Published in final edited form as:

Drug Saf. 2016 July ; 39(7): 697–707. doi:10.1007/s40264-016-0414-0.

Linking MedDRA®-coded Clinical Phenotypes to Biological Mechanisms by The Ontology of Adverse Events: A pilot study on Tyrosine Kinase Inhibitors (TKIs)

Sirarat Sarntivijai^{1,6,*}, Shelley Zhang², Desikan G. Jagannathan³, Shadia Zaman¹, Keith K. Burkhart¹, Gilbert S. Omenn^{5,6,7}, Yongqun He^{4,6,8}, Brian D. Athey^{6,7,9}, and Darrell R. Abernethy¹

¹ Office of Clinical Pharmacology, Office of Translational Science, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD 20904, USA

² College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI 48109, USA

³ College of Engineering, University of Michigan, Ann Arbor, MI 48109, USA

⁴ Unit of Laboratory Animal Medicine, University of Michigan, Ann Arbor, MI 48109, USA

⁵ Department of Internal Medicine and Human Genetics and School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA

⁶ Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

⁷ Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

⁸ Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA

⁹ Psychiatry Department, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

Introduction—A translational bioinformatics challenge lies in connecting population and individual's clinical phenotypes in various formats to biological mechanisms. The Medical Dictionary for Regulatory Activities (MedDRA®) is the default dictionary for Adverse Event (AE) reporting in the FDA Adverse Event Reporting System (FAERS). The Ontology of Adverse Events (OAE) represents AEs as pathological processes occurring after drug exposures.

Objectives—The aim is to establish a semantic framework to link biological mechanisms to phenotypes of AEs by combining OAE with MedDRA® in FAERS data analysis. We investigated the AEs associated with Tyrosine Kinase Inhibitors (TKIs) and monoclonal antibodies (mAbs) targeting tyrosine kinases. The selected 5 TKIs/mAbs (i.e., dasatinib, imatinib, lapatinib,

* Corresponding author – siiraa@umich.edu.

Compliance with Ethical Standards

Conflicts of Interest: Sirarat Sarntivijai, Shelley Zhang, Desikan Jagannathan, Shadia Zaman, Keith Burkhart, Gilbert Omenn, Yongqun He, Brian Athey, and Darrell Abernethy have no conflicts of interest that are directly relevant to the content of this study.

cetuximab, and trastuzumab) are known to induce impaired ventricular function (non-QT) cardiotoxicity.

Results—Statistical analysis of FAERS data identified 1,053 distinct MedDRA[®] terms significantly associated with TKIs/mAbs, where 884 did not have corresponding OAE terms. We manually annotated these terms, added them to OAE by the standard OAE development strategy, and mapped them to MedDRA[®]. The data integration to provide insights into molecular mechanisms for drug-associated AEs is performed by including linkages in OAE for all related AE terms to MedDRA[®] and existing ontologies including Human Phenotype Ontology (HP), Uber Anatomy Ontology (UBERON), and Gene Ontology (GO). Sixteen AEs are shared by all 5 TKIs/mAbs, and each of 17 cardiotoxicity AEs was associated with at least one TKI/mAb. As an example, we analyzed ‘cardiac failure’ using the relations established in OAE with other ontologies, and demonstrated that one of the biological processes associated with cardiac failure maps to the genes associated with heart contraction.

Conclusion—By expanding existing OAE ontological design, our TKI use case demonstrates that the combination of OAE and MedDRA[®] provides a semantic framework to link clinical phenotypes of adverse drug events to biological mechanisms.

1. Introduction

Advances in systems pharmacology have presented a growing challenge and opportunity to utilize complex bioinformatics and systems biology methods for pharmacovigilance research. The solution requires the development of an ontology-based framework that describes clinical (adverse drug) events, that goes beyond a simple classification of adverse events (AEs) (1). The Medical Dictionary for Regulatory Activities (MedDRA[®]) is used in standard AE reporting in the United States, Europe, and Japan (2). The usage of MedDRA has greatly facilitated AE report standardization and cross-report data analysis. There are a number of other ontologies and standard vocabularies that attempt to capture clinical phenotype for AEs and clinical disease including the Mammalian Phenotype Ontology (MP) (3), Systematized Nomenclature of Medicine – Clinical Term (SNOMED-CT) (4), the International Classification of Diseases (ICD) (5), and the Common Terminology Criteria for Adverse Events (CTCAE) (6, 7). However, MedDRA and the other data structures do not provide links to other information needed to explore and understand biological mechanism associated with the AEs.

