Towards a Consistent and Scientifically Accurate Drug Ontology

William R. Hogan*, Josh Hanna, Eric Joseph, and Mathias Brochhausen
Division of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

ABSTRACT
Our use case for comparative effectiveness research requires an ontology of drugs that enables querying National Drug Codes (NDCs) by active ingredient, mechanism of action, physiological effect, and therapeutic class of the drug products they represent. We conducted an ontological analysis of drugs from the realist perspective, and evaluated existing drug terminology, ontology, and database artifacts from (1) the technical perspective, (2) the perspective of pharmacology and medical science (3) the perspective of description logic semantics (if they were available in Web Ontology Language or OWL), and (4) the perspective of our realism-based analysis of the domain. No existing resource was sufficient. Therefore, we built the Drug Ontology (DrOn) in OWL, which we populated with NDCs and other classes from RxNorm using only content created by the National Library of Medicine. We also built an application that uses DrOn to query for NDCs as outlined above, available at: http://ingarden.uams.edu/ingredients. The application uses an OWL-based description logic reasoner to execute end-user queries. DrOn is available at http://code.google.com/p/dr-on.

1 INTRODUCTION

A coherent ontology of drugs can serve many purposes. Similar resources have been proposed for clinical decision support (Broverman, 1998; Sperzel, 1998; Kim, 2001), interoperability of drug data (Broverman, 1998; Nelson, 2011; Palchuk, 2010; Parrish, 2006; Kim, 2001), comparative effectiveness research or CER (Olsen, 2011), translational research (Pathak, 2011; Palchuk, 2010; Chute, 2003), and pharmacovigilance (Merrill, 2008; Saunders, 2005).

The chief use case driving our work on drug ontology at present is support of CER (a branch of translational research). A recent Institute of Medicine Report recommends semantic technologies in support of CER (Olsen, 2011). Also, author WRH was part of a research team (Kelkar, 2012a, 2012b) whereby a student had to manually identify all drug products containing acetaminophen. This team studied IMS LifeLink, a large, proprietary database of pharmacy claims in the United States (Pharmetrics Inc). The only non-proprietary drug codes in IMS Life-Link are National Drug Codes (NDCs). Had the ability to query historical NDCs been available, the research would have been more efficient.

Our initial requirement is thus to automate the generation of a list of NDCs representing products that contain any particular ingredient(s) using an ontology that is non-proprietary and open to investigators. Additional requirements include the ability to query drug products with a particular therapeutic indication (e.g., antihypertensive) or that contain ingredients with a particular mechanism of action (such as non-activating beta-adrenergic receptor blockade).

No existing resource is sufficient for several reasons. Besides other problems we report here, our technical requirement for a historically comprehensive list of NDCs was not met. Each version of RxNorm (Nelson, 2011)—a standard drug terminology curated by the National Library of Medicine (NLM)—when released contains only NDCs from recent versions of its source drug knowledge bases (KBs).

Thus, our goal is to create a correct and consistent ontology of drugs in OWL that includes historical NDCs and enables querying NDCs of drug products with particular ingredient(s) and that have particular properties. Here, we describe the theoretical foundations for the ontology. Our hypothesis was that a realism-based approach would avoid systematic errors in our representation of drugs.

2 METHODS

Our goal was to have an ontology of drugs that (1) enables query of historical pharmacy claims and electronic health record data, (2) is correct from the perspective of pharmacy and biomedical science, (3) has a set of logical axioms that do not entail untrue or inconsistent inferences, (4) and is consistent and interoperable with other ontologies across numerous levels of granularity for translational science.

2.1 Analysis of drugs and existing artifacts

To begin, we analyzed drugs and their parts from a realist perspective using the methodology of Smith and Ceusters (Smith & Ceusters, 2010) and Smith and Brochhausen (Smith & Brochhausen, 2010). Specifically, we studied (1) the composition of drugs including notions of parthood (to support querying drugs with particular ingredients), (2) the molecular mechanisms of action of what are referred to as "active ingredients" of drugs (to support query for beta-adrenergic receptor blockade), and (3) therapeutic indications (to support query for antihypertensive drugs).

