Perspective

Formalizing drug indications on the road to therapeutic intent

Stuart J Nelson,[1] Tudor I Oprea,[1] Oleg Ursu,[1] Cristian G Bologa,[1] Amrapali Zaveri,[2] Jayme Holmes,[1] Jeremy J Yang,[1] Stephen L Mathias,[1] Subramani Mani,[1] Mark S Tuttle,[3] and Michel Dumontier[2]

[1]Translational Informatics Division, Department of Internal Medicine, University of New Mexico School of Medicine, Albuquerque, NM, USA, [2]Institute of Data Science, Maastricht University, Maastricht, The Netherlands, and [3]Center for Digital Health Innovation, University of California, San Francisco, CA, USA

Corresponding Author: Tudor I Oprea, Translational Informatics Division, Department of Internal Medicine, University of New Mexico School of Medicine, Albuquerque, USA. E-mail: toprea@salud.unm.edu

ABSTRACT

Therapeutic intent, the reason behind the choice of a therapy and the context in which a given approach should be used, is an important aspect of medical practice. There are unmet needs with respect to current electronic mapping of drug indications. For example, the active ingredient sildenafil has 2 distinct indications, which differ solely on dosage strength. In progressing toward a practice of precision medicine, there is a need to capture and structure therapeutic intent for computational reuse, thus enabling more sophisticated decision-support tools and a possible mechanism for computer-aided drug repurposing. The indications for drugs, such as those expressed in the Structured Product Labels approved by the US Food and Drug Administration, appears to be a tractable area for developing an application ontology of therapeutic intent.

Key words: therapeutic intent, drug indications, drug labels

PROBLEM: THERAPEUTIC INTENT AND THE NEED FOR PRECISION

The logic of why a physician chooses to take a particular action goes far beyond diagnosis. Guidelines for medical practice may specify the precise conditions and aspects of care for an individual, which are not adequately captured in current information systems. These aspects include disease progression, other complicating conditions, and concurrent or past treatments. To illustrate, consider olanzapine, an antipsychotic that is indicated for the treatment of schizophrenia and bipolar disorder. The US Food and Drug Administration's (FDA) approved indication is treatment of agitation in the context of these disorders. All of these various aspects can be thought of as providing a context of practice, which we are choosing to refer to as the therapeutic intent. As we progress toward a form of practice where treatment is becoming increasingly precise

with the use of genomic, proteomic, microbiomic, and metabolomic data, there is a simultaneous need to ensure that computational algorithms that use such information to suggest a course of action have access to the logic of medical practice in an accurate, computer-accessible form. Indeed, the lack of good quality information will have a negative impact on any effort to create powerful health analytics and precision clinical decision support. It is time to develop a framework for accurately and precisely representing therapeutic intent that is also FAIR (Findable, Accessible, Interoperable, and Reusable).[1]

WHERE TO START

Our previous experience with formalization[2,3] has taught us that it is often practical to take a manageable task in hand before

attempting to extend it to larger and more complicated domains. We suggest that representing the therapeutic intent of indications for medications is such an area, since the number of approved medicines worldwide is manageable by a team of human curators, assisted by computational tools.

Drug indications[4] (referred to as indications) are listed on approved drug labels (ADLs), which are provided by the sponsor with approval by the FDA. We believe that these indications are an appropriate place to start. With the exception of "grandfathered" drugs that entered the market before 1962, the legislation introduced at that time enabled the FDA to mandate drug efficacy and safety requirements. All medications are now subject to rigorous evaluation of clinical trial results for safety and efficacy. ADLs are legal documents that list medically valid reasons to administer specific drugs, with specified (maximum) dosages. Indications must be supported by evidence that specific drug-dosage combinations benefit patients with particular conditions, as documented by the sponsor (typically) following clinical trials. Indications are subject to revision, amendment, and withdrawal. They are not just disease- and population-specific, but also formulation-specific: 20 mg tablets of sildenafil citrate (Revatio®) are indicated for pulmonary hypertensive arterial disease, but the indication is male erectile dysfunction for the 50 mg tablets (Viagra®).

PREVIOUS EFFORTS

Previous efforts at representing the indications for medicines have focused on the disease involved rather than the therapeutic intent. The failure of these efforts to recognize intent in addition to the disease is easily recognized. Examples abound; without a representation of the intent, this failure will continue. Multiple attempts to track indications are listed in Table 1.

One such example, the Ensemble MEDication Indication Resource (MEDI),[5] aggregates indications from the Vanderbilt Electronic Medical Records System (not open access), RxNorm,[3] MedlinePlus,[6] Side Effect Resource (SIDER2),[7] and Wikipedia,[8] for a total of 63 344 drug-indication pairs. Of these, 13 380 are the "high-precision set." Examples of failure to represent intent from the MEDI high-precision set are shown in Table 2. In MEDI, morphine is indicated for acute myocardial infarction, but the therapeutic intent is to relieve the acute pain associated with acute myocardial infarction.

