OPINION

Needs for an Expanded Ontology-Based Classification of Adverse Drug Reactions and Related Mechanisms

PE Zhichkin[1], BD Athey[2], MI Avigan[3] and DR Abernethy[1]

The growing significance of bioinformatics and systems biology in drug safety research requires a system of adverse-event classification that goes beyond a simple vocabulary. This opinion piece outlines the need for development of an ontology-based framework of describing adverse drug reactions (ADRs) and describes the potential applications for such a framework.

The standard classification for ADRs is the Medical Dictionary for Regulatory Activities (MedDRA) terminology, developed under the auspices of the International Conference on Harmonisation. Among other users, regulatory authorities worldwide, including the US Food and Drug Administration, employ it in the evaluation of clinical trials and postmarketing surveillance reports. With more than 90,000 terms structured in five levels of hierarchy, the MedDRA provides both granularity of detail and the higher level descriptions by grouping adverse effects into classes.[1]

This has served the public health well; however, we believe that, to continue to advance the science of drug safety, ADR classifications must evolve beyond relatively simple vocabularies and toward being the knowledge framework for a systematic organization of all ADR-related data and information. This capability promises to enable new discoveries, inform researchers and regulators, and create new biosurveillance capabilities.

Our proposed ontology-integration approach offers a model of such organization and is consistent with the approach recommended in the recent National Research Council report on the new taxonomy of disease.[2] Distinctive characteristics of an ontology are logical consistency of definitions of classes in the biological and clinical hierarchies and rigorous formal rules of organization, in which links between terms are usually specified using OWL Web Ontology Language and the Protégé authoring tool (http://protege.stanford.edu). A key underlying principle is that the terms composing an ontology have properties (or attributes) defined through, or compatible with, other ontologies from other related areas of knowledge.[3] This enables the creation of an overlapping network of terms and definitions between different specialized areas that would considerably simplify computational treatment of complex multilevel interdisciplinary problems. The prime example of such a unifying ontology is Gene Ontology, which has become indispensable in many areas of bioinformatics.[4]

Similarly, an ontological organization of the ADR-related knowledge can lead to the creation of a sharable infrastructure enabling and expanding the applications of bioinformatics, next-generation sequencing, the "-omics," and systems biology, to drug safety. As systems biology extends its reach to enable predicting and explaining of ADRs,[5] the need for an ADR ontology becomes apparent. Particularly useful will be an ontology providing the framework for linking ADRs with biological mechanisms and function, establishing correlations between them, and allowing statistically meaningful inference from one to the other.

Considering the established role of the MedDRA, the adverse reactions and mechanisms (ARM) ontology will be founded on MedDRA terms and consistent with MedDRA general classifications whenever possible. Some of the MedDRA terms (e.g., the codes for medical procedures and for medical/social/family background and circumstances) can be omitted or de-emphasized. Concentrating on medical terms that are used in clinical trials and medical practice will reduce the number of terms and allow for adding new attributes and expanding the ontology.

[1]Office of Clinical Pharmacology, Office of Translational Sciences, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; [2]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, USA; [3]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA. Correspondence: DR Abernethy (darrell.abernethy@fda.hhs.gov)

To support clear logical links to biological mechanisms, the hierarchies in the ARM ontology have to be rooted in the underlying biology and physiology of adverse events. This may cause some classes and groupings of ARM ontology to differ from those in the MedDRA because the MedDRA uses anatomy as its primary organizing principle.[1] Future research and development can extend this to physiology.

In addition to providing standardized terms and classification hierarchy, an important function of an ontology is to supply the consensus definitions of the terms, leading to a meaningful set of Common Data Elements (CDEs). These standardized terms and definitions will be based on openly available medical dictionaries. For example, many ADR definitions can be automatically imported from the National Library of Medicine (NLM) Unified Medical Language System (UMLS) and its Medical Subject Heading vocabulary (MeSH), provided in a machine-readable format by the NLM. The decreased ambiguity of classification will reduce misclassification errors and should improve the precision of the ARM ontology-based applications.

