Drug Safety (2019) 42:657–670
https://doi.org/10.1007/s40264-018-0767-7

ORIGINAL RESEARCH ARTICLE

Development of a Controlled Vocabulary-Based Adverse Drug Reaction Signal Dictionary for Multicenter Electronic Health Record-Based Pharmacovigilance

Suehyun Lee[1,2] · Jongsoo Han[1] · Rae Woong Park[3] · Grace Juyun Kim[1] · John Hoon Rim[4,5,6] · Jooyoung Cho[5,7] · Kye Hwa Lee[1,8] · Jisan Lee[9] · Sujeong Kim[10] · Ju Han Kim[1,8]

Published online: 16 January 2019

Abstract

Introduction Integration of controlled vocabulary-based electronic health record (EHR) observational data is essential for real-time large-scale pharmacovigilance studies.

Objective To provide a semantically enriched adverse drug reaction (ADR) dictionary for post-market drug safety research and enable multicenter EHR-based extensive ADR signal detection and evaluation, we developed a comprehensive controlled vocabulary-based ADR signal dictionary (CVAD) for pharmacovigilance.

Methods A CVAD consists of (1) administrative disease classifications of the *International Classification of Diseases* (ICD) codes mapped to the *Medical Dictionary for Regulatory Activities* Preferred Terms (MedDRA® PTs); (2) two teaching hospitals' codes for laboratory test results mapped to the Logical Observation Identifiers Names and Codes (LOINC) terms and MedDRA® PTs; and (3) clinical narratives and ADRs encoded by standard nursing statements (encoded by the International Classification for Nursing Practice [ICNP]) mapped to the World Health Organization–Adverse Reaction Terminology (WHO-ART) terms and MedDRA® PTs.

Results Of the standard 4514 MedDRA® PTs from Side Effect Resources (SIDER) 4.1, 1130 (25.03%), 942 (20.86%), and 83 (1.83%) terms were systematically mapped to clinical narratives, laboratory test results, and disease classifications, respectively. For the evaluation, we loaded multi-source EHR data. We first performed a clinical expert review of the CVAD clinical relevance and a three-drug ADR case analyses consisting of linezolid-induced thrombocytopenia, warfarin-induced bleeding tendency, and vancomycin-induced acute kidney injury.

Conclusion CVAD had a high coverage of ADRs and integrated standard controlled vocabularies to the EHR data sources, and researchers can take advantage of these features for EHR observational data-based extensive pharmacovigilance studies to improve sensitivity and specificity.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40264-018-0767-7) contains supplementary material, which is available to authorized users.

Suehyun Lee and Jongsoo Han contributed equally to this work.

✉ Ju Han Kim
juhan@snu.ac.kr

1 Division of Biomedical Informatics, Seoul National University Biomedical Informatics (SNUBI), Seoul National University College of Medicine, Seoul, Korea

2 Department of Biomedical Informatics, College of Medicine, Konyang University, Daejeon, Korea

3 Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

4 Department of Laboratory Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

5 Physician-Scientist Program, Department of Medicine, Yonsei University Graduate School of Medicine, Seoul, Korea

6 Department of Pharmacology, Yonsei University College of Medicine, Seoul, Korea

7 Department of Laboratory Medicine, Wonju Severance Christian Hospital, Yonsei University Wonju College of Medicine, Wonju, Korea

8 Precision Medicine Center, Seoul National University Hospital, Seoul, Korea

9 College of Nursing, Catholic University of Pusan, Busan, Korea

10 College of Nursing, Seattle University, Seattle, USA

Key Points

Large-sale electronic health record (EHR)-based computational approaches are an important part of pharmacovigilance. As the use of EHRs expand, research on adverse drug reaction (ADR) signal detection using real patient records is increasing.

To enable the use of EHR data for adverse event reporting or clinical research we resolved the terminological differences between EHRs and standard ADR terminology.

An ADR signal dictionary for pharmacovigilance provide exploration applicable to extensive electronic medical record data.

