HL7_XPARADIGM_ALLERGY_VS_R1_I2_2018JAN

HL7 Cross Paradigm Specification: Allergy and Intolerance Substance Value Set(s) Definition Release 1

January 2018

HL7 Informative Ballot

Sponsored by:
Patient Care Work Group
Vocabulary Work Group

IMPORTANT NOTES:

HL7 licenses its standards and select IP free of charge. **If you did not acquire a free license from HL7 for this document,** you are not authorized to access or make any use of it. To obtain a free license, please visit http://www.HL7.org/implement/standards/index.cfm.

If you are the individual that obtained the license for this HL7 Standard, specification or other freely licensed work (in each and every instance "Specified Material"), the following describes the permitted uses of the Material.

A. HL7 INDIVIDUAL, STUDENT AND HEALTH PROFESSIONAL MEMBERS, who register and agree to the terms of HL7's license, are authorized, without additional charge, to read, and to use Specified Material to develop and sell products and services that implement, but do not directly incorporate, the Specified Material in whole or in part without paying license fees to HL7.

INDIVIDUAL, STUDENT AND HEALTH PROFESSIONAL MEMBERS wishing to incorporate additional items of Special Material in whole or part, into products and services, or to enjoy additional authorizations granted to HL7 ORGANIZATIONAL MEMBERS as noted below, must become ORGANIZATIONAL MEMBERS of HL7.

B. HL7 ORGANIZATION MEMBERS, who register and agree to the terms of HL7's License, are authorized, without additional charge, on a perpetual (except as provided for in the full license terms governing the Material), non-exclusive and worldwide basis, the right to (a) download, copy (for internal purposes only) and share this Material with your employees and consultants for study purposes, and (b) utilize the Material for the purpose of developing, making, having made, using, marketing, importing, offering to sell or license, and selling or licensing, and to otherwise distribute, Compliant Products, in all cases subject to the conditions set forth in this Agreement and any relevant patent and other intellectual property rights of third parties (which may include members of HL7). No other license, sublicense, or other rights of any kind are granted under this Agreement.

C. NON-MEMBERS, who register and agree to the terms of HL7's IP policy for Specified Material, are authorized, without additional charge, to read and use the Specified Material for evaluating whether to implement, or in implementing, the Specified Material, and to use Specified Material to develop and sell products and services that implement, but do not directly incorporate, the Specified Material in whole or in part.

NON-MEMBERS wishing to incorporate additional items of Specified Material in whole or part, into products and services, or to enjoy the additional authorizations granted to HL7 ORGANIZATIONAL MEMBERS, as noted above, must become ORGANIZATIONAL MEMBERS of HL7.

Please see http://www.HL7.org/legal/ippolicy.cfm for the full license terms governing the Material.

Ownership. Licensee agrees and acknowledges that **HL7 owns** all right, title, and interest, in and to the Materials. Licensee shall **take no action contrary to, or inconsistent with**, the foregoing.

Licensee agrees and acknowledges that HL7 may not own all right, title, and interest, in and to the Materials and that the Materials may contain and/or reference intellectual property owned by third parties ("Third Party IP"). Acceptance of these License Terms does not grant Licensee any rights with respect to Third Party IP. Licensee alone is responsible for identifying and obtaining any necessary licenses or authorizations to utilize Third Party IP in connection with the Materials or otherwise. Any actions, claims or suits brought by a third party resulting from a breach of any Third Party IP right by the Licensee remains the Licensee's liability.

Following is a non-exhaustive list of third-party terminologies that may require a separate license:

Terminology	Owner/Contact
Current Procedures Terminology (CPT) code set	American Medical Association http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt/cpt-products-services/licensing.page?
SNOMED CT	SNOMED International https://www.snomed.org/
Logical Observation Identifiers Names & Codes (LOINC)	Regenstrief Institute
International Classification of Diseases (ICD) codes	World Health Organization (WHO)
NUCC Health Care Provider Taxonomy code set	American Medical Association. Please see 222.nucc.org. AMA licensing contact: 312-464-5022 (AMA IP services)

Acknowledgments

This material contains content from SNOMED CT® (http://www.ihtsdo.org/snomed-ct/). SNOMED CT is a registered trademark of the International Health Terminology Standard Development Organization (IHTSDO).

Project Co-Chairs

- Elaine Ayres, National Institutes of Health Clinical Center, NIH
- Jay Lyle, JP Systems

Patient Care WG Co-Chairs

- Stephen Chu
- Laura Heermann-Langford
- Emma Jones
- Jay Lyle
- Michelle Miller
- Michael Tan

Subject Matter Experts

- Russell Leftwich (Immunology)
- Robert Hausam (Vocabulary)

The Patient Care WG also is grateful to the organizations that provided invaluable data sets for this work:

- Cerner Population Health – Larry McKnight and Michelle Miller
- Veterans Administration – Catherine Hoang
- Department of Defense (CHCS)– David Parker
- Kaiser Permanente – Rita Barsoum
- Intermountain Health – Susan Matney
- University of Nebraska – Jim McClay
- Cleveland Clinic – Sue Kent
- NIH Clinical Center – Elaine Ayres

The following individuals provided assistance with data annotation and analysis:

Olivier Bodenreider – NIH National Library of Medicine	David Parker - Defined IT, Inc
Della Dunbar – DM&A	Donna Quirk – Lexington Health
Jennifer Harward – US Air Force	Kate Russell – NIH Clinical Center
Clare Hicks – CBORD, Inc.	Sharon Solomon – Morrisons
Lindsey Hoggle – Academy of Nutrition and Dietetics	Jim Coates – Wolters Kluwer
Sue Kent – Cleveland Clinic	Larry McKnight – Cerner