The Ontology of Adverse Events (OAE), previously known as Adverse Event Ontology – AEO (8, 9), is a community-based open source ontology that supports the representation and analysis of clinical adverse event phenotypes (10). In the OAE, an adverse event is considered a pathological bodily process (a class defined as an *occurent* in the Basic Formal Ontology (BFO) (11)) and is described as an unpleasant or unexpected clinical phenotype that occurs after a medical treatment. This results in an ontological definition of phenotype that differs from other ontologies describing phenotypes or disorders as a disposition (BFO’s *continuant* (11)) such as those of the Human Phenotype Ontology (11, 12), Mammalian Phenotype Ontology (13), Disease Ontology (14), the Ontology for General Medical Science (15), or the Experimental Factor Ontology (16). OAE captures other related components such as a medical intervention (e.g. drug exposure) that are temporally associated with the

occurrence of an adverse event. An adverse drug event is defined as a pathological bodily process that is preceded by a drug exposure. An adverse event (AE) may also resemble a disease process, but it needs to be defined and distinguished from a normal biological process. However, it is important to note that this temporal association does not imply causation.

The application of OAE has been previously demonstrated in a bioinformatics analysis of influenza vaccine adverse events (17), temporal analysis of vaccine AEs (18), representation of genetic susceptibility to vaccine AEs (19), and ontological representation of AEs of licensed vaccines recorded in FDA drug label documents (20).

The OAE class organization links an AE to a biological process(es) potentially associated with the AE, and the biological process can be further linked to other contents such as the anatomical location of the adverse process. Therefore, OAE semantically captures the relationships of AEs to the medical intervention leading to the occurrence of AEs, and provides a possible framework to link clinically reported drug AEs to the underlying mechanisms (10). The expansion and further development of OAE described here allows it to be used to explore mechanisms for drug-associated AEs as defined by MedDRA[®] terms. It is necessary to use the MedDRA[®] dictionary because it is the standard terminology for AE reporting. However, there has not been a clear and specific use case published to demonstrate the feasibility of such a framework for linking clinically reported MedDRA AE terms to biological mechanisms of clinical phenotypes.

This study aims to demonstrate the use of the ontological framework provided by the OAE to link and hypothesize the linkage of MedDRA[®]-coded clinical phenotypes to underlying biological mechanisms. Our use case explored a drug exposure -AE pair, impaired contractility cardiotoxicity associated with the inhibition of tyrosine kinases by TKIs and mAbs. Although TKIs and mAbs mechanisms of action are slightly different, both work to inhibit the downstream activities in the tyrosine kinase signaling cascade. mAbs compete with ligand binding dimerization at the membrane receptor, while TKIs block ATP from binding to the intracellular tyrosine kinase domain preventing its phosphorylation and leading to a non-activated tyrosine kinase cascade (in the case of TKIs)(21). We have taken 5 drugs from these two classes for this study; dasatinib, imatinib, lapatinib, cetuximab, and trastuzumab. All five drugs are loosely bundled and referred to as TKIs/mAbs in this study.

2. Methods

In this study, we expanded the existing OAE to include TKI-associated adverse events. We focused our studies on 5 TKIs/mAbs - dasatinib, imatinib, lapatinib, cetuximab, and trastuzumab-because these agents are associated with a high frequency of cardiotoxicity-related AEs (21-23). The FDA Adverse Event Reporting System (FAERS) was mined based on these 5 TKIs/mAbs. Currently, there are over 3,900 MedDRA[®] terms in use in the FAERS database. Adverse event terms of interest were normalized to conform to ontology construction guidelines. This permits implementation with a formal representation having a proper definition, synonyms, cross-reference to MedDRA[®] and other reference terminologies, and relationships to other knowledge components such as anatomical site and

pathological process. This step includes both cardiotoxic and non-cardiotoxic adverse events as the drug may induce adverse events other than those related to the heart. A diagram summarizing this process is illustrated in Figure 1.

In the TKI queries in FAERS, AE terms for each drug were selected for the OAE-MedDRA® mapping process based on a Proportional Reporting Ratio (PRR) score of 2.0 and a minimum number of reports of three when the total number of reports is smaller than 1,500; or two in a thousand cases (0.2%) when the total number of reports is larger than 1,500 (17). These thresholds were selected to distinguish between a true adverse event signal and background noise based on the FDA's experience with AE signal detection. A PRR of 2.0 implies that the AE was observed for a given drug-AE pair twice as frequently as was observed for the AE in the background comprised of all FAERS reports included in the analysis. The FAERS data are a heterogeneous collection of reports from clinical experts, consumers, and manufacturers. Therefore, this screening process is necessary to the study to identify the "true" signal from the background noise.

Adverse events reported in FAERS are coded by MedDRA® terminology. When the result of mining FAERS with TKI queries was an AE term that corresponded to an existing OAE class, but the OAE class was previously missing a cross-reference to MedDRA®, we added the cross reference of that MedDRA® term to the particular OAE class. When the adverse events mined from FAERS did not previously exist in OAE, the new AE terms were added to OAE along with the MedDRA® cross-reference information. For this study, data mining was performed on the public FAERS dataset available as of April 2013.

Following the addition of the TKI-associated cardiac AE terms to OAE, we manually curated their classification and ontological relationship among the classes in OAE to ensure the accuracy of data representation. We recruited the help of clinicians with expertise in cardiotoxicity to determine the proper classification of terms before they were inserted into the ontology structure. A complete description of the methods have been detailed in the previously published work by He et al. (10).