1 Introduction

An adverse drug reaction (ADR) is defined as an appreciably harmful or unpleasant reaction, resulting from an intervention related to the use of a medicinal product, which predicts hazards for future administration and warrants prevention or specific treatment, alteration of the dosage regimen, or withdrawal of the product [1]. In many countries, the number of ADRs reported has increased dramatically. For instance, in the USA, the number of serious ADRs increased to 807,270 in 2014 from 264,227 in 2006 [2]. Therefore, finding an effective pharmacovigilance modality is one of the most imperative issues for patient safety.

As the electronic health record (EHR) systems expand, research into ADR signal detection is increasing [3–5]. The Semantic integrAtion and reasoning Framework for pharmacovigilancE signals Research (SAFER) project [3] provided a semantically enriched platform for combinatorial ADR signal detection using combinations of diverse open-source ADR signal detection methods and publicly available data sources.

The SAFER project focused on three main data sources: spontaneous reporting systems (SRSs), observational databases, and free-text resources for semantic harmonization of computational ADR signal detection methods for pharmacovigilance. In particular, the PharmacoVigilance Signal Detector Ontology (PV-SDO) supports the development of an integrated platform for ADR signal detection [6]. PV-SDO categorizes the different data sources employed for signal detection and classifies signal detection methods in terms of their types, underlying computational models, inputs, outputs, analysis parameters, requirements of use, strengths, and weaknesses. Declerck et al. [7] developed a method to extract patient data from EHRs for automatically populating adverse event (AE) reporting forms in the scope of the Scalable, Standard-based Interoperability Framework for Sustainable Proactive Post Market Safety Studies (SALUS) project. SALUS built an interoperability platform and a dedicated toolkit to enable secondary use of EHR data for post-marketing drug surveillance [7]. To enable the use of EHR data for AE reporting, clinical research, or both, the terminological differences between EHRs and standard ADR vocabularies need to be addressed. Terminological reasoning also needs to be designed to ensure automatic conversion of source EHR terminologies to standardized ADR vocabularies [7]. Laboratory test results are the most commonly used EHR data resources for the detection of ADR signals [8]. ADR signal detection from laboratory results requires systematic mapping to translate laboratory abnormalities into ADR signals. Therefore, it is necessary to systematically map laboratory test names (or codes) and their results to symptom-based terms, such as those contained in the *Medical Dictionary for Regulatory Activities* (MedDRA®). In contrast, narrative data such as doctor's notes or nursing records capture ADR symptoms adequately but are not commonly used for ADR signal detection because of poor codification, standardization, or both. An ideal ADR signal detection method should be able to investigate all drugs by all signs, all symptoms, and all abnormalities.

Therefore, systematic mappings between various EHR data resources and ADR-related standardized terms are essential for extensively detecting ADR signals from EHR data generated and stored daily in the course of clinical practice. Nurses are believed to play a more important role than doctors and pharmacists in discovering and spontaneously reporting ADRs [9, 10]. This may be because nurses regularly perform clinical observations/recordings with more standardized nursing statements rather than with the diagnosis codes and test results recorded by doctors [9–12]. Nursing records contain ADR signals in the form of clinical symptoms and signs, such as dizziness, dry mouth, and weight gain, which are not detectable by laboratory tests. For all nursing documents, the Seoul National University Hospital (SNUH) (Seoul, Korea) EHR has applied standard nursing statements (SNSs) encoded by the International Classification for Nursing Practice (ICNP) for 14 years at the institutional level (Electronic Supplementary Material Table 1).

Observational Health Data Sciences and Informatics (OHDSI) proposed the Observational Medical Outcomes Partnership (OMOP), which is to be employed to study AEs of medical products using observational healthcare databases. The OMOP Common Data Model (CDM) enables researchers to integrate healthcare data from diverse sources in a consistent and standardized way. OHDSI tools are under development. HERMES (Health Entity Relationship and Metadata Exploration System) is a web-based vocabulary

browsing tool in which the user is able to search for a term and explore related concepts [13].