We then studied existing ontological and terminological artifacts from four perspectives, including (1) *technical*, whether it meets the requirements of our use case, practices good version control, and is easy to use and understand; (2) *scientific,* including correctness of drug knowledge; (3) *description logic semantics*, analogous to the analysis of Schulz (Schulz, 2010) and Boeker (Boeker, 2011), and (4) *realism,* based on our realist analysis of drugs.

* To whom correspondence should be addressed: hoganwr@gmail.com

Through this process, we defined key terms, including 'drug', 'drug product', 'solution', 'cream', 'ointment', etc. However, we did not give a textual definition of every drug product represented by an NDC, because (1) there are hundreds of thousands of them and (2) there is insufficient historical information about the manufacturer, number of tablets in a bottle (the 50 tablet bottle and 100 tablet bottle of Tylenol have different NDCs), etc., to differentiate them.

We analyzed from all 4 perspectives the representations of molecules and formulations in RxNorm, the National Drug File Reference Terminology (NDF-RT), SNOMED CT, Chemical Entities of Biological Interest (ChEBI), and an OWL conversion of the Anatomical and Therapeutic Chemical (ATC) classification system. All 5 artifacts represent "ingredients", such as acetaminophen. The latter 4 classify ingredients and formulations by molecular mechanisms of action and therapeutic indications. We also reviewed DrugBank and PharmGKB. Finally, the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) makes terminological resources available for free (Overhage, 2012). We excluded several resources that were not freely available (Broekstra, 2004; Chute, 2003; Doulaverakis, 2012; Merrill, 2008; Senger, 2011).

2.2 Creation of a new ontology

Because no existing ontology was suitable, we created a new ontology: the Drug Ontology (DrOn). We reused Universal Resource Identifiers from ChEBI and the Protein Ontology (PRO). We developed an automated process to create in DrOn a class for each historical NDC and tablets, solutions, ointments, etc. and their active ingredients—represented in RxNorm.[1] But given the mistakes in RxNorm (see below), we did not reproduce all its relationships.

3 RESULTS

We first outline our analysis, which includes our definitions of key terms. We then present the results of our evaluation of existing artifacts. Finally we describe DrOn, created to meet our requirements and address the shortcomings of existing artifacts with respect to all four perspectives.

3.1 Realist analysis of the portion of reality

Anything that is a drug is a material entity as defined by Basic Formal Ontology (BFO). The vast majority of drugs are tablets, capsules, solutions, suspensions, creams, etc. Even when the "dose form" of a drug is given as 'inhaler', the inhaler itself is not the actual form: it is rather typically a solution that the inhaler delivers into the lungs as an aerosol: the "presentation" vs. "administration" form (Senger, 2011).

However, drugs are material entities with a special purpose: not every tablet or ointment is a drug. Drugs are used in medicine to diagnose, prevent, treat, and/or study disease.

Drug Role

Thus, drugs are material entities that bear a particular role. The Ontology of Biomedical Investigations (OBI) defines 'drug role' as *a role borne by a molecular entity and is realized in a process of absorption by an organism alters, or effects (or is assumed to effect) a function(s) which inhere in an organism* [sic]. ChEBI similarly defines 'drug role', stating that it *…always inheres in a small molecule, and as such is…the role played by the active ingredient in a pharmaceutical formulation, because drugs are complicated substantial entities containing not only the active ingredient…* (Batchelor, 2010) Thus per OBI and ChEBI, in a 20 milligram tablet of furosemide, absolutely every molecule of furosemide in the tablet is the bearer of its own unique drug role. Neither the tablet itself, nor the aggregate of furosemide molecules in it, has a drug role of its own per se.