Allergy and Intolerance Substance Value Set(s) Definition

Table of Contents

Allergy and Intolerance Substance Value Set(s) Definition 4
1. Problem Statement 5
2. Goal of this Project 5
3. History and Context 7
4. Considerations 9
a. This is not pharmacovigilance 9
b. Class definitions 9
c. Mixtures 10
d. Terminology system selection 10
e. Criticality 12
f. Length 14
g. Vaccines 14
h. List crossover 15
5. Approach 15
a. Collection 15
b. Analysis of Medication, Medication Classes, Vaccine and Biologics . 15
e. Analysis of Negation Terms 17
f. List Output 17
6. Issues 17
a. Source data discrepancies 17
b. Discrepancies between best practice and reality 18
7. Quality assurance for medication data 18
8. Findings 19
a. Medications 19
b. Food 19
c. Environmentals 20
9. Guidance 20
a. Seafood. 20
b. Iodine. 20
c. Penicillin. 21
d. Vaccines 21
e. Oils 22
f. Environmental allergens 22
g. Nickel 22
Appendix A: Record Counts by Data Source and Type 23
Appendix B: International Labeling Regulations for Food Allergens 24

1. Problem Statement

Different clinical systems use different assumptions when recording patient allergies and intolerances. Because it is conceivable that a patient may be sensitive to virtually any substance, the assumptions tend to support the capture of a wide variety of substances, and they do so in a variety of ways. The Consolidated CDA (R2.1)[1] specification, for instance, specifies a rule[2] whereby substances can be recorded as any substance identified by SNOMED CT, UNII, NDF-RT, or RxNorm. The sheer number of concepts involved means:

1. It may be difficult to find a concept that is appropriate, making it likely that an approximation will be used,
2. The use of such approximations may mean that the same condition may be recorded multiple times, in terms that are difficult to reconcile,
3. The use of multiple terminology systems introduces synonymy, making redundant and possibly confusing records likely, and
4. Automated systems may choose to adopt their own shorter, more tractable lists, making interoperability more challenging.

2. Goal of this Project

Two approaches seem to promise better results: better semantic engineering of the list, and results-driven heuristics.

The first is to fix or constrain the specified code systems to reduce the number of concepts to a more manageable level. This is the approach of the US Core FHIR specification, which begins to address this issue by sub-setting the constituent code systems, e.g., using the SNOMED CT substance hierarchy, but excluding certain sub-branches. This may be a feasible approach to limiting synonymy, and it suggests an approach that can be used programmatically to validate content. However, the

[1] http://www.hl7.org/implement/standards/product_brief.cfm?product_id=408

[2] *Value Set: Substance-Reactant for Intolerance urn:oid:2.16.840.1.113762.1.4.1010.1*
A substance or other type of agent (e.g., sunshine) that may be associated with an intolerance reaction event or a propensity to such an event. These concepts are expected to be at a more general level of abstraction (ingredients versus more specific formulations). This value set is quite general and includes concepts that may never cause an adverse event, particularly the included SNOMED CT concepts. The code system-specific value sets in this grouping value set are intended to provide broad coverage of all kinds of agents, but the expectation for use is that the chosen concept identifier for a substance should be appropriately specific and drawn from the available code systems in the following priority order: NDFRT, then RXNORM, then UNII, then SNOMED CT. This overarching grouping value set is intended to support identification of drug classes, individual medication ingredients, foods, general substances and environmental entities. Value set intensionally defined as a GROUPING made up of: Value Set: Medication Drug Class (2.16.840.1.113883.3.88.12.80.18) (NDFRT drug class codes); Value Set: Clinical Drug Ingredient (2.16.840.1.113762.1.4.1010.7) (RxNORM ingredient codes); Value Set: Unique Ingredient Identifier - Complete Set (2.16.840.1.113883.3.88.12.80.20) (UNII ingredient codes); Value Set: Substance Other Than Clinical Drug (2.16.840.1.113762.1.4.1010.9) (SNOMED CT substance codes).

respective systems are not designed to classify substances by likely cross-reactivity, so significant overlap and spurious concepts are likely to remain.

The other approach, and that taken by this project, is to define a short list of likely concepts that should be used preferentially. This short (preferably under 1000) list of substances, substance classes, and mixtures will be chosen purely based on observed frequency of use. These concepts will support almost all allergy and intolerance records, and they will do so in a form that will allow clinicians to develop familiarity with the list and avoid confusion.

We recognize that standard representations of the concepts on this list would also be useful. While this was not a primary goal, it not only would support unambiguous use of the identified concepts, but it turned out to be a necessary part of the analysis process. Aggregating data from diverse systems required us to identify a system of record in order to disambiguate similar concepts consistently.

The intended use for these subsets follow:

1) When capturing information, the user should attempt to select an appropriate value from this list—and if more than one value fits, the most granular. For example, use the RxNORM ingredient level code as opposed other RxNORM codes (e.g. BCD, SCD, BN, SY, UNII ingredient, NDFRT code, etc.) that might include this ingredient. This makes it easier for downstream systems to interpret this code correctly without complex inferencing.

2) When sending data to some other system, send the originally captured text (and local encoding if available) for human review, but use the value from list (if appropriate) as the standard code. For example, if sending data to represent the patient statement “I am allergic to Percocet,” send “Percocet” and the local coding (perhaps 'RxNORM:42844|Percocet|'), but send the RxNorm value “214183|Acetaminophen / oxyCODONE” as the standard value so that downstream users might clearly understand both the information as it is captured, and what decision support or reconciled equivalent it should be matched with.

3) When interpreting data from other systems, be able to understand and trigger appropriate any appropriate logic for the value in this set *at a minimum*. In other words, if a DSS system can create an allergy alert for a user placing an order for "Percocet" it should recognize the code RxNORM:214183 and generate the alert.

4) This list in no way restricts the recording of substances not included in this list, and they may be encoded and used in DSS; this list represents a kind of threshold for semantic interoperability. So, when sending values not in this

subset, no expectation of decision support or reconciliation can be reasonably expected.