In this study, we developed a controlled vocabulary-based ADR signal dictionary (CVAD) for pharmacovigilance that encompasses disease classifications, laboratory tests, and clinical narratives for use in extensive EHR data exploration. The aim of CVAD is to make this information available for ADR researchers to apply to heterogeneous EHR data resources for pharmacovigilance. CVAD is an ADR signal dictionary that maps various terms in EHR resources to commonly used ADR-controlled vocabularies. For the evaluation, we used CVAD to detect ADR signals from the EHR data resources of teaching hospitals, using the ADR controlled vocabulary. CVAD enables daily EHR-based pharmacovigilance in a standardized manner, with systematic ADR signal detection for patient safety.

2 Methods

We conducted the present study in two phases. First, we developed the CVAD by reviewing numerous pharmacovigilance studies. This first phase involved reviewing the ADR sources and proposing a concept for the CVAD. The review of the pharmacovigilance studies (Table 1) and mapping between standard ADR terms used MedDRA®, and the controlled vocabularies used the *International Classification of Diseases* (ICD), Logical Observation Identifiers Names and Codes (LOINC), and World Health Organization-Adverse Reaction Terminology (WHO-ART). Secondly, we applied the CVAD to the two local EHR systems in Korean tertiary hospitals; we mapped the terminologies in the EHR to the controlled vocabularies. The applicability of CVAD for studying the drug–ADR signals of linezolid-induced thrombocytopenia, warfarin-induced bleeding tendency, and vancomycin-induced acute kidney injury was then shown.

2.1 Review of the Studies for Pharmacovigilance

We reviewed numerous articles and chose disease classifications, laboratory tests, and clinical narratives for standardization to develop the CVAD for EHR-based pharmacovigilance (Table 1). We searched for terms such as "Pharmacovigilance", "EHR", "Controlled vocabularies" and "Adverse drug reaction" in PubMed in 2018 for related articles. As a result, we extracted papers from 2010 to 2018. After that, we determined which data categories to use in this study. The ICD is the most widely used disease classification for EHR-based pharmacovigilance [14]. Laboratory test names are well-encoded, and numerous laboratory results are in numerical formats, enabling researchers to set objective criteria for ADRs. Free-text clinical narratives are much more difficult to exploit than other data sources. Clinical narratives are hindered by issues such as misspelling, grammatical errors, and acronyms. However, it is believed to be more valuable than other data sources because it covers a wide portion of clinical information in daily practice and, thereby, provides rich ADR information. Many researchers have attempted to use clinical narratives in ADR by applying natural language processing (NLP) to map the extracted terms to ontologies, reference terminologies, or controlled vocabularies.

2.2 Controlled Vocabulary Mapping Strategy

We developed CVAD to enable the use of a variety of data sources in pharmacovigilance studies. CVAD consists of standard ADR terms, controlled vocabularies, and EHR data. We employed the subset of MedDRA® PTs that were utilized at the Side Effect Resources (SIDER) 4.1 project to form standard ADR terms in the CVAD [33]. SIDER covers almost all known drug-ADR pairs because it reviews several well-established ADR databases. SIDER 4.1 contains 1345 drugs, 4251 ADRs, and 163,207 drug-ADR pairs. The core concept of developing CVAD is to map various data sources of EHR, i.e., disease classifications, laboratory tests, and clinical narratives, to controlled vocabularies in ICD, LOINC, and WHO-ART, respectively, and then to map them to the standard ADR terms (Fig. 1).