However, a single molecule by itself cannot diagnose, prevent, study, or treat disease. In humans and other animals at least, it is instead the collective action of molecules that typically number on the order of $>10^{22}$.[2] Even ChEBI recognizes that: *the granularity of realization of a drug role is at the bulk level of granularity…* (Batchelor, 2010) However, how to recognize a "bulk realization" of a single molecule's drug role is not clear. Nor is it clear what such an entity is ontologically. In reality, the realizable entity (that ChEBI calls 'role') inheres in an aggregate, not a single molecule. Thus, at a minimum, a drug role as defined here inheres in the scattered aggregate of furosemide molecules.

Furthermore, the term 'inactive ingredient' implies that such entities have no function, which is untrue. A better term is 'excipient'. Excipients primarily serve the role of aiding delivery of the active ingredient into the organism. Tablets and capsules enable swallowing the active ingredient. Solvents such as saline solution enable intravenous injection of the active ingredient. Excipients also stabilize the active ingredient chemically, prevent inflammation at the administration site, and optimize absorption into the body.

Because the active ingredient(s) cannot treat, prevent, etc. disease, as intended, without being combined with excipients, it is the finished drug product that bears the drug role, as we define it. Here we use the term 'product' to refer to the output of a production process, not in a narrower commercial sense. The Federal Food, Drug and Cosmetic Act (FD&C Act) also defines 'drug product' this way: *a finished dosage form, for example, tablet, capsule, solution, etc., that contains an active drug ingredient generally…in association with inactive ingredients* (Food and Drugs, 2012). To differentiate the drug role of drug products from molecule-based drug role(s) in OBI/ChEBI, we call it 'clinical drug role'.

Our definitions thus are:

Clinical drug role: *the role of a material entity to prevent, diagnose, treat, or study disease and/or its effects*.

[1] All RxNorm-sourced content in the Unified Medical Language System is Category 0, meaning that derivative works are not prohibited.

[2] 20 milligrams of furosemide consists of ~3.64×10^{22} molecules.

FDA approved drug role: *a clinical drug role conferred by the United States Food and Drug Administration that permits the production, marketing, sale, prescribing, and consumption of its bearer in the United States.*

Drug product: *a material entity (1) containing at least one scattered molecular aggregate as part (the active ingredient) and (2) that is the bearer of a clinical drug role.*

"Forms" of drug product

As we noted above, drug products come in a variety of forms, which can differ from the intended or planned administration form (e.g., an effervescent tablet dissolved into solution and then drunk). In DrOn, we currently represent the presentation form (it always exists). Representing administration forms when different is future work.

Drug tablet: *a solid object typically of discoid, spheroid, or elliptic-cylindrical shape or approximations thereof, that bears a clinical drug role.*

Drug solution: *portion of solution that bears a clinical drug role.*

We similarly define 'drug suspension', 'drug colloid', 'drug cream', 'drug ointment', and 'drug lotion'. We define solution, suspension, etc. per chemistry and pharmacology:

Portion of pure substance: *an object all of whose parts that are atoms or molecules are of the same type (as determined by a unique structure).*

Portion of element: *a portion of pure substance all of whose atomic parts are of the same type (as determined by a unique structure).*

Portion of compound: *a portion of pure substance whose molecular parts are of the same type (as determined by a unique structure), but whose atomic parts are of different types (as determined by a unique structure).* The latter clause excludes diatomic molecules of oxygen, chlorine, etc.

Portion of mixture: *a material entity that contains two or more scattered object aggregates as its only parts, where the grains of each object aggregate are of different types and evenly distributed throughout, such that any two parts of the entity, each of which is spatially contiguous and of the same size, contain nearly equal numbers of grains of the aggregates*. Typically, the grains of the object aggregates are molecules (as in a solution of salt), but a mixture made of gravel mixed in a heap of soil also meets the definition: the grains are individual rocks (gravel) and particles of soil.

Portion of solution: *a portion of mixture of two or more portions of pure substances whereby one or more of its pure substance parts called solute(s) is (are) dispersed evenly in another pure substance part called the solvent, whereby (1) the grains of the solvent and solute(s) are of size $\leq 10^{-9}$m, and (2) the solvent typically has a much greater mass and volume than the solute(s).* The phase (solid, liquid, gas) of the solution is usually the phase of the solvent.