Elkin et al.[9] produced a set of DailyMed "drug has indication" semantic triplets but did not capture the depth, complexity, and diversity of data types and relationships present in ADL "indications." The National Drug File–Reference Terminology (NDF-RT)[10] effort of the Veterans Administration used extractions from published literature to establish "may_treat" relationships between drugs and diseases. Once again, therapeutic intent was not captured.

The Drug Ontology[11] links drugs to many classes and ontologies, and carries only a few therapeutic indications. These indications are not strictly mapped to a disease, yet the context of therapeutic use is not fully captured.

The example of olanzapine was drawn from LabelIn.[12] An automatic and manual curation of indications from ADLs, it covers 250 drugs. In LabeledIn, olanzapine is indicated for agitation and for schizophrenia, but the actual indication stated on the label is treatment of agitation in the context of schizophrenia.

MalaCards,[13] which provides drug indications mined from clinical trials, literature, and ADLs, also has misleading indications. For example, it associates doxazosin with pheochromocytoma. However, the indications specify hypertension and benign prostatic hyperplasia only, not pheochromocytoma. It is true that doxazosin has been prescribed[14] to prevent hemodynamic instability before surgery for pheochromocytoma. The therapeutic intent is to lower blood pressure caused by pheochromocytoma prior to surgical removal of the tumor. Pheochromocytoma surgery is a context in which doxazosin is used, but the tumor is not the reason for using the medication.

The above examples serve as reminder that we need to capture therapeutic intent and contextually represent indications as accurately as possible, or else we run the risk that algorithms based on only diagnosis will result in less than useful, potentially harmful results, given incorrect information.

THERAPEUTIC SPECIFICITY

Another aspect of the need to represent intent is specificity. Genetic abnormalities in cystic fibrosis, such as Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) mutations, have medications developed for specific patient subpopulations. Lumacaftor works for CFTR F508del, whereas Ivacaftor is effective for G551D and 9 other mutations. Ivacaftor is marketed as single ingredient (Kalydeco®), but Lumacaftor is only available in combination with Ivacaftor (Orkambi®). These 2 drugs address 11 of the approximately 1700 CFTR gene mutations and are effective for about 50% of the US cystic fibrosis population.[15] They are, however, not effective in patients who have other CFTR mutations. Such specific information, eg, mutations or preexisting medical conditions or treatments, is not captured by the existing sources. We are not aware of any online resource that provides such precision, except for the ADL document itself.

In a sense, classifications of drugs, of which there are many, might also be thought of as representing intent. The therapeutic classifications are based on what diseases or processes a drug may target. Thus, anti-epileptic drugs target epilepsy, while anti-inflammatory drugs act on the inflammatory process. Such classifications are not sufficiently robust in their present configuration to support precision medicine. Pharmacologic classifications, as expressed in such terminologies as the FDA Expressed Pharmacologic Class, are often formed by a compound of structural information (eg, thiazide) with physiologic effect (eg, diuretic). Some Expressed Pharmacologic Classes express classifications in terms of what they are not (eg, nonsteroidal anti-inflammatory agents are described partly as being "nonsteroidal"). The NDF-RT effort tried to separate the components of the classification into major groups, thus structure, physiologic effects, and molecular mechanisms of action were expressed separately (Erlbaum M, personal communication).

A successful representation of therapeutic intent of drug indications might be useful in several different areas. Most doctors appear to use only about 200 drugs for the conditions they treat.[16] Using clinical decision support tools that exploit the representation might assist practitioners in finding more specific and precise therapies for their patients. The representation could also be of use in accurately mapping off-label uses of drugs for computational reuse, or in computationally[17–19] identifying potentially novel indications for old drugs (computational drug repurposing[20]).

Table 1. Examples of online resources that specifically discuss or capture drug indications

Medi-Span from WoltersKluwer (for fee) *Not available for evaluation*	MEDI, an Ensemble MEDication Indication Resource (open-access dataset) *This resource contains unstructured text only; discussed in text*
MedKnowledge from First Databank (for fee) *This resource lacked context in 2012, when it was deposited in Observational Medical Outcomes Partnership*	LabeledIn (open-access dataset) *This resource contains unstructured text only; discussed in text*
Gold Standard Drug Database from Elsevier (for fee) *Not available for evaluation*	MalaCards (open access) *This resource contains unstructured text only; discussed in text*
Approved drug uses from PubMed Health (open access) *This resource contains unstructured text only*	DailyMed (open access) *This serves as primary source for indications* The text is structured only to the extent that the indications are in a separate section of the Structured Product Label

Table 2. Examples of MEDI-HPS indications in need of adjustment and clarification