This document represents implementation guidance to US realm for existing interoperability standards that represent allergies. Other realms may adapt this set to country specific terminologies using alternative equivalent coding where license restrictions or ingredient code systems may be in conflict with local needs.

The output of this project is a collection of domain-specific value sets of substances based on the values identified here. The Patient Care workgroup will publish and maintain these value sets within the Value Set Authority Center (VSAC) maintained by the NIH National Library of Medicine (NLM).[3] To link the various domain specific value sets created, a grouping value set will be used to link drug, food, environmental and negation value sets.

Changes will follow the rules used in this iteration unless and until reasons are identified to modify those rules.

3. History and Context

This project is driven in part by the need emphasized by the US realm adoption of Consolidated CDA and the unwieldy value set it implies, as outlined above.

Other drivers include concerns about the quality of allergy data in patient records, most notably the concern around alarm fatigue resulting from inaccurate information or information concerning subcritical risks.

The US Pharmacopeia is conducting a similar effort. Key methodological differences include that effort's focus on a small initial set of drug classes (statins, nsaids, opioids, and antibiotics) and the effort to establish substance-based classes of manifestation, including criticality.

The collection of allergy data poses a number of issues which must be considered when evaluating data from electronic health records.[4] We describe the process allergy information capture and use in order to identify relevant assumptions and issues.

It is possible to be allergic or have intolerance to almost any substance. Substances causing reactions are commonly medications, but may also include foods or environmental substances. In many cases, a reaction is recorded to a prescribed or administered medication: in these cases, identification of the causative agent is straightforward. However, in most cases, this information is captured as a patient's

[3] NIH National Library of Medicine Value Set Authority (VSAC): https://vsac.nlm.nih.gov/

[4] Comments courtesy of Larry McKnight

response to a question, and it may be vague or inaccurate. For instance, many patients state they were told they have a Penicillin allergy as a child, but have no memory of the event or reaction. Because of the vague provenance, many substances are recorded as allergies which the patient clearly tolerates.

Human entry of data is typically supported by picklists designed to ensure that sensitivities are captured in a way that decision support systems—specifically, drug and diet order check rules—can screen orders for contraindications. The terms in these lists follow a Zipf[5] distribution, where most records can be described with a small set of values, but where higher percentages of coverage require exponentially higher numbers of terms in a "long tail" that is impractical to encode.

Once captured, this information may also be provided to other systems in structured documents or messages, but the quality of the data depends on the source provenance. During an encounter, this information may be confirmed by query, e.g., "I see that you have allergy to statins; is that true?"

Allergy records are used to inform decision support rules. However, the variety of possible allergies, options for encoding in different systems, and the uncertain specificity of information available to the clinician regarding the exposure, (not to mention common misconceptions regarding sensitivity) all mean that the rules have low specificity, resulting in many spurious alerts.

The variety and breadth of the substance concepts also means that reconciling allergy data from one or more systems is time consuming because of the many ways data may be recorded.

This understanding points out several issues, and some tactics for addressing them:

- We expect that using a short heuristic list will reduce the number of alerts for redundant encodings; it will not, of course, reduce alerts for incorrect or low-criticality records.
- We address some of the more common immunology and sensitivity misconceptions in the guidance section of this document, and we exclude from our list concepts that are clinically not actionable.
- We also observe that there are no "drug class" concepts designed to capture cross-reactive substances. Cross-reactivity may be adequately represented by a corresponding NDF-RT class--e.g., a class based on chemical structure or method of action--but this basis is certainly not always valid, and where it may be, it has not been proven.
- We note that the capture of combinations of substances (such as "Percocet") may be followed by information about a patient's ability to tolerate one

[5] George K. Zipf (1935) The Psychobiology of Language. Houghton-Mifflin.

component. The ability to remove items from the allergy list when this happens will also reduce the burden on providers resulting from spurious alerts.

4. Considerations

We point out several considerations that may help inform readers about the goal and the constraints we encountered in the process.

a. This is not pharmacovigilance

The primary purpose of these substance value sets is *not* to support pharmacovigilance (e.g. monitoring adverse events or reactions). The pharmacovigilance case requires identifiers for administered substances far more specific than those for sensitivity risk. Specific reactions to substances should record as much detail as possible about the substance, including dose, brand, manufacturer, and lot number. The substance concept used in the allergy record, however, is a more general concept used to identify other products that might contain the substance. If the precise substance is known, a good drug knowledge base check product ought to be able to determine whether a proposed ordered substance contains a relevant active moiety, but the substance list also aims to inform the clinician of substances to avoid prior to order, as well as to support cases where the ordered product is not known.

b. Class definitions

There is no system that defines cross-reactive substance classes. NDF-RT and ATC define classes, but they do so by enumeration. Any intensional[6] semantics in these groups are presumed to be supplied by the stated axes for some axes (mechanism of action, chemical structure, etc.), or not, for others (Established pharmacologic class, e.g.).

SNOMED CT could provide classes defined specifically for cross-reactivity, but it does not do so now. We adopt the general SNOMED CT substance classes as an interim measure, and we observe that substance classifications designed for cross-reactivity remain a gap in our informatics landscape.

Note that classifications provided in our list for food and environmental allergens are used to assist with searching and are not at this time meant to imply cross-reactivity or any other biological or clinically relevant relationships.

[6] Value sets defined by intension are value sets that are defined by a computable expression that can be resolved to an exact list of codes. HL7 Value: Set Definitions: http://wiki.hl7.org/index.php?title=Domain_and_Value_Set_Definitions_and_Binding#Value_Sets

c. Mixtures

Since allergy statements may be captured from the user in various forms, it is not infrequent to have patients state allergy to brand forms that contain multiple ingredients, where it is in general not known which ingredient is the offending agent. Currently systems may send this either as separate allergy statements (e.g. allergy to 'Oxycodone', and separately allergy to 'Acetaminophen'), or as a single code representing the multiple ingredients (e.g. RxNORM MIN).