2.3 Mapping Between Standard Adverse Drug Reaction (ADR) Terms and Controlled Vocabularies

The mapping between the ICD and MedDRA® is already well-established via the Unified Medical Language System (UMLS). We included considerable human resources in the MedDRA®–LOINC and MedDRA®–WHO-ART mappings. We used two steps to map the ICD, 10th edition (ICD-10) to MedDRA® Preferred Terms (PTs). We first mapped Hohl et al.'s [34] 608 ICD-10 terms related to ADRs [34] to the UMLS. SIDER provides mapping relations between the UMLS Concept Unique Identifier (CUI) and MedDRA® PTs, and, therefore, We extracted existing mapped contents using information regarding existing MedDRA® UMLS CUIs and ICD UMLS CUIs provided by SIDER. LOINC was manually mapped to MedDRA® by three physicians. Two of the physicians were in residency training in the Department of Laboratory Medicine at their respective institution and one was a general physician. They chose the MedDRA® PTs to be mapped to LOINC. Table 2 shows the inclusion and exclusion criteria. A MedDRA® PT that fitted any of the inclusion criteria was considered a candidate. Otherwise, the term was considered excluded based on the exclusion criteria. If a term did not meet the inclusion or exclusion criteria, the physicians attempted to map the term

Table 1 Pharmacovigilance studies using electronic health records

Study	EHR data used	Data categories used			Controlled vocabularies
		Disease classification	Laboratory tests	Clinical narrative	
Park et al. [15]	Admission, discharge, drug prescription, laboratory tests	×	O	×	N/A
Liu et al. [16]	Laboratory tests, drug prescription	×	O	×	N/A
Ji et al. [17]	Drug prescription, disease classification	O	×	×	ICD-9
Yoon et al. [18]	Admission, discharge, drug prescription, laboratory tests, disease classification	×	O	×	ICD-10
LePendu et al. [19]	Clinical notes, drug prescription, disease classification, devices, procedures	O	×	O	RxNorm, ATC, UMLS, MedDRA®, ICD-9
Overhage et al. [20]	Drug prescription, disease classification, procedures	O	O	×	SNOMED, HCPCS, ICD-9, CPT-4, LOINC
Coloma et al. [21]	Drug prescription, disease classification	O	×	×	ATC, ICD-9-CM, ICD-10, ICPC
Eriksson et al. [22]	Drug prescription, disease classification, laboratory test, clinical narrative	×	O	O	N/A
Stausberg [23]	Disease classification	O	×	×	ICD-10, national procedure classification, ICD-9-CM, ICD-10-GM
Neubert et al. [24]	Drug prescription, disease classification, laboratory test	×	O	O	ATC, WHO-ART, LOINC
Patel and Kaelber [25]	Drug prescription, disease classification, findings, procedures, laboratory test	O	O	×	UMLS, SNOMED-CT, RxNorm, LOINC
Haerian et al. [26]	Laboratory data, medication lists, and information in discharge summary notes	×	O	O	UMLS
Li et al. [27]	Admission notes, discharge summaries, laboratory tests, structured diagnosis	×	O	O	UMLS, ICD-9
Li et al. [28]	Retrospective narrative outpatient visits, admission notes, discharge summaries, and structured medication orders and laboratory results	×	O	O	UMLS, ICD-9
Reich et al. [29]	Inpatient, outpatient, and pharmacy services supplemented with clinical laboratory results	O	O	×	ICD-9, SNOMED-CT, UMLS
Reisinger et al. [30]	Demographics, drug exposures, condition occurrence	O	×	×	MedDRA®, SNOMED-CT, NDC, CPT, HCPCS, ICD-9
Ryan et al. [31]	Medical (inpatient, outpatient, and emergency room), pharmacy, and laboratory data (including test results)	O	O	×	MedDRA®, SNOMED-CT, NDC, CPT, HCPCS, ICD-9
Wang et al. [32]	Unstructured clinical notes	×	×	O	RxNorm, UMLS, MedDRA®, WHO-ART, STITCH

ATC Anatomical Therapeutic Chemical, *CPT* Current Procedural Terminology, *EHR* Electronic Health Record, *HCPCS* Healthcare Common Procedure Coding System, *ICD* International Classification of Diseases, *ICD-9* ICD, 9th edition, *ICD-9-CM* ICD, 9th edition, Clinical Modification, *ICD-10* ICD, 10th edition, *ICD-10-GM* ICD, 10th edition, German Modification, *ICPC* International Classification of Primary Care, *LOINC* Logical Observation Identifiers Names and Codes, *MedDRA®* Medical Dictionary for Regulatory Activities, *SNOMED* Systematized Nomenclature of Medicine, *SNOMED-CT* SNOMED Clinical Terms, *STITCH* Search Tool for Interactions of Chemicals, *UMLS* Unified Medical Language System, *WHO-ART* World Health Organization–Adverse Reaction Terminology

