pharmacist, and 2 additional clinical informaticians. When updating the

clinician-preferred terms, we used a data-driven approach so that the majority of selected terms were based on the frequency with which they were entered in the database, although some preferred terms were based on the experts’ choices and agreed upon by at least 2 experts. When reducing granular terms, we not only consolidated terms of similar meanings but also merged some child-level (ie, more specific) concepts to their parent terms based on the experts’ opinions. For example, while there are a total of 38 SNOMED CT concepts related to pain, we maintained only those concepts describing pain in a major body location with distinct and meaningful clinical interpretations (eg, “back pain,” “chest pain,” “abdominal pain,” “joint pain,” “neck pain”), and merged pain in other locations (eg, “pain in face,” “foot pain,” “pain in upper limb,” “aching pain”) with the concept of “pain.” Applying this rule, the number of pain concepts was reduced from 38 to 11. Although we believe these changes resulted in an improved value set, further investigation is required to understand how these reaction changes impact EHR allergy documentation.

The allergen consolidation process, although conducted to facilitate development of dynamic reaction picklists, revealed problems with our EHR’s existing knowledge base, including duplicate entries and similar entries. About one-third of allergens were able to be mapped to other allergens. The benefit of allergen consolidation is that by combining the frequencies of reactions that co-occurred with multiple individual allergens that all refer to the same substance, the dynamic picklist can be more accurately ranked for that substance.

It was not surprising that *support* was the best association rule-mining measure for suggesting relevant reactions for allergy reaction documentation. *Support* worked well because it suggested reactions based on allergen-reaction co-occurrence in the development dataset, and the distribution of allergen-reaction pairs in the validation dataset may follow similar patterns. In addition, *lift* and *tfidf*, which tended to raise the associations between allergens and rare reactions, yielded lower recall at top k, perhaps because our validation dataset has an uneven distribution of allergen-reaction pairs with a high frequency of common reactions. However, regardless of the measure used, all 3 dynamic reaction picklists outperformed the static reaction picklist. Based on these results, we recommend the healthcare systems that employ EHR systems that use static picklists consider the adoption of dynamic reaction picklists to facilitate and improve allergy documentation.

It is worth mentioning that in our current EHR allergy module, 3 reaction types (ie, allergy, intolerance, and contraindication) are available for clinicians to choose from when entering a reaction. This field was rarely used by clinicians, however, resulting in most reactions being recorded as the “allergy” reaction type. Contraindications are conditions such as glucose-6-phosphate dehydrogenase (G6PD) deficiency or renal failure in which case certain medications should be avoided. While these contraindications should perhaps be entered elsewhere in the EHR—such as the problem list—in this work, we did not remove contraindications but used the original data for our analysis. Our ongoing effort is to better classify these reaction entries to facilitate further analyses. Once adverse reaction types and contraindications are well-defined and differentiated, we can build additional functions to inform clinicians dynamically of where to document these items.

This study has several limitations worth noting. First, although we used a diverse team of experts and a data-driven approach, some changes (eg, choosing preferred terms, mapping, removal) were subjective and may be biased by the experts’ training, background, and experiences. Second, during the allergen consolidation, we used the

mappings between FDB medications and RxNorm concepts as a source to identify the allergens' ingredients. However, due to differences in the coverage of these medication terminology systems, many allergens could not be identified by their ingredients in RxNorm and thus were left unconsolidated. Allergens with complicated ingredients (eg, vitamins, food) may be less likely to be consolidated due to the challenge of listing all their ingredients. Third, some newer allergens (eg, recently approved drugs) listed in the EHR reference list did not have allergy entries in the allergy database, and such allergens would temporarily need to have static reaction lists. However, dynamic picklists for those allergens could be developed in the future; as we noticed a growth in the number of allergens with allergy entries in the database, for example, there were 743 new allergens with allergy entries in the dataset's final year (October 9, 2018 through October 9, 2019). Finally, this study was conducted in a single healthcare system, and the dynamic reaction picklist was generated in the context of an Epic EHR system and a terminology system for allergens and medications provided by FDB. Although both Epic and FDB are among the top EHR vendors in the US, the portability of the dynamic reaction picklist to other institutions may be impacted due to differences in allergy-reaction frequency distribution across different patient populations as well as in EHR vendors. We therefore suggest that the dynamic picklist be modified by considering these factors prior to adoption.

CONCLUSION

In summary, we identified an enhanced reaction value set of 490 reactions and found that dynamic reaction picklists developed using data-mining association rules and retrospective allergy reaction entries can suggest relevant reactions for improved allergy module reaction entries. Future work includes enhancing reaction type and severity definition and evaluating whether these enhancements improve reaction documentation and/or reduce free-text entries using a prototype user interface that implements the dynamic reaction picklist.

FUNDING

This research was supported with funding from the Agency for Healthcare Research and Quality (AHRQ) grant R01HS025375 and the National Institute of Allergy and Infectious Diseases (NIAID) of National Institute of Health (NIH) grant 1R01AI150295.

AUTHOR CONTRIBUTIONS

LZ obtained funding. LW, FG, and LZ led the conceptualization and design of the work. LW, SY, KGB, FG, YL, SNS, ZTK, and LZ contributed to the refinement of the reaction picklist. SB performed the reaction consolidation. LW conducted the data collection, statistical analysis, and interpretation of the results and drafted the manuscript. SB, SY, KGB, and LZ critically revised the manuscript. All authors provided feedback and take responsibility for the final approval of the version to be published and are accountable for all aspects of the work.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

The authors thank Adam Wright, PhD for valuable advice with this research.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