Often there may be common assumption or other direct evidence that the offending substance cannot be one of the ingredients. However, a gap exists in methods to assert positive tolerance to an agent (for example, the patient was witnessed to tolerate the 'Acetaminophen' without adverse reaction). Codification rules are therefore needed to ensure that interpretation of multiple ingredient forms are given correct interpretation in relation to their individual components, and that when a substance has been vindicated, it can avoid generating inappropriate alerts and causing alert fatigue.

For the purposes of frequency, we capture “multiple ingredient” substances as asserted, understanding that there is likely one sensitivity, but that which substance is responsible is unknown. We leave the important question of how this record is maintained to the application designers.

d. Terminology system selection

We evaluated the following systems for use in providing standard identifiers for substances. In Table 1, columns 2-6 assess the systems’ respective coverage of the identified domains; 7-10 capture other important requirements. Systems selected for this iteration are in green.

System	Substance	Mixture	Class	Food	Env	Realm	Li
SNOMED CT	Yes	For products	Yes	Yes	Yes	International	Re
RxNorm	Yes	Yes	No	No	No	US	Fre
UNII	TBD	No	No	Specific	Specific	US	Fre
G-SRS	TBD	TBD	No	TBD	No	International	Fre
NDF-RT	No	No	Possible	No	No	US	Fre
ATC	No	No	Possible	No	No	International	Fre

Table 1: Candidate substance terminologies

We found NDF-RT and ATC to offer many classes that seemed useful. On closer examination, however, the boundary definitions were problematic. Two classes with the same name ("opioid agonist," e.g.) might refer to different constituent substances. The bases for classification are specified, however, and they are not cross-reactivity. For this reason, we chose to use SNOMED CT for substance classes: even though the SNOMED concepts also list children, they do so in an "open world" context: there is no implication that the enumerations constitute the semantics of the class, and no consequent opportunity for inadvertent contradiction.

SNOMED CT seemed best for food and environmentals for similar reasons. It offers classifying concepts, and it does so without implying specific memberships.
We found UNII to be often too precise for the patient safety use case, providing concepts for many kinds of rockfish, e.g., but not rockfish in general. SNOMED CT offers more variable granularity, which supports these cases well, especially for foods, environmentals, and drug classes. It does not, however, define the level of granularity, making the more predictable RxNorm the preference for specific drug substances.

For specific drug substances, UNII is a viable option. However, it does not support mixtures, as RxNorm does. Because we needed RxNorm for mixtures, we felt it best to minimize the number of systems and use RxNorm for specific substances as well. We expect this choice to be easily reversible, as there are maps from UNII to RxNorm.

G-SRS is an effort by the US Food & Drug Administration (FDA), the European Medicines Agency (EMA), and affiliated organizations to establish an international registry of substances, primarily but not exclusively for pharmacovigilance. The current pilot,

available at https://tripod.nih.gov/ginas/, seems to use UNII identifiers and would, we expect, face similar issues for the patient safety use case. The stated goal of leveraging the ISO Identification of Medical Products (IDMP) suite of standards suggests that mixtures and classes may be supported in the future. When the project goes live, we will be able to assess these hypotheses.

We use RxNorm for drug substances, including the ingredient (IN) and multiple ingredient (MIN) term types. We do not use precise ingredient (PIN) terms, as they tend to identify salts without immunological relevance.

The WHO's International Nonproprietary Names (INN) does not seem to have an accessible aggregate publication, nor does it publish concept identifiers for its preferred strings.

SNOMED CT is our current choice for foods, environmentals, and drug classes. [7] It has broad international reach, but licensing issues remain for many jurisdictions; we will continue to review our choices as events unfold.

e. Criticality

The goal of the work is to make it easier to work with the substance list, and consequently to provide better care and data interoperability. One of the threats facing patient safety is alarm fatigue, and a shorter list may help with this problem by making it easier to find common representations of common substances and reducing the likelihood of redundant records. Another tactic to address alarm fatigue may be to identify which sensitivities are critical in order to focus time and attention on risks more likely to cause harm. It has been suggested that, in some cases, substances can be categorically associated with criticalities, and that this inferred criticality can then be used to grade alerts. This is not an objective of this project, but other projects are investigating this question, and there may be opportunities for constructive engagement.

A full discussion on criticality, as well as the difference between severity and criticality in the documentation of an adverse reaction can be found in the Patient Care WG Allergy and Intolerance Domain Analysis Model[8]. Below is a brief explanation provided by Dr. Russell Leftwich on this topic taken from Appendix A of the Allergy and Intolerance Domain Analysis Model:

[7] Goss et al. agree that SCT & RxNorm "can satisfy most criteria" for allergy substance records based on coverage (Goss FR, et al. J Am Med Inform Assoc 2013;20:969–979. doi:10.1136/amiajnl-2012-000816).
[8] HL7 Version 3 Domain Analysis Model: Allergy and Intolerance Release 1 http://www.hl7.org/implement/standards/product_brief.cfm?product_id=308 (Accessed on July 27, 2017)

Discussion of Criticality – Russell B. Leftwich, MD
Severity and criticality are two related but distinct concepts in the domain of allergic and intolerance reactions.

Severity is an attribute of a symptom or a sign that is part of a reaction or an attribute of the constellation of signs and symptoms that constitute an episode of a reaction. Since there are a variety of different signs or symptoms and a variety of different reaction types, it would not be plausible to have a single rating scale that could be applied to different symptoms or two different types of reactions. It is true that rating scales have been established for research purposes to compare different episodes of a reaction type, such as anaphylaxis. It is also true that symptoms or reactions themselves are considered to have a range of severity and this is often divided intuitively into mild, moderate, and severe with mild and severe intuitively representing the two ends of the spectrum.