It is a standard pharmacovigilance practice to group different terms related to the same pathological phenomenon independently from MedDRA hierarchies. These MedDRA terms groupings, called Standardized Medical Queries (SMQs), are routinely used for mining the ADR databases in search of events marked by medical conditions with multiple possible manifestations.[1] The ARM ontology, being biologically based, will naturally incorporate many of the groupings defined by SMQs and is likely to make the construction of future SMQs a more logical and streamlined process. Perhaps more importantly, in the future it is expected that the ARM would enable more powerful computationally adaptive *ad hoc* querying strategies.

Another potential application of ARM ontology is an automatic expansion of ADR searches to progressively broader concepts, with the desired level of commonality defined by the user by specifying the level of hierarchy for the search.

A prototype of such query expansions can be seen in the way PubMed, by default, expands simple searches by using related MeSH synonyms.

As with any ontology, significant utility of the ARM ontology will stem from its integration with other medical classification systems so that the data gathered from different databases can be readily imported and seamlessly used together. The ARM ontology terms will need to map to the standard medical diagnostic terms (International Classification of Diseases, ICD), to terms used in literature searches (MeSH), and to electronic medical records using the Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT). The mapping of MedDRA to ICD and SNOMED-CT performed at the National Center for Biomedical Ontology Bioportal (http://bioportal.bioontology.org/mappings) considerably simplifies this task.

Carrying out this alignment of classifications will require employing another feature of ontologies: their ability to store additional information in the term-to-term links between ontologies. These interontological links are specified by assigning terms from one ontology to be additional attributes of a term in the second ontology. The most obvious attributes to assign to an ADR term are the name of drugs causing a particular ADR, ADR frequencies for each drug and the molecular mechanisms responsible for the ADR. In this manner, a system integrating relevant fields can be designed around the evolving ARM ontology.

As noted above, the connections between terms in the ARM ontology will be encoded using OWL. This is very important for consistency and interoperability, as most other ontologies of biological pathways and mechanisms such as Gene Ontology also use OWL or a compatible language. This should allow structured computational-based reasoning services that can be programmed, not only about ADRs and their links with each other but also about ADR connections with outside terms. The reasoning software may ultimately be capable of semiautomatically inferring the set of ADRs from the

drug mechanism of action, thus laying one of the foundational elements for the advancement of predictive toxicology.

The ARM ontology is likely to further serve as a template for building classifications in adjacent areas of biomedical science. For example, drugs can be classified based on the classes of ADRs they cause; genes and biological mechanisms can be grouped based on the classes of ADRs for which they are responsible. As such classifications are being developed, they will make more apparent the relationships between drugs, their targets, biological pathways they alter, and ADRs they cause, making the details of these interactions visible to various communities of users from basic researchers to regulators.

A variety of models, from simple path tracing to Bayesian networks, can be built on top of this foundation. If such a network of drugs, biological pathways, and corresponding ADRs is further linked to patients' genetic and genomic profiles *and* pretreatment phenotype data, a future goal will be to quantify the contributions of drug molecular targets and an individual patient's characteristics to the likelihood of an ADR in individuals or patient subgroups.

We believe the development of the ARM ontology with an integrated biosystems approach will prove invaluable in the analysis of a variety of important regulatory questions. Such a capability, as it is developed and gains acceptance, may also eventually augment current methods used in the identification and evaluation of drug-related risk. This concept has yet to be rigorously and formally tested, and its areas of applicability and limitations have to be more fully defined. Moreover, the drawing of conclusions from purely bioinformatics-based approaches, without a careful understanding and analysis of the range and quality of individual ADR cases, may be fraught with hazard.

It is clear that ontology-based search strategies will need to be complemented by an analysis of cases for appropriate case inclusion or exclusion, and would employ uniformly accepted case definitions. This is particularly important in the assessment of ADRs that are

clinically serious, because these events are most likely to have the strongest impact in gauging overall drug-related risk. Some essential information about this risk can only be gathered from an assessment of single ADR cases to characterize the range of clinical and laboratory findings on presentation, the clinical course and complications that may ensue, and the statistical likelihood of drug-induced causality.