Portion of suspension*: a portion of mixture whereby the grains of at least one object aggregate are ≥1 μm in size and the grains of at least one other object aggregate are $\leq 10^{-9}$m in size.* The two aggregates can be separated by gravitational settling of the grains ≥1 μm. Example: portion of blood (red blood cells and other large grains will settle out in a static portion of blood).

Portion of colloid: *a portion of mixture whereby the grains of at least one object aggregate are $>10^{-9}$m and <1 μm in size (the dispersed material) and the grains of at least one other object aggregate are $<10^{-9}$m in size (continuous medium).* The aggregates cannot be separated by gravitational settling. Note that in certain cases the grains are relatively large contiguous portions of liquid or gas dispersed in a medium.

Portion of emulsion: *a portion of colloid where both the dispersed material and the continuous medium are liquids*.

Portion of cream: *a portion of emulsion of oil and water in nearly equal proportions with high viscosity.*

Portion of lotion: *a portion of emulsion of oil and water in nearly equal proportions with low viscosity.*

Portion of ointment: *a portion of emulsion of oil and water where the proportion of oil is significantly greater than proportion of water, typically 80% oil and 20% water*.

Portion of gel: *a portion of colloid whereby a liquid is dispersed in a solid that is either gelatin, jelly, or agar.*

Molecular disposition

Although the clinical drug role inheres in the drug product, each individual molecule of the active ingredient nevertheless bears a certain biological disposition. For example, each metoprolol molecule bears a disposition to bind to a beta-1 adrenergic receptor. Note that the tablet does not bear this disposition because it does not bind to receptors. When a metoprolol molecule binds (thereby realizing the disposition), it does not activate the receptor (as would an epinephrine molecule). Furthermore, once it binds, it prevents the binding of epinephrine. Thus, a metoprolol molecule has a *function-inhibiting beta-1 adrenergic receptor-binding disposition*. Similarly, each furosemide molecule bears a disposition to bind to sodium-potassium-chloride (Na-K-Cl) cotransporter 2 protein (NKCC2) such that NKCC2 cannot function. Thus we say *furosemide* **bearer_of** *function-inhibiting NKCC2 binding disposition*.

In the current version of DrOn, we include these two dispositions plus (1) function-inhibiting hydrogen/potassium adenosine triphosphatase enzyme (H^+/K^+ ATPase) binding disposition (a.k.a. proton pump inhibition), (2) function-inhibiting L-type voltage-gated calcium channel binding disposition (a.k.a calcium channel blocking), (3) function-inhibiting vitamin K epoxide reductase binding disposition (a.k.a. vitamin K antagonizing).

The dispositions are function inhibiting because their realization is the suppression of the functioning (i.e., suppressed realization of the function) of the proteins, receptors, enzymes, etc. that they bind. Some dispositions, by contrast,

are function inducing. For example, carbamazepine induces the activity of cytochrome P450 2C19 (CYP2C19) enzyme.

Therapeutic potentiality

When drugs are given to treat a disease, its symptoms (angina), or its other effects (such as fever), it is common to classify them as antihypertensives, antianginals, antipyretics, etc. A particular drug product may have multiple potentialities in this regard. For example, metoprolol 50 mg tablet might be used as an antihypertensive in one patient, an antianginal in another patient, an antiarrhythmic in another patient, and to treat myocardial infarction (MI) in another.

The entire finished drug product has these potentialities, not just the active ingredient. For example, timolol ophthalmic solution can treat glaucoma but timolol tablets cannot. Thus, depending on the presentation and/or administration form, certain therapeutic potentialities exist or not. This situation further argues against assigning therapeutic drug roles to molecules. None of the timolol molecules in a tablet bear a "role" (or other realizable entity) to treat glaucoma.

Furthermore, the quantity of active ingredient also determines therapeutic potentiality. For example, finasteride is used in a dose of 5 mg per day for benign prostatic hypertrophy (BPH) but 1 mg per day for androgenetic alopecia.