The list of allergies and intolerances for an individual is a list of conditions that represent a propensity to have a reaction if exposed to a specific substance in the future. This is based on a history of one or more past reactions. The potential seriousness of a future reaction is an attribute referred to as criticality. This represents a clinical judgment about the worst case scenario for a future reaction. It would be based on the severity of past reactions, the dose and route of exposure that produced past reactions, and the life-threatening or organ system threatening potential of the reaction type.

Although the list of allergies and intolerances for an individual might refer to a severe penicillin allergy or severe bee sting allergy, and the meaning is clear, this is not appropriate from a modeling standpoint. The model breaks down when the reaction type is not the presumed anaphylactic reaction of the penicillin allergy or the bee sting allergy.

As an example to contrast severity and criticality, an individual might have severe vomiting as an intolerance reaction for sulfa drugs. This reaction would be listed as a sulfa drug intolerance with low criticality, since the potential for serious injury from this is low. An individual who had a reaction immediately after a bee sting consisting of generalized itching, hives, and wheezing, which resolved without treatment would be considered to have had a mild anaphylactic episode. That individual's condition of anaphylactic sensitivity to bee stings would be considered of high criticality, because of the life-threatening potential.

High criticality does not equate to a future severe reaction, but rather the potential for a severe and life-threatening reaction. Most reaction types are dose dependent, including anaphylaxis. Therefore, although they have a sensitivity of high criticality, exposure to a small dose of the substance to which they are sensitive might result in only a mild reaction. Severity of the reaction is also dependent on the route of exposure, but criticality since it applies to the condition, is not.

f. Length

In order for the list to be useful, it must be long enough to meet most content requirements without becoming cumbersome. For medications, we found that mapping all concepts down to a count of 1000 (out of a total of 81MM) resulted in coverage of 97.7%. This number is lower for vaccines (84.1%), as there is great deal of specificity in the vaccine strings; however, the dozen vaccine concepts cover the great majority of lower-frequency unmapped strings.

For foods, all concepts with a frequency of > 1000 (from a total sample size of 3 M records) as well as concepts included on international food allergy labeling lists we used to develop the value set. This inclusion criteria resulted in a value set representing 98.5% of food reported in over 100 million allergy records.

g. Vaccines

Vaccine concepts in RxNorm are based on the Centers for Disease Control's CVX codes. These codes are very specific, and we felt they would not be optimal for our heuristic purpose.

The International Consensus (ICON) on allergic reactions to vaccines provides a comprehensive evaluation of research on this topic[9]. While the reference does not discuss the representation of vaccines in an EHR, it does provide an international consensus on the evaluation and management of allergic reactions to vaccines, and it does so following a general classification.

We have chosen to use SNOMED concepts at the level of granularity suggested in the ICON report.

[9] Dreskin, S.C. etal. International Consensus (ICON): allergic reactions to vaccines. World Allergy Organization Journal 2016;9:32 https://doi.org/10.1186/s40413-016-0120-5

h. List crossover

Some substances (notably caffeine, ethanol) appear on both drug and food sensitivity lists. The respective food and drug lists address these substances with their own methodologies. For the purpose of this publication, we include such substances identified with both RxNorm and SNOMED CT identifiers.

5. Approach

Our basic approach was to solicit allergy frequency counts from several large healthcare systems, harmonize their record strings to standard concept codes, aggregate counts for the values so identified, and select a threshold for high-frequency substances.

a. Collection

We solicited counts of allergy records from several large health systems, classified by substance. Lists were compiled by domain – drugs, foods, environmentals and statements of absence—and then analyzed by frequency. E.g., if 1000 patients have allergies to 100 substances, but have on average 1.5 allergies each, the system would provide a list of 100 substance names with counts totaling 1500. This approach was used to determine substances that are most frequently found on allergy lists to best inform the creation of constrained lists of substances.
The following institutions responded: (See Appendix A for counts from each data source)

- US Veterans Administration (VA)
- Cerner Population Health
- US Department of Defense (DOD)
- Kaiser Permanente
- Intermountain Healthcare
- University of Nebraska
- Cleveland Clinic (food only)
- NIH Clinical Center (food only)

b. Analysis of Medication, Medication Classes, Vaccine and Biologics

i. We asserted mappings from the reported strings to concept codes in standard terminologies based on maps provided by data contributors, the NLM's RxMix mapping tool, and Health Language's Language Engine. We ensured mappings for strings occurring with a combined data set frequency of at least 1000 observation counts: this threshold resulted in coverage of 97.7% of instances.

ii. We confirmed the mappings by multiple independent sources, manually reviewing cases with too few or divergent maps. (See quality assurance passage below.)

iii. We summed the counts over the standard mappings to produce the lists to come up with relative frequencies of medications, medication classes, vaccines and biologics.

c. Analysis of Food

i. Each data set was annotated for food substances and products by subject matter experts representing the Academy of Nutrition and Dietetics. (Note that a small data set from Nebraska was not included in the analysis.)
ii. Text strings were parsed into component foods.
iii. Lists from each source were then combined and normalized, e.g., combining counts for hazelnuts and filberts (a regional term difference), combining oranges and orange juice, as well as combining singular/plural terms and misspellings.
iv. Food substances with a frequency of 1000 or greater were then classified by type: additives, alcohol, artificial sweetener, bee products, chocolate, coconut, dairy, dye, eggs, fish, flavorings, fruit, gluten, grains, legumes, meat, molluscs, nuts, oil, other, poultry, seafood, seeds, shellfish, soy, spices and vegetables. In addition, foods that appear on food labels (either regulatory or voluntary) were added to the value sets. . These classes differ from classes proposed by Plasek et al. at Harvard Partners[10], where fruits and vegetables were combined, extracts included oils, and fungus included mushrooms and yeast.
    a. NOTE: classes for food allergens are used to assist with searching and are not at this time meant to imply cross-reactivity or any other biological relationships.
v. Foods and classes were then mapped to SNOMED CT using Apelon TermManager[11].
    a. When a term was not available via TermManager, the IHTSDO SNOMED CT browser was used to manually retrieve terms.
vi. Quality Assurance
    a. Lists and data mappings were reviewed and verified by subject matter experts

d. Analysis of Environmental Allergens

a. Three large data sets (VA, DOD, Cerner Population Health) were combined to create a discrete set of environmental allergens.
b. Frequencies of greater than 100 records were mapped to SNOMED CT using Apelon Term Manager.
c. Quality assurance compared mappings provided by contributory systems to mappings provided by Apelon Term Manager.
d. NOTE changes in R2
    1. Seasonal allergies were removed from the environmental allergen value set. The term seasonal allergies is not considered to be an actionable concept.
    2. The SNOMED CT mapping for nickel was changed to nickel sulfate. It is the salt of nickel that can cause an adverse reaction on contact (e.g. jewelry).