- Gandhi TK, Weingart SN, Borus J, *et al*. Adverse drug events in ambulatory care. *N Engl J Med* 2003; 348 (16): 1556–64.
- Zhou L, Dhopeswarkar N, Blumenthal KG, *et al*. Drug allergies documented in electronic health records of a large healthcare system. *Allergy* 2016; 71 (9): 1305–13.
- Ariosto D. Factors contributing to CPOE opiate allergy alert overrides. *AMIA Annu Symp Proc* 2014; 2014: 256–65.
- Bryant AD, Fletcher GS, Payne TH. Drug interaction alert override rates in the Meaningful Use era: no evidence of progress. *Appl Clin Inform* 2014; 05 (03): 802–13.
- Isaac T, Weissman JS, Davis RB, *et al*. Overrides of medication alerts in ambulatory care. *Arch Intern Med* 2009; 169 (3): 305–11.
- Jani YH, Barber N, Wong IC. Characteristics of clinical decision support alert overrides in an electronic prescribing system at a tertiary care paediatric hospital. *Int J Pharm Pract* 2011; 19 (5): 363–6.
- McCoy AB, Thomas EJ, Krousel-Wood M, Sittig DF. Clinical decision support alert appropriateness: a review and proposal for improvement. *Ochsner J* 2014; 14 (2): 195–202.
- Slight SP, Beeler PE, Seger DL, *et al*. A cross-sectional observational study of high override rates of drug allergy alerts in inpatient and outpatient settings, and opportunities for improvement. *BMJ Qual Saf* 2017; 26 (3): 217–25.
- Topaz M, Seger DL, Lai K, *et al*. High override rate for opioid drug-allergy interaction alerts: current trends and recommendations for future. *Stud Health Technol Inform* 2015; 216: 242–6.
- Topaz M, Seger DL, Slight SP, Goss F, *et al*. Rising drug allergy alert overrides in electronic health records: an observational retrospective study of a decade of experience. *J Am Med Inform Assoc* 2016; 23 (3): 601–8.
- Blumenthal KG, Wickner PG, Lau JJ, Zhou L. Stevens-Johnson syndrome and toxic epidermal necrolysis: a cross-sectional analysis of patients in an integrated allergy repository of a large health care system. *J Allergy Clin Immunol Pract* 2015; 3 (2): 277–80 e1.
- Goss FR, Lai KH, Topaz M, *et al*. A value set for documenting adverse reactions in electronic health records. *J Am Med Informatics Association* 2018; 25 (6): 661–9.
- Liu S, Ma W, Moore R, Ganesan V, Nelson S. RxNorm: prescription for electronic drug information exchange. *IT Prof* 2005; 7 (5): 17–23.
- Donnelly K. SNOMED-CT: the advanced terminology and coding system for eHealth. *Stud Health Technol Inform* 2006; 121: 279–90.
- Zhou L, Plasek JM, Mahoney LM, *et al*. 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.
- 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; 2014: 580–8.
- Wang L, Del Fiol G, Bray BE, Haug PJ. Generating disease-pertinent treatment vocabularies from MEDLINE citations. *J Biomed Inform* 2017; 65: 46–57.
- Wright A, Chen ES, Maloney FL. An automated technique for identifying associations between medications, laboratory results and problems. *J Biomed Inform* 2010; 43 (6): 891–901.
- Demsar J. Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res* 2006; 7: 1–30.

Supplementary Online Content

eTable 1. The current EHR reaction picklist of 46 reaction concepts used by a large Massachusetts healthcare system.

eTable 2. Refined reaction picklist of 490 reaction concepts with mapping to SNOMED CT concepts

eTable 3. The Top 15 Most Relevant Reactions Ranked by *Support* for Common Drug Allergens

eTable 1. The current EHR reaction picklist of 46 reaction concepts used by a large Massachusetts healthcare system.

Rank	Reaction
1	Acute Generalized Exanthematous Pustulosis
2	Anaphylaxis
3	Anemia
4	Angioedema
5	Anxiety
6	Arrhythmia
7	Bronchospasm
8	Cough
9	Dermatitis
10	Diarrhea
11	Dizziness
12	Dystonia
13	Eosinophilia
14	Erythema Multiforme
15	Fever
16	Fixed Drug Eruption
17	Flushing
18	GI Upset
19	Headaches
20	Hepatotoxicity
21	Hives
22	Hypotension
23	Itching
24	Lightheadedness
25	Maculopapular Rash
26	Mental Status Change
27	Musculoskeletal Pain
28	Myalgia
29	Nausea and/or Vomiting
30	Palpitations
31	Photosensitivity
32	Rash
33	Rash with Joint Pains (Serum Sickness)
34	Rash with Mucosal Lesions (SJS, TENS)
35	Rash with Organ Involvement - Specify in comments
36	Rash with Skin Desquamation
37	Renal Toxicity
38	Rigor
39	Seizures
40	Shortness of Breath

Rank	Reaction
41	Sneezing
42	Swelling
43	Thrombocytopenia
44	Tinnitus
45	Vasculitis
46	Wheezing

eTable 2. Refined reaction picklist of 490 reaction concepts with mapping to SNOMED CT concepts

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
1	Abdominal bloating	Abdominal bloating	116289008
2	Abdominal pain	Abdominal pain	21522001
3	Abnormal breathing	Abnormal breathing	386813002
4	Abnormal gait	Abnormal gait	22325002
5	Abnormal liver function test	Liver function tests abnormal	166603001
6	Acidosis	Acidosis	51387008
7	Acne	Acne	11381005
8	Acute generalized exanthematous pustulosis	Acute generalized exanthematous pustulosis	702617007
9	Acute kidney injury (AKI)	Acute injury of kidney	14350001000004108
10	Aggression	Aggressive behavior	61372001
11	Agitation	Feeling agitated	24199005
12	Agranulocytosis	Agranulocytosis	17182001
13	Airway closure	Airway constriction	44416002
14	Airway swelling	Swelling of respiratory structure	300875005
15	Akathisia	Akathisia	285145004
16	Allergic reaction	Allergic reaction	419076005
17	Allergy test positive	Allergy test positive	165014009
18	Alopecia	Alopecia	56317004
19	Amenorrhea	Amenorrhea	14302001
20	Anaphylaxis	Anaphylaxis	39579001
21	Anemia	Anemia	271737000
22	Anger	Feeling angry	75408008
23	Angioedema	Angioedema	41291007
24	Anxiety	Anxiety	48694002
25	Aplastic anemia	Aplastic anemia	306058006
26	Apnea	Apnea	1023001
27	Arrhythmia	Cardiac arrhythmia	698247007
28	Arthritis	Arthritis	3723001
29	Ataxia	Ataxia	20262006
30	Back pain	Backache	161891005
31	Black hairy tongue	Black hairy tongue	81934005
32	Bleeding	Bleeding	131148009
33	Blepharitis	Blepharitis	41446000
34	Blindness	Blindness AND/OR vision impairment level	105597003
35	Blister	Blister	417237009
36	Blood dyscrasia	Disorder of cellular component of blood	414022008