to the full extent of their clinical knowledge. The physicians mapped up to six LOINC terms to a MedDRA® term when we used the top 2000 common laboratory results of LOINC for the mapping process. We classified the laboratory results as normal versus abnormal. In the case of anomalies, they are classified as high, low, or other depending on the characteristics of each test. In the case of hepatitis, when the value of alanine aminotransferase test increases, it is indicated as high. In the case of anemia, where the value of the hemoglobin test is decreased, it is indicated as low. In the case of chromaturia, it is noted as other and has a positive or negative meaning. We marked each laboratory result as high, low, or other for each LOINC term according to the associated MedDRA® PTs and their threshold values. Two medical

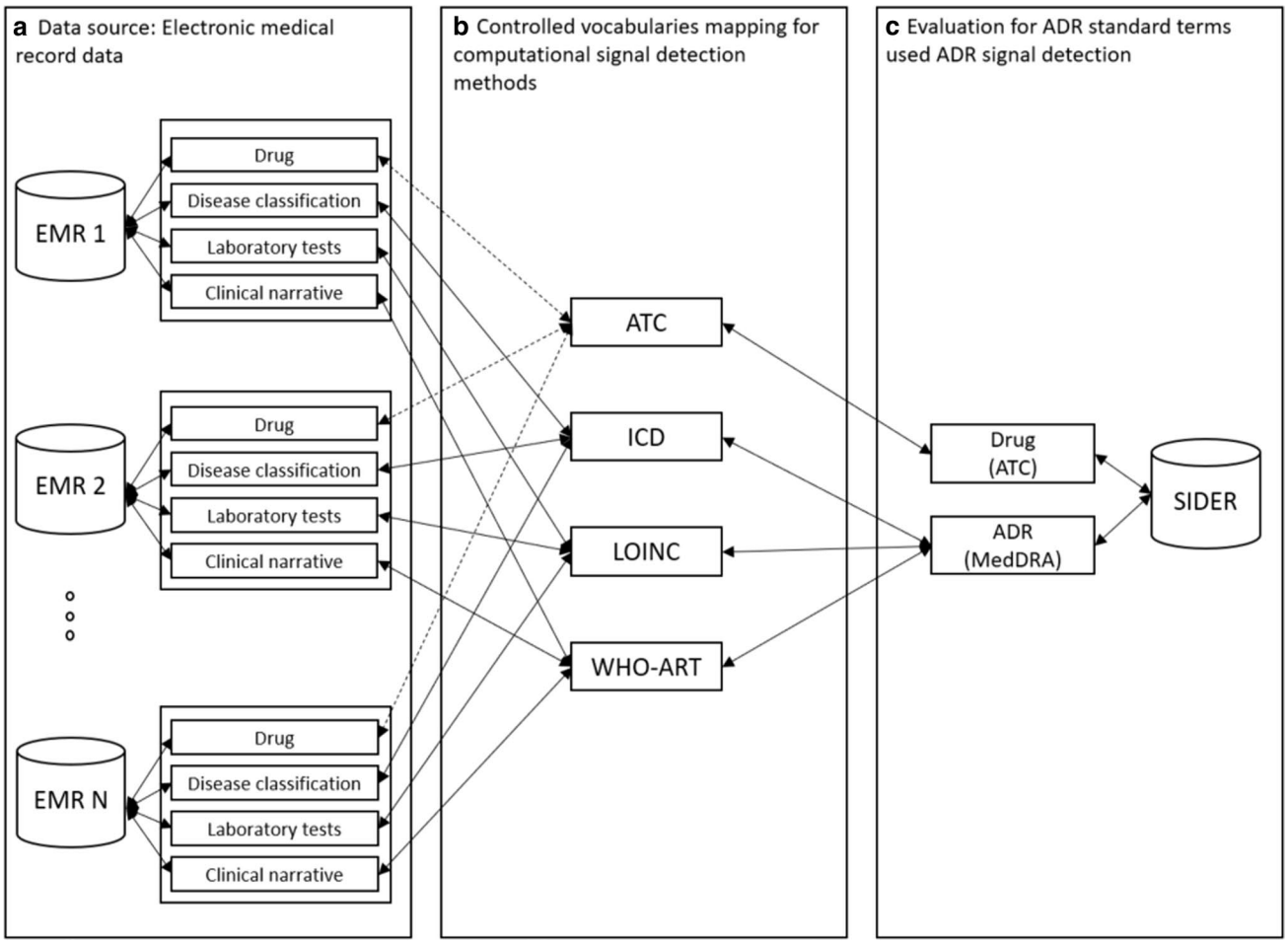

Fig. 1 Controlled vocabulary mapping strategy of CVAD. **a** Extraction–Transformation–Loading of EHR data into CVAD data structure. **b** Data sources of EHR, i.e., drugs, disease classifications, laboratory results, and clinical narratives are mapped to STITCH, ICD, LOINC, and WHO-ART, respectively. **c** Evaluation and integration of ADR signals using the SIDER database. *ADR* adverse drug reaction, *CVAD* controlled vocabulary-based ADR signal dictionary, *EHR* electronic health record, *ICD* International Classification of Diseases, *LOINC* Logical Observation Identifiers Names and Codes, *SIDER* Side Effect Resources, *STITCH* Search Tool for Interacting Chemicals, *WHO-ART* World Health Organization–Adverse Reaction Terminology

residents created the primary mapping table between LOINC and the MedDRA® PTs, and then a specialist created the secondary mapping table. Lastly, a consensus meeting finalized the mapping results.