[10] Plasek, JM Food Entries in a large allergy data repository. J Amer Med Assoc 2016 Apr;23(e1):e79-87. https://www.ncbi.nlm.nih.gov/pubmed/26384406
[11] Apelon Term Manager Version 1.6 Copyright © 2017 Apelon Inc., 750 Main St #1500, Hartford, CT 06103, SNOMED CT version SNOMED CT [2017.01.16AB]

e. Analysis of Negation Terms

Data sets contained records reflecting no allergies, no drug allergies, no food allergies as well as other negation terms. Rather than include these terms on specific domain substance lists, a negation value set was developed and mapped to SNOMED CT.

f. List Output

The Patient Care WG will after the completion of this informative ballot publish five value set lists (Medications, Food, Environmental and Negation, and Consolidated) in a globally accessible location that supports automated retrieval – currently the National Library of Medicine's Value Set Authority Center (VSAC).

Note that existing value set specifications do not support metadata, e.g., frequencies.

The resulting lists of substances most frequently found on system allergy lists can be viewed in the spreadsheet attachments to this ballot. The four spreadsheets include

a. Drugs
b. Food
c. Environmental
d. Negation terms

6. Issues

a. Source data discrepancies

Source data was collected opportunistically. Donors did not always have precise characterizations of their populations or time frames. Also note that the provenance of data is also unknown, e.g., whether data was added to an allergy list by a clinician based on observation vs. a patient reported allergy or intolerance. One data source provided associated reactions with their data set.

The largest set (Cerner) contains data based on health information exchanges, and as such certainly contains a significant number of duplicated records, but the contributors were unable to provide information for de-duplicating this data. We suspect that such duplication may magnify the contributions of medications for critical chronic conditions (e.g., statins), but we have no remedy. (Note, however, that statins had a significantly higher representation in VA data, which does not suffer from this uncertainty, than in the Cerner data.) Dataset sizes range by over two orders of magnitude, from ~150k to ~60MM.

b. Discrepancies between best practice and reality

We observed several cases where data records seem to reflect common misconceptions. Particularly in the food domain, we were able to disambiguate these terms and provide better concepts. In the medication domain, we were unable to discern intent and untangle the issues. For both cases, we offer guidance in section 9.

7. Quality assurance for medication data

The value sets provided are a heuristic tool for user interface support: we are not providing mappings of actual patient conditions for clinical use. The misranking of one or more concepts, in other words, is unlikely to cause serious harm to patients or serious inconvenience to providers. That said, we do wish to provide data that is as accurate and useful as possible.

For foods and environmentals, quality assurance is managed by multiple reviewers. For medications, the volume of data made this impractical.

Most of the medication data contributors provided mappings of their own to standard terminologies, so we assume that the intent of these strings is accurately reflected by the provided mapping. However, the selected system is not always the system we have chosen. In these instances, we have supplied new mappings. We have also mapped strings for which no mappings have been provided. Our mappings have been provided by a pilot implementation of Health Language's Language Engine[12], queries against the US National Library of Medicine's RxMix medication terminology API, and manual search of the terminologies.

In cases where we have mappings for a string from multiple sources to the same concept, and no mappings that disagree, we assume that the mapping is correct.

In cases where we have multiple mappings to the same concept and at least one mapping disagrees, and in cases where we have only one mapping, we have manually reviewed the mappings and selected the appropriate one, based on these guidelines:

1. RxNorm for substances (IN) and mixtures (MIN)
2. SNOMED CT (substance) for classes
3. No salts (PIN) or brands

In the example below, row 1 breaks rule 3; row 2 seems appropriate; row 3 contains an error; and row 4 maps to the wrong system. We ran another query against the HL (Health Language) Language Engine to generate row 5, which confirms row 2.

[12] Language Engine Access Portal by Health Language. Accessed 9/28/16. Copyright Wolters Kluwer (http://wolterskluwer.com/).

Num	Provided string	Map source	Map name	Map code	Map System	Comment
1	Dihydroergotamine	Provider	Dihydroergotamine mesylate	203176	RxNorm	Salt
2	Dihydroergotamine	Provider	Dihydroergotamine	3418	RxNorm	Appropriate
3	DIHYDROERGOTAMINE	Provider	Dihydroergotoxine	3419	RxNorm	Error in supplied map
4	Dihydroergotamine	Provider	Dihydroergotamine	387267005	SNOMED CT	Not chosen system
5	DIHYDROERGOTAMINE	HL LE	Dihydroergotamine	3418	RxNorm	Appropriate

Table 2: Quality Assurance Example

8. Findings

Our primary result is the list of substance values in the attached spreadsheet, which will be used to create service-accessible value sets in the NLM's VSAC publication.

a. Medications

About half of medication records assert no known allergy. After negations, medications are the most common class of assertions recorded in allergy lists.

After a relatively high peak (67% of instances covered with 38 values; 90% with 152), the approach of mapping common (>1000 instances) strings covers 97.6% of instances with 628 standard values.