Thus, the finished drug product combined with the therapeutic intent of the physician prescribing the drug ultimately determines its therapeutic use or role. A patient taking metoprolol for control of abnormal heart rhythms, but passes out due to hypotension, has hypotension as an adverse reaction. Whereas a patient taking metoprolol as an antihypertensive, but whose athletic performance is inhibited because she cannot generate an adequate heart rate, has bradycardia as an adverse reaction.

The therapeutic potentialities of drug products are at their essence dispositions. Drug products have these potentialities as a result of their physical makeup, and losing them would necessitate a change in their physical makeup. The BFO definition of disposition is thus applicable here. Because drugs are often used to block the realization of dispositions (including diseases as dispositions per the Ontology for General Medical Science), these therapeutic dispositions are often blocking dispositions (Goldfain, 2011).

3.2 Review of existing artifacts

Analysis from the technical perspective

Only RxNorm has a freely publicly available set of historical NDCs; thus none of the other artifacts met this requirement. However, RxNorm does not maintain all historical NDCs in its current version. Thus, we had to process all historical versions of RxNorm beginning with June 2008 (1st one with NDCs). This process was complicated by the fact that we also had to historically trace the "concept unique identifiers" (or RXCUIs) to which NDCs are attached. RXCUIs are often retired, and it can be difficult to uncover their history in older versions of RxNorm.

Only NDF-RT and ChEBI are available as an OWL artifact from their respective developers. None of RxNorm, SNOMED CT, ATC, DrugBank, PharmGKB, OMOP CDM is available as OWL. We did find a third-party conversion of ATC into OWL (Croset, 2012).

Analysis from the perspective of pharmacology

NDF-RT makes incorrect assertions from the perspective of pharmacology and medical science. For example, it incorrectly asserts for timolol oral tablet that it "may treat" glaucoma, something true only for the ophthalmic form. NDF-RT also incorrectly asserts that vancomycin capsules "may treat" bacterial endocarditis and pneumococcal meningitis. However, vancomycin is not absorbed through the gastrointestinal tract and if taken orally cannot treat these diseases: intravenous (IV) vancomycin must be used. Similarly, NDF-RT incorrectly states that IV vancomycin "may treat" pseudomembranous colitis, which true only when the IV form is administered into the GI tract (not intravenously).

SNOMED CT makes an incorrect assertion about timolol ophthalmic solution, but through a series of **is_a** relations instead of "may treat". Specifically, it asserts *ophthalmic form timolol* **is_a** *timolol* **is_a** *non-selective beta blocking agent* **is_a** *beta-Blocking agent* **is_a** *Hypotensive agent*. By transitivity then, ophthalmic timolol is a hypotensive (antihypertensive) agent. SNOMED CT also erroneously asserts through a series of **is_a** relations that furosemide and 9 other diuretics are antimycobacterial agents. Furosemide's structure includes a sulfonamide group, through which it is related to some antimicrobials. But it is not an antimicrobial.

These incorrect assertions in NDF-RT and SNOMED CT are not comprehensive. We did not manually review their entirety for accuracy. However, these examples were easy to find. The ontological analysis below identifies the source of such errors and suggests they may be systemic.

ChEBI, as we have seen, incorrectly assigns therapeutic dispositions to individual molecules (e.g., antibiotic to vancomycin) and represents molecular dispositions as roles.

ATC in OWL makes incorrect assertions such as *cyclophosphamide* is_a *anti-neoplastic agent*, which is incorrect because some instances of cyclophosphamide treat autoimmune disorders. As stated above, finasteride treats alopecia as well as BPH, but ATC says all finasteride molecules are instances of *Drugs used in benign prostatic hypertrophy*. These errors stem from ontologization of a classification.

The semantics of DrugBank and PharmGKB are not explict, and thus whether they attribute properties of drug products to molecules is uncertain.