2.1. Data clean-up and term insertion

For each term that was prioritized by the rules set above, the reference definition used in the ontology is taken from a medical dictionary or terminology resources available from the National Library of Medicine (NLM) MedlinePlus® (24). The definition is the driver for assigning the proper classification for the term based on the descriptor that can be linked to a pathophysiological process (e.g. edema, inflammation) with the anatomical location of the AE used as a secondary classification. This is a technique called axiomatisation in ontology practice. For example, *cardiac failure congestive AE* is asserted as a subclass of *cardiac failure AE* via *is_a* (SubClassOf) relation. This is the primary classification by the pathological process of congestive heart failure. At the same time, axiomatisation places *cardiac failure congestive AE* as having the manifestation site at the heart via the axiom *occurs_in* the heart by inheriting from the *cardiac disorder AE* parent. Cross-references to MedDRA:10007559, HP: 0001635, and SIDER:C0018802 are also annotated to *cardiac failure congestive AE* as described in Figure 2. Having this infrastructure also allows for the consolidation of the related clinical observations that are often used interchangeably by the

clinicians. For example, cardiomyopathy is used to indicate the cause of a heart failure observation, while in relational terms cardiomyopathy links to congestive heart failure via the intermediate term of cardiomegaly. Clearly separating these terms by an ontology mechanism helps identify the associated biological process that underlies the heart failure occurrence as demonstrated in Figure 2. Linking out to the permanent URL (PURL) for each cross-reference resource can be implemented with minimal scripting as the cross-reference annotation in OAE conforms to the OBO-dbXref annotation recommendation.

In this OAE TKI-AE expansion process, the AE terms were manually examined by clinical experts for their biological and pathological relevance, and synonyms were identified to ensure the AE term was represented with multiple alternative terms, which are commonly used in clinical settings (e.g. *cardiac failure congestive AE* is also annotated with the alternative terms of *congestive heart failure*, and *congestive cardiac failure*, that is more common in practice despite the preferred label *cardiac failure congestive* in MedDRA® dictionary – Figure 2). These synonyms were captured with each AE class with the ontology standard annotation property *alternative_term*. The prepared terms were then verified for data completeness and the proper location in the ontology hierarchical arrangement (asserted classification). Cross-references of the source definition were recorded with the annotation property *definition_source*. Information on term editor, editor's note, and other related metadata details were also recorded in the term's annotation properties. Other annotation, data, or object properties, that were not OAE-specific, were described with concepts inherited from upper-level ontologies that mandate OAE's ontology structure based on the Open Biological and Biomedical Ontologies (OBO) Foundry ontology development principles (25).

An AE term is a subclass of the overarching class *adverse event*. The primary classification favors biological and pathological process over the anatomical organs and tissues. Because the clinical manifestation may differ depending on the anatomic location that is primarily affected (e.g., mitochondrial toxicity manifests as pancreatitis in some cases, hepatotoxicity in others, and cardiotoxicity in yet others), the anatomic location is used as a secondary classifier, for which the classification hierarchy can be constructed (i.e. inferred) via the ontology reasoning mechanism, and therefore classification by organ system is also captured with OAE. Each AE class contains a reference to its corresponding MedDRA® number via the *seeAlso* annotation property. When applicable, reference was made to the corresponding Human Phenotype Ontology (HP)'s uniform resource identifiers (URI) (3) and/or the Mammalian Phenotype Ontology (MP)'s URIs (13) and annotated with a *seeAlso* annotation property.

New AE terms that were not previously in the OAE were manually defined and structurally aligned to OAE via the software program Ontorat (<http://ontorat.hegroup.org>) (26). Ontorat programmatically constructs and defines relationships of the newly-created AE classes by writing ontology syntax codes (axioms) to connect and align with the structure of the OAE. Final consistency checking and adjustment of the overall output were done using Protégé version 4.2 (<http://protégé.stanford.edu/>).

2.2. Adverse event classification and analysis

Different subgroups of AEs were classified by obtaining the terms and the related hierarchical terms and relations from the OAE using the OntoFox tool (27). The additional TKI-related AEs not previously in OAE were merged into the OAE in an ontology web language (OWL) format (28), an ontology standard file that uses the eXtensible Markup Language (XML) syntax for interoperability across different computer operating platforms. The individual AE classes in the resulting OAE expansion were visualized in the Protégé-OWL editor and manually verified by clinical scientists familiar with drug adverse event terminology.