There is noticeable variance in the frequencies across contributed data sets. For this reason, we provide, in addition to the value set in the main attachment, a source variation spreadsheet to illustrate the differences, available on the HL7 Patient Care wiki. We do not speculate here about the reasons for these differences.

b. Food

Food allergies are subject to international food labeling regulations based on frequency data. The data from 7 million subject records reflects the known frequency of eight major food categories in the US which serve as the basis for the FDA FALCPA food labeling laws[13] [14] [15]. The data also supports the inclusion of additional foods found in international labeling regulations. A summary of these labeling regulations and associated substances can be found in Appendix B. The food value set spreadsheet also presents findings that relate to US and international food labeling regulations.

What the data show are the other major foods that are not represented in US food labeling laws, but that are included in allergy and intolerance lists. Food sensitivities include not only protein-based foods, but a wide range of legumes, fruits, vegetables

[13] https://www.fda.gov/downloads/food/ingredientspackaginglabeling/ucm192048.pdf

[14] https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/Allergens/ucm106890.htm Accessed 7/25/17

[15] Hefle, SL, Nordlee, JA and Taylor, SL. Allergenic Foods Critical Reviews in Food Science and Nutrition 1996; 36(S):569-589

and food additives. The food data was uniform across sources and did not show any significant variability in composition.

c. Environmentals

Environmental data shows lower total frequencies than other categories. As they are recorded for clinical management and not for patient safety, we have brought more scrutiny to food and medication categories.

9. Guidance

In general, we advise against terms that are not actionable, i.e., that cannot be used to guide decisions unambiguously; e.g., “mold” (environmental respiratory issue, or a specific sensitivity?), “antibiotics” (in the unlikely event all classes of antibiotics are contraindicated, they should be contraindicated explicitly), “tape” (unclear whether there is an adhesive substance sensitivity or a skin integrity concern).

a. Seafood.

Patients have been confirmed to have allergies to mollusks, crustaceans, and certain varieties of fish. “Seafood” may be an attempt to generalize one or more of these concepts, but the term seafood is ambiguous and therefore a non-actionable term. The term seafood would exclude snails, which are mollusks, or may be taken to exclude freshwater fish. Its use should be avoided. However, the data show an overwhelming use of this term.

In addition, “shellfish” should be disambiguated as either mollusks, crustaceans, or both.

RECOMMENDATION:

i. Do not use the term seafood. Use specific terms for fish, mollusks or crustaceans.
ii. Do not use the term shellfish. Use specific terms for mollusks or crustaceans.

b. Iodine.

Patients with “seafood” allergies are often assumed to have iodine sensitivities, which may result in a contraindication to using radiocontrast media. This may result in either unnecessary use of more expensive low osmolality media or reduced use of radiocontrast media altogether with decreased diagnostic sensitivity for the patient.

It is true that the use of low-osmolality radiocontrast media carries an approximately ⅔ reduction of risk of serious reactions. The reductions is not however related to the presence of iodine since low-osmolality radiocontrast media is also an iodinated

compound[16] [17]. The sensitivity, however, is not allergic but related to activation of the complement system; it is only patient-specific because driven by the patient's condition.

The frequency numbers for iodine, iodinated contrast media, and contrast media are almost certainly inflated.

RECOMMENDATION:

i. Do not record "iodine" for mollusk, crustacean, or fish sensitivities.

c. Penicillin.

Studies show that penicillin allergies are vastly over-reported, but the only data we have is that they are reported, so we have no basis for modifying the data.

Two tactics may improve the situation. First, the creation of a more accurate set of cross-reactive classes may support more precise recording of actual sensitivities. Given the frequency of a reported penicillin allergy in the US population (8%) vs. the actual number of those who demonstrate a sensitivity to penicillin (1%)[18] [19], the clinical safety of assuming cross reactivity based on a documented penicillin allergy is ill-founded. The risk of a penicillin allergy should be based on clear documentation of what specific medication caused the reaction (penicillin or other beta-lactams) without the use of a medication class. Second, the ability to record an override reason, and to persist that reason with an allergy record, may support grading the criticality of the records and reduction in alarm fatigue.

RECOMMENDATION:

i. Do not use a medication class to represent penicillin or other beta-lactams. Be as specific as possible about the product a patient is believed to have reacted to and enter that on the allergy list, not a class or category.

d. Vaccines

Many reaction records list very specific vaccines. At the reaction level this is perfectly appropriate, but it's unclear at the allergy level whether a reaction to a vaccine with a specific set of antigens is a useful indicator for other vaccines, and if so which ones. See footnote #5 for further information on this issue.

---

[16] Schabelman, E, Whitting, M. the relationship of radiocontrast, iodine and seafood allergies: a medical myth exposed. J Emerg Med 2010 Nov; 39(5):701-7. https://www.ncbi.nlm.nih.gov/pubmed/20045605

[17] Huang, SW. Seafood and iodine: an analysis of a medical myth. Allergy Asthma Proc 2005 Nov-Dec; 26(6):468-9. https://www.ncbi.nlm.nih.gov/pubmed/16541971

[18] https://www.aaaai.org/global/latest-research-summaries/New-Research-from-JACI-In-Practice/penicillin-label

[19] Vyles D, Adams J, Chiu A, et al. Allergy Testing in Children With Low-Risk Penicillin Allergy Symptoms. Pediatrics. 2017;140(2):e20170471

e. Oils

The analysis of data indicated an adverse sensitivity to a number of edible oils. Based on the literature, edible oils as a derivative of a nut (e.g. almond oil), seed (sesame seed oil), legume (e.g. peanut oil) or vegetable (e.g. vegetable oil, olive oil) do not contain protein and therefore would not cause a Type 1 immediate hypersensitivity reaction.[20] [21] [22]

f. Environmental allergens

Non-actionable terms were included in the frequency data but are not recommended for inclusion on an allergy list. Terms such as "seasonal allergies" do not convey adequate specificity.