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
37	Blurred vision	Hazy vision	246636008
38	Boil	Furuncle	416675009
39	Bradycardia	Bradycardia	48867003
40	Breakthrough bleeding	Break-through bleeding	56169002
41	Breast tenderness	Breast tenderness	55222007
42	Bronchospasm	Bronchospasm	4386001
43	Bruising	Contusion	125667009
44	Bruxism	Bruxism (teeth grinding)	191983006
45	Burning sensation	Burning sensation	90673000
46	Burping	Burping	271834000
47	Canker sores	Recurrent aphthous ulcer	398870000
48	Cardiomyopathy	Cardiomyopathy	85898001
49	Cardiorespiratory arrest	Cardiorespiratory arrest	410430005
50	Catatonia	Catatonia	247917007
51	Celiac disease	Celiac disease	396331005
52	Changes in skin texture	Changes in skin texture	274672009
53	Chemical burn	Chemical burn	426284001
54	Chest discomfort	Chest discomfort	279084009
55	Chest pain	Chest pain	29857009
56	Chest tightness	Tight chest	23924001
57	Childhood	Childhood	68433009
58	Chills	Chill	43724002
59	Chloasma	Chloasma	36209000
60	Choking	Choking	249489001
61	Choking sensation	Choking sensation	373909009
62	Cholestasis	Cholestasis	33688009
63	Chronic kidney disease	Chronic kidney disease	709044004
64	Claustrophobia	Claustrophobia	19887002
65	Clostridium difficile infection (C DIFF)	Clostridioides difficile infection	186431008
66	CNS symptom	CNS symptom	246557006
67	Cold extremities	Complaining of cold extremities	161996001
68	Cold sores	Herpes labialis	1475003
69	Cold sweat	Cold sweat	83547004
70	Colitis	Colitis	64226004
71	Coma	Coma	371632003
72	Combativeness	Combativeness	609586001
73	Confusion	Clouded consciousness	40917007
74	Conjunctivitis	Conjunctivitis	9826008

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
75	Constipation	Constipation	14760008
76	Contact dermatitis	Contact dermatitis	40275004
77	Contraction	Contraction	78558002
78	Cough	Cough	49727002
79	Crazy	Fear of going crazy	247822006
80	Crying	Crying	304534000
81	Cyanosis	Cyanosis	3415004
82	Cyanotic attack	Cyanotic attack	301795004
83	Cystitis	Cystitis	38822007
84	Cytopenia	Cytopenia	50820005
85	Deep venous thrombosis (DVT)	Deep venous thrombosis	128053003
86	Dehydration	Dehydration	34095006
87	Delirium	Delirium	2776000
88	Delusions	Delusions	2073000
89	Depression	Depressive disorder	35489007
90	Dermatitis	Inflammatory dermatosis	703938007
91	Diabetes insipidus	Diabetes insipidus	15771004
92	Diarrhea	Diarrhea	62315008
93	Diarrhea and vomiting	Diarrhea and vomiting	249519007
94	Difficulty speaking	Difficulty speaking	47004009
95	Diplopia	Diplopia	24982008
96	Discomfort	Discomfort	247347003
97	Disinhibition	Disinhibited behavior	247973004
98	Disorder of filtering bleb	Disorder of filtering bleb	410724009
99	Disorientation	Disorientated	62476001
100	Dissociation - mental defence mechanism	Dissociation - mental defence mechanism	286669009
101	Dizziness	Dizziness	404640003
102	Dream disorder	Dream disorder	85418005
103	Drooling	Dribbling from mouth	62718007
104	Drug reaction with eosinophilia and systemic symptoms (DRESS)	Drug reaction with eosinophilia and systemic symptoms	702809001
105	Dry eyes	Dry eyes	162290004
106	Dry mouth	Xerostomia	87715008
107	Dry skin	Dry skin	16386004
108	Dry throat	Pharyngeal dryness	102618009
109	Dysesthesia	Dysesthesia	279079003
110	Dyskinesia	Dyskinesia	9748009
111	Dyspepsia	Indigestion	162031009