Analysis from the perspective of OWL DL semantics

This analysis applies only to NDF-RT, ChEBI, and ATC in OWL (nothing else was in OWL). In NDF-RT, the "may treat" relation is problematic in the same manner as the **disease_may_have_finding** relation in the NCI Thesaurus (Schulz, et al., 2010). Namely, the assertion that *vancomycin*

125 mg oral capsule **may_treat** SOME *pseudomembranous colitis* implies that for every vancomycin capsule in existence, there also exists an actual instance of pseudomembranous colitis, to which it is related by **may_treat**. Clearly, this assertion is false. There are 48,241 asserted **may_treat** relations in the NDF-RT OWL file (1/14/13) that are therefore incorrect from the perspective of OWL-DL semantics.

Boeker et al. analyzed the DL semantics of existentially quantified relations in OBO ontologies, including ChEBI (Boeker, et al., 2011). They found that ~62% of existentially quantified clauses in ChEBI are incorrect (based on a small sample). The most problematic relations were structural relations such as **is_tautomer_of** and **has_parent_hydride**. However, they also found that role assertions in ChEBI were problematic: namely, asserting that each molecule bears the role of, e.g. anti-ulcer drug, leads to problems when it is used for a different purpose.

ATC in OWL makes no existential restrictions.

Analysis from the perspective of our ontological analysis

NDF-RT, SNOMED CT, ChEBI, OBI, and ATC in OWL all make the fundamental mistake of assigning properties of drug products to molecules of a particular type. In particular, NDF-RT assigns to *Vancomycin* the **may_treat** relation to *pseudomembranous colitis* and *bacterial endocarditis*. Thus, classes for oral and intravenous vancomycin, which are descendants of *Vancomycin*, inherit the **may_treat** relation to the two diseases. SNOMED CT makes similar confusions, whereby both timolol (substance) and timolol (product) are types of *Hypotensive agent*. ATC in OWL also asserts numerous incorrect **is_a** relations.

General ontological issues

RxNorm makes use-mention mistakes via its **tradename_of** relation. For example, it asserts *Vicodin oral tablet* **tradename_of** *acetaminophen 300 mg / hydrocodone 5mg tablet*. However, *Vicodin* tablets are not names, let alone tradenames of acetaminophen/hydrocodone tablets.

ATC makes numerous ontological and terminological errors of the same nature as other classifications, such as the International Classification of Diseases. It contains dozens of classes whose terms begin with 'other'; it has "is_a overloading"; it has hierarchical codes; it contains redundancies.

3.3 The Drug Ontology (DrOn)

Because existing artifacts failed to meet numerous requirements and contained systematic factual and ontological errors, we constructed DrOn. It is an OWL 2.0 artifact, with a manually-curated upper layer (including terms defined here) and automatically-created layers based on RxNorm (but on only RxNorm content curated by the NLM (source abbreviation of RXNORM). We created classes in DrOn for each ingredient or IN (*furosemide*), semantic clinical drug form or SCDF (*furosemide oral tablet*), semantic clinical drug or SCD (*furosemide 20 mg oral tablet*), and semantic branded drug or SBD (*Lasix 20 mg oral tablet*) term in the Feb 2013 version of RxNorm. Ontologically, the correct representation is *Lasix 20 mg oral tablet* **is_a** *furosemide 20 mg oral tablet* **is_a** *furosemide oral tablet* (SBD **is_a** SCD **is_a** SCDF). We also created the ingredient relationship as *furosemide oral tablet* **has_proper_part** SOME *(scattered molecular aggregate* AND **has_grain** SOME *furosemide*). We mapped as many IN terms to ChEBI URIs as we could.

Because an NDC typically represents a packaged form of multiple tablets, vial of intravenous solution, tube of ointment, etc. it is the case that the tablet, portion of solution/ ointment/etc. is a proper part of the packaged product. Thus, we created a class for each NDC, and related it to the tablet, portion of solution, etc. using the **has_proper_part** relation.

The Feb 2013 version of RxNorm has 188,716 RxNorm-curated NDCs, whereas our historical processing extracted 394,830 such NDCs. The current version of RxNorm has <½ the required NDCs (47.8%). Of the 394,830 NDCs, there were 6,644 NDCs for which it was difficult to determine the RXCUI to which it should be related. Of those, we found an RXCUI for 4,475. Thus, the remaining 2,169 NDCs (0.55%) are not in DrOn.