RECOMMENDATION

i. Use actionable substances such as "cat dander" for inclusion on the allergy list.

g. Nickel

The term "nickel" was prevalent but to reflect the actual reactive compound the term was mapped to nickel sulfate, the salt form. Nickel in it's pure form does not cause an adverse reaction on contact.

RECOMMENDATION

i. The term nickel should be mapped to nickel sulfate, the substance which engenders the adverse reaction.

[20] Taylor, SL etal. Peanut oil is not allergenic to peanut-sensitive individuals. J. allergy Clin. Immunol. 1981 Vol. 68; No. 5: 372-375

[21] Fremont, S. etal Allergenicity of oils. Allerg Immunol (Paris) 2002 Mar;34(3):91-94.

[22] Crevel, R.W. R., etal. Allergenicity of refned vegetable oils. Food and Chemical Toxicology. 2000 38:385-393.

Appendix A: Record Counts by Data Source and Type

Source	Null	Biologic	Environmental	Food	Negatives	Medications	Medication Classes	Supply	Vaccines	TOTAL
Cerner Population Health	296,556	3,925	1,719,054	3,888,006	16,147,817	29,603,050	6,157,854	1,098,880	208,702	59,123,844
US Department of Defense (DOD)	7,433		42,927	72,599	1,909,269	561,231	287,107	818	7,006	2,888,390
Intermountain Healthcare System	30,153	102	885,001	876,886	1,505	2,190,724	237,751	27,677	7,403	4,257,202
Kaiser Permanente	38,470	1,227	139,709	1,259,320	988,036	4,137,845	1,140,079	116,997	41,153	7,862,836
Nebraska	4,863	-	4,319	4,704	1,638	107,187	29,081	4,971	919	157,682
US Veterans Administration (VA)	841,395	-	404,969	873,103	24,861,191	5,162,600	1,030,970	106,417	666	33,281,311
National Institutes of Health Clinical Center	-	-	-	23,141	-	-	-	-	-	23,141
Cleveland Clinic	-	-	-	3,995	-	-	-	-	-	3,995
TOTALS	**1,218,870**	**5,254**	**3,195,979**	**7,001,754**	**43,909,456**	**41,762,637**	**8,882,842**	**1,355,760**	**265,849**	**107,598,401**

Appendix B: International Labeling Regulations for Food Allergens

http://farrp.unl.edu/IRChart

http://farrp.unl.edu/documents/Regulatory/International%20Allergens%205-25-17.pdf

Citation: University of Nebraska – Lincoln, Institute of Agriculture and Natural Resources, Food Allergy Research and Resource Program. Website accessed on July 23, 2017

In order to engineer a better representation of allergy list semantics, certain use case assumptions were made:

To address these issues, this document proposes the attached subset(s), derived from actual allergy statements on real patients from several large health systems to determine the frequency distribution of allergy statements in raw text form, and therefore derive a relative concept distribution (were 2 or more terms may indicate common ingredient or substance class of concern, but the class is sufficiently granular to trigger important downstream alerting). This concept list has a goal of representing the 99% use case in a relatively small list (less than 1000 concepts). Secondarily it represents the preferred coding of those concepts, and methods to determine a preferred coding, even if not in this subset. The 99% use case is taken so that given some user input, or external records from some external system, a reconcile algorithm or decision support system might recognize at least these 'common' (>=99%) items, but with understanding that in a small percent of cases (e.g. <=1%) the record might only be understood by a human reading the free text because the item falls in a very long tail of items the computer can't understand reliably. In other words, in 100% of cases, a human should be able to record and read back some information, but 99% of the time that information can be represented in this subset, and these codes should therefore be used preferentially so that multiple computer systems might act on it reliably.

The intended use for these subsets are therefore:
1) When capturing information from a user, that if the allergen substance or allergy statement can accurately be codified using a code in this, then it should be codified to the most granular code in this list preferentially, rather than choosing the from the many variations of granularity available to represent any possible substance. For example, use the RxNORM ingredient level code as opposed other RxNORM codes (e.g. BCD, SCD, BN, SY, UNII ingredient, NDFRT code, etc) that might include this ingredient. This makes it easier for downstream systems to interpret this code correctly without complex inferencing.
2) When sending data to some other system, where the information can be interpreted as an item from this list send the originally captured text (and local encoding if available) for human review, but use the code from list as the 'Standard encoding'. For example, if sending data to represent the patients statement 'I am allergic to "Percocet"' send 'Percocet' (and granular UI terminology, local code, or perhaps 'RxNORM:42844|Percocet|') the local coding, but send 'RxNORM:214183|Acetaminophen / oxyCODONE|' as the standard encoding so that downstream users might clearly understand both the information as it is captured, and what decision support or reconciled equivalent it should be matched with.
3) When interpreting data from some other system, be able to understand and trigger appropriate any appropriate logic with _at_minimum_ the codes in this set. In otherwords, if a DSS system can create an allergy alert for a user placing an order for "Percocet" it should recognize the code RxNORM:214183 and generate the alert.
4) Codification and interpretation of other items in the 1% use case is permitted as local encoding that may be more granular than listed in this subset, however this subset represents a kind of 'minimum bar' for semantic interoperability. So, if sending codes not in this subset, no expectation of reconcile or decision support can be reasonably expected. However, guidance is given on what would likely be the most appropriate code based on more general terminology choice preferences.

This document represents implementation guidance to US realm for existing standards that represent allergies in FHIR, CDA, or HL7 V2. Other realms, may take this set and adapt it to country specific terminologies using alternative equivalent coding where license restrictions or ingredient code systems may be in conflict with local needs.

Generation of this subset has additionally identified a standards gap in representing that the patient has

explicit tolerance to a given substance, and therefore some part of an allergy interpretation can be excluded. For example, to record that the patient has allergy to 'Percocet', but tolerates 'Acetaminophen', thus implicating the oxycodone, and avoiding the alert fatigue from inappropriate interpretation as allergy to acetaminophen.