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
112	Dysphagia	Dysphagia	40739000
113	Dysphoria	Dysphoric mood	30819006
114	Dystonia	Dystonia	15802004
115	Dysuria	Dysuria	49650001
116	Easy bruising	Easy bruising	424131007
117	Ecchymosis	Ecchymosis	302227002
118	Edema	Edema	267038008
119	Electrolyte imbalance	Electrolyte imbalance	105593004
120	Elevated creatinine	Serum creatinine raised	166717003
121	Elevated liver function test	Elevated liver enzymes level	707724006
122	Encephalopathy	Disorder of brain	81308009
123	Eosinophilia	Eosinophil count raised	386789004
124	Eosinophilic esophagitis	Eosinophilic esophagitis	235599003
125	Epistaxis	Epistaxis	249366005
126	Erythema	Erythema	247441003
127	Erythema multiforme	Erythema multiforme	36715001
128	Erythema nodosum	Erythema nodosum	32861005
129	Erythroderma	Erythroderma	399992009
130	Esophageal spasm	Diffuse spasm of esophagus	79962008
131	Esophageal ulcer	Ulcer of esophagus	30811009
132	Esophagitis	Esophagitis	16761005
133	Euphoria	Euphoria	85949006
134	Excessive somnolence	Excessive somnolence	372947007
135	Excitability	Excitability	102893002
136	Extrapyramidal sign (EPS)	Extrapyramidal sign	43378000
137	Eye inflammation	Scleritis	78370002
138	Eye irritation	Sensation of irritation of eye proper	705083008
139	Eye pain	Pain in eye	41652007
140	Eye swelling	Swelling of structure of eye	45177002
141	Facial edema	Edema of face	445088006
142	Facial numbness	Numbness of face	309557009
143	Facial swelling	Facial swelling	278528006
144	Falls	Falls	161898004
145	Fatigue	Fatigue	84229001
146	Fecal incontinence	Incontinence of feces	72042002
147	Feeling intoxicated	Feeling intoxicated	102897001
148	Fever	Fever	386661006
149	Fixed drug eruption	Fixed drug eruption	73692007

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
150	Fluid retention	Body fluid retention	43498006
151	Flushing	Flushing	238810007
152	Follicular conjunctivitis	Follicular conjunctivitis	86402005
153	Gagging	Gagging	249488009
154	Galactorrhea	Galactorrhea not associated with childbirth	78622004
155	Gas	Passing flatus	249504006
156	Gastritis	Gastritis	4556007
157	Gastroesophageal reflux disease (GERD)	Gastroesophageal reflux disease	235595009
158	Gastrointestinal hemorrhage	Gastrointestinal hemorrhage	74474003
159	Gastroparesis	Gastroparesis syndrome	235675006
160	Generalised aches and pains	Generalised aches and pains	82991003
161	GI Upset	Gastrointestinal irritation	95516005
162	Gingival hyperplasia	Gingival enlargement	54711002
163	Gingivitis	Gingivitis	66383009
164	Glaucoma	Glaucoma	23986001
165	Glossitis	Glossitis	45534005
166	Glucose-6-phosphate dehydrogenase deficiency anemia (G6PD)	Glucose-6-phosphate dehydrogenase deficiency anemia	62403005
167	Gout	Gout	90560007
168	Granulocytopenia	Granulocytopenic disorder	417672002
169	Guillain-Barre syndrome	Guillain-Barré syndrome	40956001
170	Gum hyperplasia	Hyperplasia of gingiva	441798003
171	Gynecomastia	Gynecomastia	4754008
172	Hair loss	Loss of hair	278040002
173	Hallucinations	Hallucinations	7011001
174	Hangover	Hangover	32553006
175	Headaches	Headache	25064002
176	Hearing loss	Hearing loss	15188001
177	Heart block	Heart block	233916004
178	Hematuria	Blood in urine	34436003
179	Heme positive stools	Occult blood in stools	59614000
180	Hemolysis	Hemolysis	73320003
181	Hemolytic anemia	Hemolytic anemia	61261009
182	Hemophilia	Hemophilia	90935002
183	Hemoptysis	Hemoptysis	66857006
184	Heparin-induced thrombocytopenia	Heparin-induced thrombocytopenia	73397007
185	Hepatitis	Inflammatory disease of liver	128241005

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
186	Hiccups	Hiccoughs	65958008
187	Hives	Weal	247472004
188	Hoarseness	Hoarse	50219008
189	Hot flashes	Menopausal flushing	198436008
190	Hyperactivity	Hyperactive behavior	44548000
191	Hypercalcemia	Hypercalcemia	66931009
192	Hyperemesis	Hyperemesis	444673007
193	Hyperglycemia	Hyperglycemia	80394007
194	Hyperkalemia	Hyperkalemia	14140009
195	Hyperlipidemia	Hyperlipidemia	55822004
196	Hyperpigmentation	Hyperpigmentation of skin	49765009
197	Hyperprolactinemia	Hyperprolactinemia	237662005
198	Hypersensitivity	Sensitivity	35209006
199	Hypersensitivity reaction	Hypersensitivity reaction	421961002
200	Hypersomnia	Hypersomnia	77692006
201	Hypertension	Hypertensive disorder	38341003
202	Hypertensive crisis	Hypertensive crisis	706882009
203	Hyperthermia	Body temperature above reference range	50177009
204	Hyperthyroidism	Hyperthyroidism	34486009
205	Hypertriglyceridemia	Hypertriglyceridemia	302870006
206	Hyperuricemia	Hyperuricemia	35885006
207	Hyperventilation	Hyperventilation	68978004
208	Hypoglycemia	Hypoglycemia	302866003
209	Hypokalemia	Hypokalemia	43339004
210	Hypomania	Hypomania	231496004
211	Hyponatremia	Hyponatremia	89627008
212	Hypotension	Low blood pressure	45007003
213	Hypothermia	Hypothermia	386689009
214	Hypothyroidism	Hypothyroidism	40930008
215	Hypoxia	Hypoxia	389086002
216	Idiopathic thrombocytopenic purpura (ITP)	Idiopathic thrombocytopenic purpura	32273002
217	Ileus	Intestinal obstruction co-occurrent and due to decreased peristalsis	710572000
218	Illness	Illness	39104002
219	Imbalance	Feels as though will fall	249983008
220	Incontinence	Incontinence	48340000
221	Increased appetite	Increased appetite	72405004
222	Increased creatine kinase level	Increased creatine kinase level	432352001