In conclusion, with the advances in the sciences of bioinformatics and systems analysis in medicine, the foundation for the ontology-enabled integration of drug safety with molecular biology data is quickly becoming possible. This is needed to keep drug safety science abreast of the evolving science in the wider biomedical community. We expect such a foundation built on the principles of ontology integration and services extension to enable multiple applications, ranging from more powerful search strategies in post-marketing safety to ADR predictions.

CONFLICT OF INTEREST
The authors declared no conflict of interest.

1. Maintenance and Support Services Organization. Introductory Guide, MedDRA version 14.1 (MSSO-DI-6003-14.1.0) <http://meddramsso.com/files_acrobat/intguide_14_1_English_update.pdf> (September 2011).
2. National Research Council. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease (National Academies Press, Washington, DC, 2011).
3. Shah, N. & Smith, B. Ontologies for biomedicine—how to make and use them. Proceedings of 15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB/ECCB), Vienna, 19–25 July <http://www.bioontology.org/wiki/images/d/d2/ISMB_2007_Handout.pdf> (2007).
4. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
5. Berger, S.I. & Iyengar, R. Role of systems pharmacology in understanding drug adverse events. *Wiley Interdisc. Rev. Syst. Biol. Med.* **3**, 129–135 (2011).

Healthy Volunteer Registries and Ethical Research Principles

EA Kupetsky-Rincon[1] and WK Kraft[1]

The dual enrolling of phase I volunteers is a potential risk to subjects. It can also distort study results, threaten study validity, and possibly cause harm to future patients. Existing subject registries differ in structure, funding, and governance. Although the choice of the ideal system is driven by the scope of the risk and the funding mechanism, and is ultimately a value judgment of freedom versus paternalism, none of the registries significantly impinges on the tenets of ethically based research.

The Belmont report, issued in 1978 by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, identified the key pillars of ethical research to be justice, autonomy, and beneficence. A key principle is that human-subject research has a responsibility to minimize harm and maximize benefit for participants as long as there is acceptable equipoise. There is, however, no absolute requirement of potential benefit of participation, even for those with disease. For example, although oncologists and patients participate in phase I oncology trials with a primary hope for therapeutic response, drug efficacy is not necessarily an immediate goal of these studies. The lack of understanding of this distinction by patients has been well described. Other study designs, such as those of noninferiority or comparative effectiveness, do not provide patients with a direct benefit of participation beyond access to care and/or financial compensation. Healthy-volunteer studies entail risk, with no potential for therapeutic benefit to participants. The lack of any potential health benefit outside of an evaluation of health status has often led to heightened institutional review board scrutiny for phase I studies. The focus of regula-

tion in healthy-volunteer clinical trials is typically the short-term protection of subjects from harm directly related to study procedures. Outside of cumulative limits on radiation exposure, the role of the subject outside of an individual trial is generally not considered. The National Institutes of Health (NIH) and the US Food and Drug Administration (FDA) do not strictly limit the number of studies in which a volunteer can participate. It is merely suggested that subjects not enroll in consecutive studies without adequate time for washout of drug or intervention based on the biology of the system. However, there have recently been concerns about the potential for phase I volunteer participants to enroll in multiple clinical trials concurrently, with calls for a mandatory registry to track subjects.[1]

Motivations for healthy-volunteer participants in clinical research can be altruistic, especially for disease-specific activists or those with afflicted family members. Generally, however, the prime motivation for most phase I trial enrollees who lack an underlying disease is in the financial compensation for participation.[2,3] Pursuit of compensation can incentivize subjects to enroll in multiple studies, despite the potential for personal injury, or risk of discovery and loss

[1]Department of Pharmacology and Experimental Therapeutics, Thomas Jefferson University, Philadelphia, Pennsylvania, USA. Correspondence: WK Kraft (Walter.Kraft@jefferson.edu)