3. Results

Mining the FAERS database with the five TKIs/mAbs resulted in a list of 1,053 distinct AE terms. Of these terms, 126 were already in the OAE prior to this investigation. After excluding ambiguous and non-AE terms such as medical procedures, we identified 844 new terms to insert into the OAE as newly-created classes. A spreadsheet of all the new terms was created as a common platform for collaboration between the informatics and clinical experts. The manual process of mining for the appropriate definition of all 844 adverse events as described in the Methods section, along with the candidate information of the parent term and the anatomical manifestation site was conducted by predoctoral individuals with expertise in pharmacology and engineering. The spreadsheet was reviewed by a clinical pharmacologist, and a clinical toxicologist at the FDA Office of Clinical Pharmacology to verify the accuracy of the information. These terms were then added to OAE by an ontologist. Although the manual process was labor intensive, it was necessary to provide the required accuracy for the ontology. This manual step took approximately one month to complete all 844 AE terms. Additional information about organs/tissues and MedDRA[®] terms led to adjustments in such attributes in some of the existing terms such as anatomical entity, MedDRA[®] ID, refined definition, and references from the original OAE design pattern (Figure 3). The information allowed modification of the knowledge to further improve clinical accuracy of OAE as this TKI-based study mined the information from a much more informative resource such as FAERS than mining from other sources without the clinical relevance (e.g. listing from general medical vocabularies). This adjustment of details may have changed some of the knowledge specific to certain terms in OAE, but did not affect the original hierarchy of the OAE, or the logical design patterns (the axiomatisation framework) of the ontology (*i.e.*, the design pattern as depicted in Figure 3 remained unchanged). The OAE is accessible via the Ontobee page at: <http://www.ontobee.org/ontology/OAE> and the National Center for Biomedical Ontology (NCBO) Bioportal ontology repository at <http://bioportal.bioontology.org/ontologies/1489/>. The Ontobee site (29) is the default linked data server for the OAE ontology. The OAE project homepage is located at <http://www.oae-ontology.org/>. A summary of OAE statistics is provided in Table 1. The current OAE revision dated December 18th, 2015 contains 5,274 terms including 5,109 classes, 156 object properties, and 96 annotation properties (Table 1).

3.1. OAE design pattern accommodates annotation of clinical adverse drug reactions

The OAE expansion in this study captures the cardiotoxicity AEs associated with TKIs/mAbs. The data analysis model is demonstrated in Figure 4A, with an example of imatinib-associated *pleural effusion AE* as a child of *edema AE* shown in Figure 4B. In this example, imatinib (an anticancer TKI) is orally administered, and is described by the ontological relation *occurs_in* (i.e., a drug exposure occurs in the location of a route of administration) inherited from the BFO relation (30). Components of the clinical, experimental, and biological processes were imported from the following existing resources; the Gene Ontology (GO) (31), the Ontology for Biomedical Investigation (OBI) (32), the Phenotypic Quality Ontology (PATO) (33), the Mammalian Phenotype Ontology (MP) (13), the Human Phenotype Ontology (HP) (3), the Information Artifact Ontology (IAO) (34), the Ontology for General Medical Sciences (OGMS) (15), UBERON anatomical ontology (35), the Human Disease Ontology (DOID) (14), and the Vaccine Ontology (VO) (36, 37). A summary of the imported classes is provided in Table 1. Cross-referencing with interoperable ontologies as defined in the OBO Foundry documentation allows linkages from OAE-defined AE class to other resources by importing appropriate terms and mapping them to those resources.

An AE occurs after a drug exposure (*exist_at* time *T2*), and therefore, *is_preceded_by* a drug exposure that occurs at a specific time during the treatment (*exist_at* time *T1*). The drug exposure is described through the use of class *planned_process*, which is a concept imported from the OBI. Therefore, the relationship between a reported AE and a preceding medical intervention is now established by the OAE information model. This aspect of temporal association between a treatment and its post-treatment AE differentiates OAE from other phenotype ontologies.

An AE is a subclass of the OGMS *pathological bodily process*. The AE hierarchy is then formed by an *is_a* (SubClassOf) relation that specifies a parent-child class term relation. This is an ontological relation that describes a specific AE as a type of an adverse event associated with the drug exposure without implying a causal association. The association between an AE and the corresponding anatomical entity is specified by an *occurs_in* relation. This description of OAE is encoded in the OWL format that can be coded for computational operations. The OAE .owl file is available for download on the project website <http://www.oae-ontology.org/>.

3.2. Classification of MedDRA coded content by the OAE hierarchy identifies different biological process clusters in different TKIs/mAbs

MedDRA-coded AE terms, as taken from FAERS for each TKI, were normalized and grouped into the OAE hierarchy. We also performed an intersection analysis of the AEs across the five TKIs/mAbs. There are 16 adverse event terms (counting at leaf nodes) that are common to all five TKIs/mAbs (Figure 5A) mostly involving homeostasis and inflammation responses (fever, pleural effusion, decreased blood granulocytes and potassium). One should note that cancer progression is also coded with MedDRA®, and this is often included in an analysis as an adverse event. In addition, common features of the

illness for which the patients are being treated (e.g., infection, bone marrow suppression due to other chemotherapy, etc.) occur across the five drugs.