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
	(ELEVATED CPK)		
223	Inflammatory disease of liver	Inflammatory disease of liver	128241005
224	Influenza-like illness	Influenza-like illness	95891005
225	Infusion reaction	Adverse reaction to drug or medicament administered by infusion	441940003
226	Injection site reaction	Injection site reaction	95377006
227	Insomnia	Insomnia	193462001
228	Interstitial lung disease	Interstitial lung disease	233703007
229	Interstitial nephritis	Interstitial nephritis	28689008
230	Involuntary movement	Involuntary movement	267078001
231	Iritis	Iritis	65074000
232	Irritable bowel syndrome	Irritable bowel syndrome	10743008
233	Irritation	Feeling irritable	55929007
234	Ischemic colitis	Ischemic colitis	30588004
235	Itching	Itching	418290006
236	Itchy eyes	Itching of eye	74776002
237	Itchy scalp	Scalp itchy	275921007
238	Itchy throat	Throat irritation	162400007
239	Jaundice	Jaundice	18165001
240	Jaw pain	Jaw pain	274667000
241	Joint pain	Joint pain	57676002
242	Joint stiffness	Joint stiffness	84445001
243	Joint swelling	Joint swelling	271771009
244	Keratitis	Keratitis	5888003
245	Ketogenic diet	Ketogenic diet	765060000
246	Kidney stone	Kidney stone	95570007
247	Lactation problem	Lactation problem	58219009
248	Lactose intolerant	Intolerance to lactose	782415009
249	Laryngeal edema	Edema of larynx	51599000
250	Laryngitis	Laryngitis	45913009
251	Laryngospasm	Laryngeal spasm	406444002
252	Lesion of lip	Lesion of lip	301350008
253	Leukocytoclastic vasculitis	Hypersensitivity angiitis	60555002
254	Leukopenia	Leukopenia	84828003
255	Libido disorder	Reduced libido	8357008
256	Lichen planus	Lichen planus	4776004
257	Lip swelling	Lip swelling	699376002
258	Liver failure	Hepatic failure	59927004

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
259	Loss of appetite	Loss of appetite	79890006
260	Loss of consciousness (LOC)	Loss of consciousness	419045004
261	Loss of voice	Loss of voice	44564008
262	Lower urinary tract symptoms (LUTS)	Lower urinary tract symptoms	307541003
263	Lowers seizure threshold	Lowered convulsive threshold	19260006
264	Lump in throat	Feeling of lump in throat	267103008
265	Lupus	Lupus erythematosus	200936003
266	Lymphadenopathy	Lymphadenopathy	30746006
267	Maculopapular rash	Maculopapular eruption	247471006
268	Malaise	Malaise	367391008
269	Malignant hyperthermia	Malignant hyperthermia	405501007
270	Mania	Mania	231494001
271	Memory impairment	Memory impairment	386807006
272	Meningitis	Meningitis	7180009
273	Menorrhagia	Menorrhagia	386692008
274	Menstrual spotting	Menstrual spotting	9126005
275	Mental status change	Altered mental status	419284004
276	Methemoglobinemia	Methemoglobinemia	38959009
277	Migraine	Migraine	37796009
278	Mood swings	Mood swings	18963009
279	Morbilliform rash	Morbilliform eruption	247470007
280	Mouth numbness	Perioral numbness	102601005
281	Mouth swelling	Swelling of oral cavity structure	300878007
282	Movement disorder	Movement disorder	60342002
283	Mucositis	Inflammatory disease of mucous membrane	95361005
284	Multiple symptoms	Multiple symptoms	38276004
285	Muscle cramps	Cramp	55300003
286	Muscle fatigue	Muscle fatigue	80449002
287	Muscle spasms	Smooth muscle spasm	128526009
288	Muscle stiffness	Muscle rigidity	16046003
289	Muscle twitching	Muscle twitch	60238002
290	Muscle weakness	Muscle weakness	26544005
291	Musculoskeletal pain	Musculoskeletal pain	279069000
292	Myalgia	Muscle pain	68962001
293	Myasthenia gravis	Myasthenia gravis	91637004
294	Myelosuppression	Myelosuppression	370569000
295	Myocardial infarction	Myocardial infarction	22298006
296	Myoclonus	Myoclonus	17450006

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
297	Myopathy	Disorder of muscle	129565002
298	Myositis	Myositis	26889001
299	Nasal congestion	Nasal congestion	68235000
300	Nasal dryness	Nasal mucosa dry	42201002
301	Nasal polyp	Polyp of nasal sinus	32307003
302	Nasal symptoms	Nasal symptom	249307003
303	Nausea , vomiting , diarrhea	Nausea, vomiting and diarrhea	2919008
304	Nausea and/or vomiting	Nausea and vomiting	16932000
305	Near syncope	Near syncope	427461000
306	Neck pain	Neck pain	81680005
307	Neck spasms	Muscle spasm of cervical muscle of neck	301000119104
308	Neck swelling	Neck swelling	301777002
309	Nephritis	Nephritis	52845002
310	Nephrotic syndrome	Nephrotic syndrome	52254009
311	Neuroleptic malignant syndrome	Neuroleptic malignant syndrome	15244003
312	Neurotoxicity	Neurotoxicity	19466003
313	Neutropenia	Neutropenia	165517008
314	Night sweats	Night sweats	42984000
315	Nocturia	Nocturia	139394000
316	Normal behavior	Normal behavior	51746009
317	Numbness	Numbness	44077006
318	Ocular hypertension	Ocular hypertension	4210003
319	Oculogyric crisis	Oculogyric crisis	5332004
320	Oliguria	Oliguria	83128009
321	Opsoclonus	Opsoclonia	194177006
322	Optic neuritis	Optic neuritis	66760008
323	Oral lesion	Oral lesion	1071000119107
324	Oral thrush	Candidiasis of mouth	79740000
325	Oral tingling	Numbness of tongue	95528002
326	Oropharyngeal swelling	Edema of pharynx	2129002
327	Orthostatic hypotension	Orthostatic hypotension	28651003
328	Osteonecrosis	Bone necrosis	240196003
329	Ototoxicity	Ototoxicity	9062008
330	Out of body experience	Out of body experience	307050007
331	Outbursts of anger	Outbursts of anger	274951009
332	Pain	Pain	22253000
333	Palpitations	Palpitations	80313002
334	Pancreatitis	Pancreatitis	75694006