Drug	MEDI Indication	SPL indication
Denosumab	Malignant neoplasm of prostate	Treatment to increase bone mass in men at high risk for fracture receiving androgen deprivation therapy for nonmetastatic prostate cancer
Vancomycin	Other and unspecified noninfectious gastroenteritis and colitis	Treatment of serious or severe infections caused by susceptible strains of methicillin-resistant (betalactam-resistant) staphylococci
Lamotrigine	Rash and other nonspecific skin eruption	Epilepsy and bipolar disorder
Benzocaine	Infective otitis externa, unspecified	Relief of pain and reduction of inflammation
Morphine	Acute myocardial infarction, unspecified site; episode of care unspecified	Relief of moderate to severe acute and chronic pain
Orlistat	Diabetes mellitus	Obesity management, including weight loss and weight maintenance
Sargramostim	Myeloid leukemia, acute	Prevention of neutropenia from chemotherapy

AN APPROACH: ONTOLOGY-BASED FORMALIZATION OF THERAPEUTIC INTENT

In representing therapeutic intent, it will be important to perform the work in a manner that can be updated quickly. DailyMed, as the name implies, is updated daily. Every new label will require processing to express the therapeutic intent. While natural language processing (NLP) has improved greatly over the past years, there remain limitations of NLP systems, particularly low recall under certain conditions, and difficulty in correctly parsing statements where critical information is present in preceding sentences. Multiple text-mining tools have been applied[12,21,22] to extract indications, and an open-source compendium mapping molecular entities to indications is available online.[23] While these tools and resources extract disease concepts and map them to existing terminologies, they do not represent the full logic of therapeutic intent.

Given current limitations of NLP systems to extract formal relations between biomedical entities of indications, human curation will be required to identify the relationships between biomedical entities tagged by NLP systems. Automated tagging of biomedical concepts and mapping to existing terminologies may reduce the burden on human curators and speed up the process of therapeutic intent formalization.

As part of the representation of therapeutic intent, a new formal model that takes into account relationships among diseases, symptoms, and other contextual information relevant to therapeutic intent is needed. Such an application-focused representation requires concept definitions for relevant entities, including drugs, therapeutic uses, diseases, symptoms, genes, mutations, anatomical entities, genetic variability, and other concepts. A high-level categorization of the different types of concepts will assist in this model. This model, an application ontology, will need to capture and express the relationships between different entities and formalize the therapeutic intent. Such a model will need to include addressed disease, relevant comorbidities, coprescribed medications, genetic abnormalities, temporal constraints (eg, pregnancy trimester), and other clinical findings (eg, body mass index for lorcaserin-managed obesity).

Some of these aspects can be implemented within the Systemized Nomenclature of Medicine (SNOMED)[24] framework. However, we believe that formalization of therapeutic intent using ontologies will require a new knowledge representation model. This must be supported by manual curation protocols and eg, software tools for automated and semiautomated annotation. By developing a knowledge extraction protocol based on an ontology and employing state-of-the-art biomedical text-mining tools, such an annotating application would manifest an appropriate workflow for extracting and representing semantically enhanced indications. We envision that validation of methods and results would need to be an ongoing effort, using the highest possible standards of ground truth. In this case, we consider the ground truth to be the consensus of well-trained and informed medical experts, especially physicians familiar with both clinical practice and medical informatics. Accordingly, accuracy metrics could be based on review by expert physicians and medical informaticians.

There is a sense of urgency in creating an appropriate framework for the development of such a system. This discussion centers on FDA-approved indications for the sake of brevity. However, we believe these arguments hold true for all drug uses, and the international perspective is of equal importance. A forum dedicated to this topic, inviting participation from stakeholders representing academia, industry, regulatory and funding agencies, health providers, and insurers, would help to ensure international participation. Such a forum would foster collective responsibility, which is more likely to result in continuously updated, maintained resources.

SUMMARY AND OUTLOOK

Formalizing the logic of therapeutic intent has thus far not been directly addressed, but has been an unvoiced difficulty in previous

attempts involving those issues, such as guideline development and clinical trial specifications.

We believe it is crucial to have access to a comprehensive gold standard for therapeutic intent that is (1) accurate with respect to context and provenance, (2) structured for computational reuse, (3) principled in its construction practices, and (4) normalized to precise concepts in standard vocabularies that enable downstream analysis. Developing a formal model as an application ontology and representing this in a machine-actionable form will lead to a marked improvement with respect to accuracy of capturing therapeutic intent of drug usage. While there are other areas that could potentially be used as a test bed for developing a formal model of therapeutic intent, such as clinical trials or clinical decision support, we believe that capturing therapeutic intent in indications is a promising starting point. With <4500 molecular entities in the prescription drug market,[23] the numbers are not as daunting as in other areas, but the lessons learned may be applicable to those larger areas. Given that indications are of a tractable size for capturing intent, and that representing this intent has multiple uses, we believe this area is ripe for exploration.