Classification by the OAE structure on each individual TKI (see Electronic Supplementary Materials – Suppl_5TKIs.jpg) reveals that the TKIs/mAbs each have many more adverse events associated with the drug in each biological system beyond the 16 shared adverse events. To focus on the drug-AE association of interest for this study, 17 cardiotoxicity adverse event terms were used in querying FAERS data (Figure 5B). Each of these was associated with at least one of the five TKIs/mAbs and were the following 17 terms: cardiac function test abnormal, congestive cardiomyopathy, cardiovascular disorder, hypertrophic obstructive cardiomyopathy, cardiotoxicity, cardiac disorder, heart injury, ischaemic cardiomyopathy, hypertrophic cardiomyopathy, myocardial rupture, hypertensive cardiomyopathy, restrictive cardiomyopathy, multiple cardiac defects, viral cardiomyopathy, non-obstructive cardiomyopathy, hypertension, and hypotension. The OAE subset shown in Figure 5B lays out the relations among these AE terms and other related AE terms. The difference in the number of adverse events reported for each TKI is not to be interpreted that one drug is safer than the others. This is in part a function of the duration that the drug has been approved and available for use, resulting in different time periods of reporting for different TKIs/mAbs.

We further performed a concept mapping by linkage to other ontologies to expand the network of knowledge to include other relevant information as exemplified in Figure 6 with *myocardial infarction AE* and *arteriosclerosis coronary AE*. The diagram shows how an AE (a clinical observation) can be linked to biological systems and anatomical location (often not explicitly stated in a patient's record). In this figure, the concept mapping diagram shows that an AE is a subclass of its parent AE class by the pathological process. It is associated with an anatomical system (UBERON ontology) identified by an *occurs_in* relation. The pathological bodily process is referenced to GO biological process necrosis (which further links up to cell death activity by the GO-defined structure). Therefore, *myocardial infarction AE is_a cardiac disorder AE*, while *arteriosclerosis coronary artery AE is_a arterial disorder AE*. However, both *cardiac disorder AE* and *arterial disorder AE* are differentiated from the common parent class *cardiovascular disorder AE*. An OAE relation *is_evidence_of* has linked myocardial infarction as an AE that occurs in the myocardium to the arteriosclerosis coronary artery AE. Similar to the usage of the Evidence Ontology (ECO) (38), the *is_evidence_of* relation is used to link back an AE as evidence to the original question. Ontology linkage of the AEs associated with TKIs/mAbs yields the network of knowledge of biological response that can be mapped back to different levels of biological organization as illustrated in Figure 2 and Figure 6.

3.3. Use case: Identification of biological processes/genes involved in cardiac failure AE

As a use case for OAE, we examined if the links introduced in OAE to other ontologies will allow identification of biological processes and genes that contribute to cardiac AEs. We focused on *cardiac failure AE* as a use case (Figure 2). In OAE, *cardiac failure AE* is linked to Human Phenotype Ontology (HP) with a *seeAlso* annotation property to HP:0001635 (congestive heart failure). Within HP, HP:0001635 is given a class relation where it is

defined as “Equivalent to: *has_part* some (*decreased amount* and *inheres_in* some *heart contraction* and *has_modifier* some *abnormal*)”. ‘Heart contraction’ is a biological process in Gene Ontology (GO:0060047) and allows the information in HP to be linked to 2,861 genes with GO’s direct annotations of *heart contraction*. These annotations were done at the source database by manual curations (see Electronic Supplementary Materials – GO_annotation_CongestiveHeartFailure.xlsx). Of the 2,861 GO annotations, there are 190 unique *Homo sapiens* genes that are associated with heart contraction and could potentially have a role in cardiac failure (see Electronic Supplementary Materials – GO_HSGenes_heartContractions.xlsx). Therefore, using OAE and the relationships to other ontologies defined for each term in OAE, we are able to link the clinical phenotype to the biological processes and genes that contribute to that phenotype.

4. Discussion

The major contributions of this paper include (i) By retrieving and statistically analyzing the FAERS data, we identified 1,053 distinct MedDRA® AE terms significantly associated with the usage of 5 TKI/mAbs (i.e., dasatinib, imatinib, lapatinib, cetuximab, and trastuzumab). (ii) We extended OAE with 884 additional clinical AE terms and mapped these terms to terms in MedDRA and existing ontologies (e.g., HP, GO). The addition of these new terms has greatly strengthened the cardiovascular AE section of OAE. (iii) Our paper is the first of its kind to apply OAE to analyze real clinical data in drug pharmacovigilance. (iv) Our study provided the first use case demonstration that OAE creates a valid semantic framework for linking MedDRA terms to biological mechanisms of clinical phenotypes. (v) The OAE-based linkages provide potentially novel relations between TKI/mAb AEs and biological processes, and thus hypotheses that can be experimentally tested.

The process for addition of more than 800 AE terms to OAE provides a framework for further extension of OAE for a better coverage in translational health research. Selection of new terms to be created in OAE has been done on a case-study basis as there are over 36,000 terms inside MedDRA®, and not every term is used with the same frequency for coding AEs. Manually refining all MedDRA® terms into an ontological infrastructure at once is not feasible for this reason. The selection of terms, however, does not introduce bias to the knowledge representation model inside OAE as the value of this approach lies in the logical skeleton of the ontology itself, and not which data are represented. The manual examination of these terms along with the PRR signal analysis ensures that the biological relevance of an AE clinical observation is properly kept in the ontological format. To better serve various case studies, in the near future we plan to add all corresponding MedDRA® concepts that are frequently used in FAERS to OAE.