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
335	Pancytopenia	Pancytopenia	127034005
336	Panic	Panic	79823003
337	Panic attack	Panic attack	225624000
338	Paralysis	Paralysis	44695005
339	Paranoia	Paranoid disorder	191667009
340	Paresthesia	Paresthesia	91019004
341	Parkinsonism	Parkinsonism	32798002
342	Paroxysmal nocturnal dyspnea	Paroxysmal nocturnal dyspnea	55442000
343	Peliosis hepatis	Peliosis hepatis	58008004
344	Pelvic pain	Pain in pelvis	30473006
345	Penile swelling	Penile swelling	335977000
346	Peptic ulcer	Peptic ulcer	13200003
347	Pericarditis	Pericarditis	3238004
348	Peripheral edema	Peripheral edema	271809000
349	Personality change	Personality change	102943000
350	Petechiae	Petechiae	271813007
351	Phlebitis	Phlebitis	61599003
352	Photosensitivity	Photosensitivity	90128006
353	Pneumonia	Pneumonia	233604007
354	Pneumonitis	Pneumonitis	205237003
355	Pollen-food allergy	Pollen-food allergy	432807008
356	Polyuria	Polyuria	28442001
357	Porphyria	Porphyria	418470004
358	Positive ana	Anti-nuclear factor positive	165850001
359	Positive skin test reaction	Positive skin test reaction	26576000
360	Priapism	Priapism	6273006
361	Proteinuria	Proteinuria	29738008
362	Pruritic rash	Pruritic rash	64144002
363	Pseudocholinesterase deficiency	Pseudocholinesterase deficiency	191397007
364	Pseudomembranous colitis	Clostridium difficile colitis	423590009
365	Pseudoporphyria	Pseudoporphyria	95565000
366	Pseudotumor cerebri	Benign intracranial hypertension	68267002
367	Psoriasis	Psoriasis	9014002
368	Psychosis	Psychotic disorder	69322001
369	Pulmonary congestion	Pulmonary congestion	67599009
370	Pulmonary edema	Pulmonary edema	19242006
371	Pulmonary embolism	Pulmonary embolism	59282003
372	Pulmonary fibrosis	Fibrosis of lung	51615001
373	Pulmonary toxicity	Toxic pneumonitis	233733000

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
374	Purpura	Purpura	423902002
375	Pustule	Pustule	271760008
376	Qt prolongation	Prolonged QT interval	111975006
377	Racing thoughts	Racing thoughts	285303006
378	Rapid pulse	Pulse fast	86651002
379	Rash	Eruption	271807003
380	Rash with Joint Pains (Serum Sickness)	Transfusion reaction due to serum protein reaction	72284000
381	Rash with mucosal lesions (sjs, tens)		
382	Rash with Organ Involvement - Specify in comments		
383	Rash with Skin Desquamation		
384	Raynaud's	Raynaud's disease	195295006
385	Red eye	Red eye	703630003
386	Red man syndrome	Drug-induced erythroderma	238992001
387	Renal failure	Renal failure syndrome	42399005
388	Renal toxicity	Toxic nephropathy	236514003
389	Respiratory depression	Decreased respiratory function	80954004
390	Respiratory syndrome	Respiratory symptom	161920001
391	Restless legs	Restless legs	32914008
392	Retinopathy	Retinopathy	29555009
393	Retrograde ejaculation	Retrograde ejaculation	50112006
394	Rhabdomyolysis	Rhabdomyolysis	240131006
395	Rhinitis	Rhinitis	70076002
396	Rigor	Rigor	38880002
397	Rosacea	Rosacea	398909004
398	Seizures	Seizure	91175000
399	Serotonin syndrome	Serotonin syndrome	371089000
400	Severe depression	Severe depression	310497006
401	Severe skin reaction	Skin reaction strong	251351004
402	Sexual dysfunction	Abnormal sexual function	56925008
403	Shortness of breath	Dyspnea	267036007
404	Sinus congestion	Congestion of nasal sinus	82297005
405	Sinus headache	Sinus headache	4969004
406	Sinusitis	Sinusitis	36971009
407	Skin breakdown	Broken skin	247442005
408	Skin discoloration	Discoloration of skin	3253007
409	Skin irritation	Skin irritation	367466007
410	Skin lesion	Skin lesion	95324001
411	Skin necrosis	Skin necrosis	95347000

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
412	Sleep disorder	Sleep disorder	39898005
413	Slurred speech	Slurred speech	289195008
414	Sneezing	Sneezing	76067001
415	Somnolence	Drowsy	271782001
416	Sore throat	Pain in throat	162397003
417	Sore tongue	Glossodynia	30731004
418	Spasm	Spasm	45352006
419	Stiff neck	Stiff neck	161882006
420	Stiffness	Stiffness	271587009
421	Stomach cramps	Stomach cramps	51197009
422	Stomach problem	Stomach problem	300306001
423	Stomatitis	Stomatitis	61170000
424	Stridor	Stridor	70407001
425	Stroke	Cerebrovascular accident	230690007
426	Stuttering	Stammering	39423001
427	Subdural hematoma	Subdural hematoma	95453001
428	Suicidal	Suicidal	267073005
429	Sweating	Sweating	415691001
430	Swelling	Swelling	65124004
431	Swelling of extremities	Swelling of limb	80068009
432	Syncope	Syncope	271594007
433	Syndrome of inappropriate vasopressin secretion (SIADH)	Syndrome of inappropriate vasopressin secretion	55004003
434	Systemic lupus erythematosus (SLE)	Systemic lupus erythematosus	55464009
435	Tardive dyskinesia	Tardive dyskinesia	102449007
436	Taste sense altered	Taste sense altered	271801002
437	Tearing	Epiphora	193982009
438	Tendinopathy	Disorder of tendon	68172002
439	Tendonitis	Tendinitis	34840004
440	Tetanus	Tetanus	76902006
441	Tetany	Tetany	10629009
442	Throat burning	Burning sensation of mucous membrane	95360006
443	Throat swelling	Pharyngeal swelling	421581006
444	Throat symptom	Throat symptom	395496008
445	Throat tightness	Feeling of throat tightness	7899002
446	Thrombocytopenia	Thrombocytopenic disorder	302215000
447	Thrombophlebitis	Thrombophlebitis	64156001
448	Thrombosis	Thrombosis	439127006