We validated DrOn by comparing our results from a query on acetaminophen (as active ingredient) with the list of NDCs manually curated by Kelkar et al. (Kelkar, 2012a, 2012b) Only one NDC was not in DrOn. However, it was a 4 digit number, not the 11-digits expected of the NDCs in LifeLink and RxNorm. Thus, it is not an NDC at all and we matched all manually-curated acetaminophen NDCs. The successful execution of this query required correct relations in DrOn as well as the mere presence of NDCs.

3.4 An Application that Utilizes the Ontology

We developed an application that utilizes DrOn to query NDCs that represent drug products with various characteristics. It is available at: http://ingarden.uams.edu/ingredients/. Currently, it allows searching for NDCs with particular active ingredients or that have active ingredients with one of six dispositions (we split calcium channel blockade into L- vs. T-type channels: ethosuximide blocks only the latter).

4 DISCUSSION

We developed an ontology—the Drug Ontology (DrOn)—and a software application that enable the query of drug products (represented by United States NDCs) based on their active ingredient(s) and their molecular and therapeutic dispositions. We found existing artifacts insufficient on multiple levels: technical, scientific, description logic semantics, and ontological.

In the process, we generated textual definitions of numerous terms (also captured in DrOn), including 'clinical drug role'. Despite being driven by the need to query NDCs, thousands of classes in DrOn apply to drug products sold in other nations (e.g., *aspirin 325 mg oral tablet*).

A key result of this work is that a frequent cause of scientific incorrectness in pre-existing artifacts was inadequate ontological analysis. A common cause of error was attributing properties of drug products to individual molecules. The fact that ontological realism avoided scientific inaccuracy in this work is a novel finding. However, because ChEBI is also realism based, its shortcomings require explanation. We believe its focus on molecules and their usage led to insufficient attention to drug products and their composition and usage. But lack of coverage of tablets/creams/ etc. is not a flaw in ChEBI: it does not claim such coverage.

Future work includes (1) adding more molecular and therapeutic dispositions to DrOn, (2) implementing monthly maintenance to update the ontology with new NDCs (we are exploring Structured Product Labels for this purpose), (3) adding other molecular dispositions such as binding to enzymes of the Cytochrome P450 system, and (4) an ontological analysis of physiological effects. With respect to the latter, for example, furosemide causes diuresis, an effect downstream of furosemide's direct NKCC2 inhibition.

ACKNOWLEDGEMENTS

This work was supported by award UL1TR000039 (National Center for Advancing Translational Sciences), award R01GM101151 (National Institute for General Medical Science), and the Arkansas Biosciences Institute. Our views do not represent those of NCATS, NIGMS, NIH, or ABI.

REFERENCES

Batchelor, C., Hastings, J., & Steinbeck, C. (2010). Ontological dependence, dispositions and institutional reality in chemistry. Paper presented at FOIS 2010, Toronto, Canada.

Boeker, M., Tudose, I., Hastings, J., Schober, D., & Schulz, S. (2011). Unintended consequences of existential quantifications in biomedical ontologies. BMC Bioinformatics, 12, 456.

Broekstra, J., et al. (2004). The Drug Ontology Project for Elsevier. Proceedings of the WWW'04 workshop on Application Design, Development and Implementation Issues in the Semantic Web.

Broverman, C., Kapusnik-Uner, J., Shalaby, J., & Sperzel, D. (1998). A concept-based medication vocabulary: an essential requirement for pharmacy decision support. Pharmacy practice management quarterly, 18(1), 1-20.

Chute, C. G., Carter, J. S., Tuttle, M. S., Haber, M., & Brown, S. H. (2003). Integrating pharmacokinetics knowledge into a drug ontology: as an extension to support pharmacogenomics. AMIA Annual Symposium proceedings, 170-174.

Croset, S., Hoehndorf, R., & Rebholz-Schuhmann, D. (2012). OWL Representation of the Anatomical Therapeutic Chemical Classification System and Mapping to Drug Bank. Paper presented at OBML 2012, Dresden, Germany.