Although this study aims to analyze the cardiotoxic adverse events, the 3,320 native OAE classes represent observations in several system organ parts outside the heart. This is because (i) OAE adverse events are a collective set of studies, and (ii) TKI/mAb associated non-cardiotoxic AEs can be observed with cardiotoxic AEs. In our future work, it will be interesting to compare the cluster of the AE terms based on the MedDRA® System Organ Class (SOC) with the cluster of AE terms classified by *occurs_in* axioms within the OAE. We expect the result to be quite similar in the anatomy classification as OAE is continually

curated and axiomatized based on the incoming knowledge. The speculated difference lies in the area of when an AE resides in multiple system organ classes (such as an inflammation in an organ that can be classified both in the physiological location (e.g., skin, eye, lung), and a pathological location (e.g., immune system). Ontological axiomatisation of these biological activities will be more robust and computationally consistent than the classification in a traditional tree-structure listing of multiple parents achieved before the concept biomedical ontology was introduced to the field.

5. Conclusions

Ontology development is an on-going effort, especially in the domain of biomedical research where new knowledge generation is continuous. The OAE is designed to accommodate this continuous evolution of information and its interpretation. The ontology is built and expanded on a case-by-case basis. The expansion of OAE with TKI/mAb-associated AEs is a result of this process (9, 10, 17). With the focus on TKI/mAb-associated AEs, we have successfully added 844 new AEs terms to OAE, most of which are cardiovascular disorders AEs or related terms. The AE terms were cross-referenced with MedDRA[®] and other reference vocabularies or ontologies (MeSH, SNOMED-CT, HP, or MP). They were carefully examined by clinicians to ensure the proper classification, term definitions, and ontology relations to other terms within OAE. Reasoning with the knowledge embedded within OAE via axiomatisation of logical rules based on pathological process, anatomical location of AE manifestation sites, and the medical intervention that precedes the occurrence of AE may make the elucidation of molecular mechanisms associated with clinical drug AEs more efficient.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Zuoshuang Xiang for his programming assistance in building OAE.

Funding: This work is supported by the Oak Ridge Institute for Science and Education (ORISE) (Sirarat Sarnitivijai), the Undergraduate Research Opportunity Program at the University of Michigan (Shelley Zhang, Desikan Jagannathan, Yongqun He), FDA Commissioner's Fellowship Program (Shadia Zaman), NIEHS P30ES017885-01A1 (Gilbert Omenn), National Institute of Health grant U54 DA021529 and UL1 TR000433-09 (Brian Athey), and National Institute of Allergy and Infectious Disease grant R01 AI081062 (Yongqun He).

References

1. Zhichkin PE, Athey BD, Avigan MI, Abernethy DR. Needs for an expanded ontology-based classification of adverse drug reactions and related mechanisms. *Clinical pharmacology and therapeutics*. 2012; 91(6):963–5. [PubMed: 22609907]
2. Brown EG, Wood L. Coding of Data–MedDRA and other Medical Technologies. *Clinical Data Management*. :177–205.
3. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics*. 2008; 83(5):610–5. [PubMed: 18950739]