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
449	Thrombotic thrombocytopenic purpura (TTP)	Thrombotic thrombocytopenic purpura	78129009
450	Tics	Tic	386783003
451	Tinnitus	Tinnitus	60862001
452	Tongue burning	Glossopyrosis	399044006
453	Tongue swelling	Tongue swelling	421262002
454	Torsades de pointes	Torsades de pointes	31722008
455	Toxicity	Poisoning	75478009
456	Transient cerebral ischemia (TIA)	Transient cerebral ischemia	266257000
457	Tremor	Tremor	26079004
458	Trismus	Trismus	87866006
459	Tunnel vision	Tunnel visual field constriction	420990001
460	Twitching	Spasmodic movement	249966004
461	Ulcer	Ulcer	429040005
462	Ulcer of mouth	Ulcer of mouth	26284000
463	Ulcerative colitis	Ulcerative colitis	64766004
464	Undifferentiated illness: Vague ill health	Undifferentiated illness: Vague ill health	248282002
465	Unusual change in behavior	Unusual change in behavior	427797008
466	Upper respiratory infection (URI)	Upper respiratory infection	54150009
467	Urinary frequency	Increased frequency of urination	162116003
468	Urinary incontinence	Urinary incontinence	165232002
469	Urinary retention	Retention of urine	267064002
470	Uveitis	Uveitis	128473001
471	Uvular swelling	Edema of uvula	232411006
472	Vaginal bleeding	Vaginal bleeding	289530006
473	Vaginal dryness	Vaginal dryness	31908003
474	Vaginal rash	Eruption of vulva	289482006
475	Vaginitis	Vaginitis	30800001
476	Vasculitis	Vasculitis	31996006
477	Vasospasm	Vasospasm	71772004
478	Vasovagal syncope	Vasovagal syncope	398665005
479	Ventricular tachycardia (VT)	Ventricular tachycardia	25569003
480	Vertigo	Vertigo	399153001
481	Violent	Physical aggression	248004009
482	Visual changes	Sight deteriorating	225581002
483	Visual hallucinations	Visual hallucinations	64269007
484	von Willebrand disorder	von Willebrand disorder	128105004

ID	Reaction Picklist	SNOMED CT Preferred Term	SNOMED CT Concept ID
485	Walking disability	Walking disability	228158008
486	Watery eye	Watery eye	420103007
487	Weakness	Asthenia	13791008
488	Weight gain	Weight gain	8943002
489	Weight loss	Weight loss	89362005
490	Wheezing	Wheezing	56018004

eTable 3. The Top 15 Most Relevant Reactions Ranked by *Support* for Common Drug Allergens

Drug class and allergen	Allergen prevalence Total = 1,683,245 [†] n (%)	Top 15 most relevant reaction ranked by <i>support</i>
Antibiotics*		
Penicillins[‡]	235 092 (14.0)	Rash, hives, anaphylaxis, swelling, itching, GI upset, angioedema, N/V, SOB, diarrhea, bronchospasm, fever, wheezing, facial swelling, throat swelling
Amoxicillin	60 841 (3.6)	Rash, hives, GI upset, N/V, itching, diarrhea, swelling, anaphylaxis, SOB, angioedema, vaginitis, fever, maculopapular rash, erythema, facial swelling
Amoxicillin-Pot Clavulanate	19 327 (1.1)	Rash, GI upset, hives, diarrhea, N/V, itching, swelling, anaphylaxis, SOB, angioedema, vaginitis, dizziness, headaches, abnormal liver function test, fever
Sulfonamide Antibiotic	139 232 (8.3)	Rash, hives, GI upset, itching, N/V, swelling, anaphylaxis, fever, angioedema, SOB, headaches, bronchospasm, diarrhea, flushing, erythema
Sulfamethoxazole- Trimethoprim	36 200 (2.2)	Rash, hives, N/V, GI upset, itching, swelling, fever, anaphylaxis, angioedema, SOB, diarrhea, headaches, dizziness, renal toxicity, flushing
Macrolide Antibiotics	27 023 (1.6)	GI upset, rash, hives, N/V, diarrhea, itching, anaphylaxis, swelling, SOB, angioedema, abdominal pain, bronchospasm, headaches, palpitations, dizziness
Erythromycin	22 345 (1.3)	GI upset, rash, N/V, hives diarrhea, itching, anaphylaxis, swelling, SOB, angioedema, abdominal pain, bronchospasm, headaches, palpitations, dizziness
Azithromycin	15 799 (0.9)	Rash, hives, GI upset, N/V, diarrhea, itching, swelling, anaphylaxis, angioedema, SOB, palpitations, dizziness, abdominal pain, headaches, facial swelling
Cephalosporins	18 113 (1.1)	Rash, hives, anaphylaxis, itching, GI upset, swelling, angioedema, N/V, SOB, diarrhea, fever, bronchospasm, wheezing, rash with joint pains (serum sickness), flushing
Cephalexin	18 134 (1.1)	Rash, hives, GI upset, itching, N/V, diarrhea, swelling, anaphylaxis, SOB, angioedema, dizziness, headaches, fever, vaginitis, erythema
Fluoroquinolones*		
Ciprofloxacin	22 819 (1.4)	Rash, hives, GI upset, N/V, itching, swelling, anaphylaxis, diarrhea, tendonitis, dizziness, musculoskeletal pain, angioedema, SOB, mental status change, headaches
Levofloxacin	15 124 (0.9)	Rash, hives, GI upset, itching, tendonitis, N/V, anaphylaxis, swelling, musculoskeletal pain, mental status change,