Doulaverakis, C., Nikolaidis, G., Kleontas, A., & Kompatsiaris, I. (2012). GalenOWL: Ontology-based drug recommendations discovery. Journal of biomedical semantics, 3(1), 14.

Food and Drugs, 21 CFR C.F.R. § 210.3 (2012).

Goldfain, A., Smith, B., & Cowell, L. G. (2011). Towards an ontological representation of resistance: The case of MRSA. Journal of Biomedical Informatics, 44(1), 35-41.

Kelkar, M., et al. (2012a). Acute and chronic acetaminophen use and renal disease: a case-control study using pharmacy and medical claims. Journal of managed care pharmacy, 18, 234-46.

Kelkar, M., et al. (2012b). Prescription-acquired acetaminophen use and the risk of asthma in adults: a case-control study. The Annals of pharmacotherapy, 46(12), 1598-1608.

Kim, J. M., & Frosdick, P. (2001). Description of a drug hierarchy in a concept-based reference terminology. Proc AMIA Symp, 314-8.

Merrill, G. H., Ryan, P. B., & Painter, J. L. (2008). Construction and annotation of a UMLS/SNOMED-based drug ontology for observational pharmacovigilance. Presented at Intelligent Data Analysis for bioMedicine and Pharmacology, Washington DC.

Nelson, S. J., Zeng, K., Kilbourne, J., Powell, T., & Moore, R. (2011). Normalized names for clinical drugs: RxNorm at 6 years. J Am Med Inform Assoc, 18(4), 441-448.

Olsen, L., Grossman, C., & McGinnis, J. M. (2011). Learning What Works: Infrastructure Required for Comparative Effectiveness Research: Workshop Summary: The National Academies Press.

Overhage, J. M., Ryan, P. B., Reich, C. G., Hartzema, A. G., & Stang, P. E. (2012). Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc, 19(1), 54-60.

Palchuk, M. B., Klumpenaar, M., Jatkar, T., Zottola, R. J., Adams, W. G., & Abend, A. H. (2010). Enabling Hierarchical View of RxNorm with NDF-RT Drug Classes. AMIA Annual Symposium proceedings, 2010, 577-581.

Parrish, F., Do, N., Bouhaddou, O., & Warnekar, P. (2006). Implementation of RxNorm as a Terminology Mediation Standard for Exchanging Pharmacy Medication between Federal Agencies. AMIA Annu Symp Proc, 1057.

Pathak, J., & Chute, C. G. (2011). Further revamping VA's NDF-RT drug terminology for clinical research. Journal of the American Medical Informatics Association, 18(3), 347-348.

Saunders, G., Ivkovic, S., Ghosh, R., & Yearwood, J. (2005). Applying Anatomic Therapeutic Chemical (ATC) and critical term ontologies to Australian drug safety data for association rules and adverse event signaling. Paper presented at the Advances in Ontologies 2005, Adelaide, Australia.

Schulz, S., Schober, D., Tudose, I., & Stenzhorn, H. (2010). The Pitfalls of Thesaurus Ontologization - the Case of the NCI Thesaurus. AMIA Annu Symp Proc, 2010, 727-731.

Senger, C., Seidling, H. M., Quinzler, R., Leser, U., & Haefeli, W. E. (2011). Design and evaluation of an ontology-based drug application database. Methods Inf Med, 50(3), 273-284.

Smith, B., & Brochhausen, M. (2010). Putting biomedical ontologies to work. Methods Inf Med, 49(2), 135-140.

Smith, B., & Ceusters, W. (2010). Ontological realism: A methodology for coordinated evolution of scientific ontologies. [10.3233/AO-2010-0079]. Applied Ontology, 5(3), 139-188.

Sperzel, W. D., Broverman, C. A., Kapusnik-Uner, J. E., & Schlesinger, J. M. (1998). The need for a concept-based medication vocabulary as an enabling infrastructure in health informatics. Proceedings AMIA Annual Symposium. 865-869.