4. Elkin, PL.; Brown, SH.; Husser, CS.; Bauer, BA.; Wahner-Roedler, D.; Rosenbloom, ST., et al., editors. Mayo Clinic Proceedings. Elsevier; 2006. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists.
5. Organization WH. International statistical classification of diseases and related health problems: World Health Organization. 2004
6. Colevas A, Setser A. The NCI Common Terminology Criteria for Adverse Events (CTCAE) v 3.0 is the new standard for oncology clinical trials. J Clin Oncol. 2004; 22(14S):6098.
7. Health UDo, Services H. Common terminology criteria for adverse events (CTCAE) version 4.0. 2012. Available on line: http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf
8. Ceusters W, Capolupo M, De Moor G, Devlies J, Smith B. An evolutionary approach to realism-based adverse event representations. Methods of information in medicine. 2011; 50(1):62. [PubMed: 21057717]
9. He, Y.; Xiang, Z.; Sarntivijai, S.; Toldo, L.; Ceusters, W., editors. ICBO. 2011. AEO: A Realism-Based Biomedical Ontology for the Representation of Adverse Events.
10. He Y, Sarntivijai S, Lin Y, Xiang Z, Guo A, Zhang S, et al. OAE: The Ontology of Adverse Events. J Biomed Semantics. 2014; 5:29. [PubMed: 25093068]
11. Smith B, Kumar A, Bittner T. Basic formal ontology for bioinformatics. Journal of Information Systems. 2005:1–16.
12. Groza T, Kohler S, Moldenhauer D, Vasilevsky N, Baynam G, Zemojtel T, et al. The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease. American journal of human genetics. 2015; 97(1):111–24. [PubMed: 26119816]
13. Smith CL, Goldsmith CA, Eppig JT. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. Genome biology. 2005; 6(1):R7. [PubMed: 15642099]
14. Schriml LM, Arze C, Nadendla S, Chang Y-WW, Mazaitis M, Felix V, et al. Disease Ontology: a backbone for disease semantic integration. Nucleic acids research. 2012; 40(D1):D940–D6. [PubMed: 22080554]
15. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. Proceedings of the 2009 AMIA Summit on Translational Bioinformatics. 2009; 2009:116–20.
16. Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, et al. Modeling sample variables with an Experimental Factor Ontology. Bioinformatics. 2010; 26(8):1112–8. [PubMed: 20200009]
17. Sarntivijai S, Xiang Z, Shedden KA, Markel H, Omenn GS, Athey BD, et al. Ontology-based combinatorial comparative analysis of adverse events associated with killed and live influenza vaccines. PLoS One. 2012; 7(11):e49941. [PubMed: 23209624]
18. Tao C, He Y, Yang H, Poland GA, Chute CG. Ontology-based time information representation of vaccine adverse events in VAERS for temporal analysis. J Biomed Semantics. 2012; 3(1):13. [PubMed: 23256916]
19. Lin Y, He Y. The ontology of genetic susceptibility factors (OGSF) and its application in modeling genetic susceptibility to vaccine adverse events. J Biomed Semantics. 2014; 5:19. [PubMed: 24963371]
20. Marcos E, Zhao B, He Y. The Ontology of Vaccine Adverse Events (OVAE) and its usage in representing and analyzing adverse events associated with US-licensed human vaccines. J Biomed Semantics. 2013; 4(1):40. [PubMed: 24279920]
21. Force T, Krause DS, Van Etten RA. Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition. Nature Reviews Cancer. 2007; 7(5):332–44. [PubMed: 17457301]
22. Force T, Kerkelä R. Cardiotoxicity of the new cancer therapeutics—mechanisms of, and approaches to, the problem. Drug discovery today. 2008; 13(17):778–84. [PubMed: 18617014]
23. Force T, Kolaja KL. Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes. Nature reviews Drug discovery. 2011; 10(2):111–26. [PubMed: 21283106]
24. (US) NLoM. MedlinePlus [Internet] 2005 [cited 2014]. Available from: <https://www.nlm.nih.gov/medlineplus/>

25. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*. 2007; 25(11):1251–5.
26. Xiang Z, Zheng J, Lin Y, He Y. Ontorat: automatic generation of new ontology terms, annotations, and axioms based on ontology design patterns. *J Biomed Semantics*. 2015; 6:4. [PubMed: 25785185]
27. Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y. OntoFox: web-based support for ontology reuse. *BMC research notes*. 2010; 3(1):175. [PubMed: 20569493]
28. McGuinness DL, Van Harmelen F. OWL web ontology language overview. W3C recommendation. 2004; 10(2004-03):10.
29. Xiang, Z.; Mungall, C.; Ruttenberg, A.; He, Y., editors. ICBO. 2011. Ontobee: A Linked Data Server and Browser for Ontology Terms.
30. Arp R, Smith B. Function, role, and disposition in basic formal ontology. *Nature Precedings*. 2008; 1941(1):1–4.
31. Consortium GO. The Gene Ontology (GO) database and informatics resource. *Nucleic acids research*. 2004; 32(suppl 1):D258–D61. [PubMed: 14681407]
32. Brinkman RR, Courtot M, Derom D, Fostel JM, He Y, Lord P, et al. Modeling biomedical experimental processes with OBI. *J Biomed Semantics*. 2010; 1(Suppl 1):S7. [PubMed: 20626927]
33. Gkoutos, GV.; Mungall, C.; Dolken, S.; Ashburner, M.; Lewis, S.; Hancock, J., et al., editors. Entity/quality-based logical definitions for the human skeletal phenome using PATO; Engineering in Medicine and Biology Society, 2009 EMBC 2009 Annual International Conference of the IEEE; 2009. IEEE
34. Ceusters W. An information artifact ontology perspective on data collections and associated representational artifacts. *Studies in health technology and informatics*. 2012; 180:68–72. [PubMed: 22874154]
35. Mungall CJ, Torniai C, Gkoutos GV, Lewis SE, Haendel MA. Uberon, an integrative multi-species anatomy ontology. *Genome biology*. 2012; 13(1):R5. [PubMed: 22293552]
36. He Y, Rappuoli R, De Groot AS, Chen RT. Emerging vaccine informatics. *BioMed Research International*. 2011; 2010
37. He, Y.; Xiang, Z. Databases and in silico tools for vaccine design. *Silico Models for Drug Discovery*; Springer; 2013. p. 115-27.
38. Chibucos MC, Mungall CJ, Balakrishnan R, Christie KR, Huntley RP, White O, et al. Standardized description of scientific evidence using the Evidence Ontology (ECO). *Database : the journal of biological databases and curation*. 2014; 2014

Key messages

- OAE was extended with 884 additional clinical AE terms and correspondingly mapped to MedDRA terms to classify AEs identified from FAERS data analysis of TKIs/mAbs associated AEs.
- OAE-based linkages find potentially useful relations between TKI/mAb AEs and biological processes.
- The combination of MedDRA and OAE in the TKI study provided the first use case of a semantic framework for linking clinical drug AE phenotype to biological processes and mechanisms.