Drug class and allergen	Allergen prevalence Total = 1,683,245 [†] n (%)	Top 15 most relevant reaction ranked by <i>support</i>
		myalgia, diarrhea, dizziness, SOB, angioedema
Tetracyclines	5645 (0.3)	Rash, hives, GI upset, N/V, itching, anaphylaxis, swelling, SOB, diarrhea, angioedema, headaches, fever, dizziness, vaginitis, mental status change
Doxycycline	14 943 (0.9)	Rash, GI upset, N/V, hives, itching, diarrhea, swelling, anaphylaxis, headaches, angioedema, photosensitivity, dizziness, SOB, esophagitis, abdominal pain
Tetracycline	7881 (0.5)	Rash, hives, GI upset, N/V, itching, swelling, anaphylaxis, angioedema, SOB, diarrhea, vaginitis, headaches, dizziness, fever, photosensitivity
Other Antibiotics*		
Clindamycin	12 855 (0.8)	Rash, hives, GI upset, diarrhea, itching, N/V, anaphylaxis, swelling, C Diff, SOB, angioedema, colitis, flushing, fever, erythema
Vancomycin	9631 (0.6)	Rash, red man syndrome, hives, itching, anaphylaxis, flushing, erythema, fever, swelling, SOB, renal toxicity, angioedema, N/V, thrombocytopenia, GI upset
Opiates*		
Codeine	75 921 (4.5)	N/V, GI upset, rash, hives, itching, mental status change, headaches, dizziness, anaphylaxis, SOB, swelling, hallucinations, syncope, palpitations, anxiety
Morphine	39 097 (2.3)	N/V, itching, GI upset, rash, hives, mental status change, hallucinations, anaphylaxis, SOB, swelling, hypotension, headaches, dizziness, confusion, anxiety
Oxycodone-Acetaminophen	30 155 (1.8)	N/V, GI upset, itching, rash, hives, mental status change, hallucinations, dizziness, headaches, SOB, anxiety, anaphylaxis, swelling, palpitations, syncope
Oxycodone	22 410 (1.3)	N/V, itching, GI upset, rash, mental status change, hives, hallucinations, dizziness, Headaches, SOB, anxiety, confusion, anaphylaxis, swelling, dream disorder
Meperidine	15 976 (0.9)	N/V, GI upset, hives, rash, mental status change, itching, hypotension, anaphylaxis, hallucinations, SOB, swelling, headaches, dizziness, seizures, syncope
Hydrocodone-acetaminophen	14 571 (0.9)	N/V, GI upset, itching, rash, hives, mental status change, headaches, dizziness, hallucinations, SOB, swelling, anaphylaxis, palpitations, syncope, anxiety
NSAIDS	21 665 (1.3)	GI upset, hives, rash, anaphylaxis, angioedema, gastrointestinal hemorrhage, swelling, renal toxicity,

Drug class and allergen	Allergen prevalence Total = 1,683,245 [†] n (%)	Top 15 most relevant reaction ranked by <i>support</i>
		bronchospasm, bleeding, N/V, itching, SOB, wheezing, peptic ulcer
Aspirin	37 619 (2.2)	GI upset, hives, rash, anaphylaxis, bleeding, swelling, N/V, angioedema, bronchospasm, gastrointestinal hemorrhage, SOB, itching, wheezing, epistaxis, bruising
Ibuprofen	20 278 (1.2)	GI upset, hives, rash, swelling, N/V, anaphylaxis, itching, angioedema, SOB, gastrointestinal hemorrhage, bronchospasm, facial swelling, bleeding, wheezing, lip swelling
Naproxen	11 040 (0.7)	GI upset, hives, rash, N/V, swelling, itching, anaphylaxis, angioedema, SOB, dizziness, gastrointestinal hemorrhage, palpitations, mental status change, diarrhea, headaches
ACE Inhibitors	22 001 (1.3)	Cough, angioedema, hyperkalemia, rash, bronchospasm, wheezing, renal toxicity, swelling, hives, anaphylaxis, hypotension, itching, GI upset, renal failure, SOB
Lisinopril	45 173 (2.7)	Cough, angioedema, rash, swelling, hyperkalemia, dizziness, hives, anaphylaxis, bronchospasm, itching, GI upset, wheezing, headaches, N/V, hypotension
Statins-HMG-CoA Reductase Inhibitors	8919 (0.5)	Myalgia, musculoskeletal pain, rash, GI upset, abnormal liver function test, pain, elevated liver function test, muscle cramps, weakness, muscle weakness, elevated CPK, hives, joint pain, N/V, myositis
Atorvastatin	17 299 (1.0)	Myalgia, musculoskeletal pain, rash, GI upset, elevated liver function test, abnormal liver function test, pain, muscle cramps, joint pain, elevated CPK, N/V, headaches, weakness, fatigue, diarrhea
Simvastatin	11 972 (0.7)	Myalgia, musculoskeletal pain, GI upset, pain, N/V, elevated CPK, elevated liver function test, muscle cramps, headaches, abnormal liver function test, itching, joint pain, fatigue, dizziness

SOB = Shortness of breath; N/V = Nausea and/or vomiting; Elevated CPK = increased creatine kinase level; C Diff = Clostridium difficile infection (C Diff)

[†] Total number of patients in the development dataset.

* Allergen group names, if not available in the database, were provided by the authors.

[‡] The count for drug class allergen was calculated based on the its own occurrence in the database rather by aggregating the counts of the child-level allergens