FUNDING

COMPETING INTERESTS

There are no competing interests.

CONTRIBUTORS

The problem was first identified by MD, TIO, SJN and evaluated by TIO, OU, CGB, JJY, SM, SJN, MST. The approach was conceived by TIO, SJN, and MD, with contributions from all other authors. All authors contributed to the writing of this manuscript and the design of the approach. All authors revised the manuscript and provided final approval of the version to be published.

REFERENCES

1. Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018.
2. Nelson SJ, Powell T, Srinivasan S, Humphreys BL. The Unified Medical Language System (UMLS) Project. In: Bates, MJ, Maack MN, eds. *Encyclopedia of Library and Information Sciences,* 3rd edn. New York: Marcel Dekker; 2009, 7: 5320–27.
3. Nelson SJ, Zeng K, Kilbourne J, *et al.* Normalized names for clinical drugs: RxNorm at 6 years. *J Am Med Inform Assoc.* 2011;18:441–48.
4. Wikipedia contributors. Indication (medicine). Wikipedia, The Free Encyclopedia. 2016. https://en.wikipedia.org/w/index.php?title=Indication_(medicine)&oldid=736851214. Accessed January 10, 2017.
5. Wei W-Q, Cronin RM, Xu H, *et al.* Development and evaluation of an ensemble resource linking medications to their indications. *J Am Med Inform Assoc.* 2013;20:954–61.
6. Miller N, Lacroix EM, Backus JE. MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. *Bull Med Libr Assoc.* 2000;88:11–17.
7. Kuhn M, Letunic I, Jensen LJ, *et al.* The SIDER database of drugs and side effects. *Nucleic Acids Res.* 2016;44:D1075–79.
8. Wikipedia. https://en.wikipedia.org/wiki/Indication_(medicine). Accessed March 23, 2017.
9. Elkin PL, Carter JS, Nabar M, *et al.* Drug knowledge expressed as computable semantic triples. *Stud Health Technol Inform.* 2011;166:38–47.
10. Carter JS, Brown SH, Erlbaum MS, *et al.* Initializing the VA medication reference terminology using UMLS Metathesaurus co-occurrences. *Proc AMIA Symp.* 2002:116–20.
11. Hanna J, Joseph E, Brochhausen M, Hogan WR. Building a drug ontology based on RxNorm and other sources. *J Biomed Semantics.* 2013;4:44.
12. Khare R, Li J, Lu Z. LabeledIn: cataloging labeled indications for human drugs. *J Biomed Inform.* 2014;52:448–56.
13. Rappaport N, Twik M, Plaschkes I, *et al.* MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res.* 2017;45:D877–87.
14. van der Zee PA, de Boer A. Pheochromocytoma: a review on preoperative treatment with phenoxybenzamine or doxazosin. *Neth J Med.* 2014;72: 190–201.
15. CFTR Modulator Therapies. Cystic Fibrosis Foundation. https://www.cff.org/Living-with-CF/Treatments-and-Therapies/Medications/CFTR-Modulator-Therapies/. Accessed March 23, 2017.
16. Taylor RJ, Bond CM. Change in the established prescribing habits of general practitioners: an analysis of initial prescriptions in general practice. *Brit J General Pract.* 1991;41:244–48.
17. Gottlieb A, Stein GY, Ruppin E, *et al.* PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol.* 2011;7:496
18. Yu L, Ma X, Zhang L, *et al.* Prediction of new drug indications based on clinical data and network modularity. *Sci Rep.* 2016;6:32530
19. Sawada R, Iwata H, Mizutani S, *et al.* Target-Based Drug Repositioning Using Large-Scale Chemical-Protein Interactome Data. *J Chem Inf Model.* 2015;55:2717–30.
20. Oprea, TI; Overington, JP. Computational and practical aspects of drug repositioning. *Assay Drug Dev Technol.* 2015;13(6):299–306.
21. Fung KW, Jao CS, Demner-Fushman D. Extracting drug indication information from structured product labels using natural language processing. *J Am Med Inform Assoc.* 2013;20:482–88.
22. Oprea TI, Nielsen SK, Ursu O, *et al.* Associating drugs, targets and clinical outcomes into an integrated network affords a new platform for computer-aided drug repurposing. *Mol Inform.* 2011;30:100–11.
23. Ursu O, Holmes J, Knockel J, *et al.* DrugCentral: online drug compendium. *Nucleic Acids Res.* 2017;45:D932–39.
24. SNOMEDCT. SNOMED Clinical Terminology, available from SNOMED international, http://www.snomed.org/snomed-ct/. Accessed May 23, 2017.