Figure 1.

The summary process diagram of curating existing OAE with additional TKI-association AE knowledge. The previously-existing OAE was further expanded with the novel 884 terms. These terms were statistically significant by PRR score and the cut-off threshold (PRR ≥ 2 , N ≥ 3). Each term was manually curated by clinician experts before it was placed at an appropriate hierarchical location inside OAE, along with the cross-references to related information such as MedDRA, and anatomical location. OAE was then used to classified TKI-associated adverse events and linked to biological processes in other ontologies such as Gene Ontology and via the organ systems.

Figure 2.

An example of using the OAE as a structured semantic framework to link a clinical observation *cardiac failure congestive AE* as a drug-associated adverse event to the underpinning biological relevance by the ontological cross-reference mechanism linking an AE to a phenotypic abnormality described in the Human Phenotype Ontology (HP). HP describes congestive heart failure as a phenotypic abnormality without any assumptions on what triggers such an observation. Combining knowledge in OAE with HP provides a seamless transition in linking a clinical observation (an AE) to biological processes and relevant human genes in the Gene Ontology via HP connection.

Figure 3.

OAE hierarchy of adverse event that facilitates the linkage of externally imported classes (blue) and the OAE-specific classes (yellow). OAE hierarchy is governed by the upper-level classes that are imported from the Basic Formal Ontology (BFO), the Ontology for General Medical Sciences (OGMS), and the Ontology for Biomedical Investigation (OBI); while the utility classes of anatomical location in which an adverse event is associated to are imported from the Uber Anatomy Ontology (UBERON).

Figure 4.

OAE class design for representing an AE. (A) Detailed OAE class structure of an AE as a subtype of another AE. The design pattern describes roles and interactions from a clinical perspective; (B) An example of the information of *pleural effusion AE* described by the OAE basic design pattern in association with imatinib treatment

Figure 5.

Classification of adverse events as hierarchical OAE subsets. (A) Classification of 16 AEs that appear in all TKIs and monoclonal antibodies (counting at leaf nodes). The 16 AEs and their associated top level OAE terms are classified in OAE. (B) Hierarchical display of 17 cardiovascular AEs associated with TKIs and monoclonal antibodies (counting at leaf nodes). Note that 'cardiovascular injury AE' (in red oval) and 'cardiovascular investigation result abnormal AE' (in blue oval) are not asserted under 'cardiovascular AE' (highlighted at the top); however, after ontology reasoning, these two AE terms and their child terms are inferred to under 'cardiovascular AE'. OntoFox was used to generate these OAE subset files, and the Protégé OWL editor was used for the visualization.

Figure 6.

Concept mapping diagram – an AE is a subclass of (defined with an *is_a* relation) a parent AE by its pathological process and is associated to an anatomical location which is linked to the UBERON anatomy ontology by an ontology relation *occurs_in*. The pathological process is mapped to the corresponding GO biological process via *is_evidence_of* relation. Knowledge integration can be drawn from the concept mapping without having to put all relevant information into a single compilation. In this example, myocardial infarction AE is classified as a type of cardiac disorder AE, while arteriosclerosis coronary artery AE is a type of arterial disorder AE. Both cardiac disorder AE and arterial disorder AE are differentiated from cardiovascular disorder AE. The myocardial infarction AE as an evidence of arteriosclerosis coronary artery AE is connected via *is_evidence_of* relation. The *occurs_in* relation has specified myocardial infarction as an AE that occurs in the myocardium in the heart. These ontological relations provide a standardized data infrastructure to create linkage of knowledge among different standard ontologies.

Table 1

Current description of OAE-specific and imported terms.

Ontology	Classes	Object Properties	Annotation Properties	Total
OAE	3220	8	1	3,229
BFO (Basic Formal Ontology)	21	19	2	42
RO (Relation Ontology)	0	89	3	92
IAO (Information Artifact Ontology)	8	1	15	9
OBI (Ontology for Biomedical Investigation)	11	2	2	13
OGMS (Ontology for General Medical Sciences)	5	0	0	5
VO (Vaccine Ontology)	7	0	0	7
GO (Gene Ontology)	330	0	0	330
UBERON Anatomy Ontology	1060	0	0	1060
PATO (Phenotypic Quality Ontology)	41	0	0	41
DOID (Disease Ontology)	1	0	0	1
Other 17 ontologies	315	37	73	442
Totals	5019	156	96	5274

OAE includes 3229 terms whose term IDs start with the “OAE_” prefix. These OAE-specific terms include medical interventions, adverse event terms, and eight ontological relations describing activities associating with an adverse event (e.g., *is_evidence_of*). Many related ontology terms that have been defined in other existing biomedical ontologies are imported from correspondent ontologies as shown in the table.