A total of three experts were involved in the clinical narrative mapping. The WHO-ART terms were mapped to MedDRA® by family medicine practitioners, a general physician, and a pharmacist. We selected the WHO-ART terms related to bedside nursing observations by matching them to the SNUH SNSs (encoded by the ICNP). In the first mapping step, the family medicine practitioner mapped the WHO-ART terms to the entire MedDRA® PTs. The general physician and pharmacist then evaluated the appropriateness of the mapping performed in the first step. The evaluation consisted of two checkpoints: first, the general practitioner and pharmacist independently evaluated the appropriateness of the mapping; and, second, all three had a consensus meeting to finalize the mapping table. The terms used in this study are from version 16.1 of the MedDRA® dictionary, ICD-10 2016 version, LOINC Mapper's Guide to Top 2000 US Lab Tests version 1.6, ICNP version 1.0, WHO-ART version 092, and *Korean Standard Classification of Diseases* (KCD)-7.

2.4 Evaluation of a Controlled Vocabulary-Based ADR Signal Dictionary (CVAD) in Two Teaching Hospitals in Korea

The evaluation of CVAD was performed with the EHR data from two local teaching hospitals, SNUH and Ajou University Hospital (AUH), with 1800 and 1030 inpatient beds, respectively. A total of three physicians participated in matching EHRs to LOINC. Mapping of ICNP-based SNUH nursing records and WHO-ART was performed by a general

Table 2 Inclusion and exclusion criteria of the *Medical Dictionary for Regulatory Activities* (MedDRA®) Preferred Terms for mapping to Logical Observation Identifiers Names and Codes

Criteria	Examples
Inclusion criteria	
Terms describing laboratory test results	Hyperbilirubinemia, hypokalemia, agranulocytosis, proteinuria
Disease that requires laboratory test results to make diagnosis	Rheumatoid arthritis, nephrotic syndrome, hemolytic uremic syndrome, tubulointerstitial nephritis, endocarditis, scleroderma, spondylitis
Disease that presents laboratory test abnormality	Dermatitis, bone cancer metastatic, hepatic cancer metastatic, abdominal abscess, virilism, otitis media acute, relapsing fever
Exclusion criteria	
Laboratory test terms without abnormalities	Hemoglobin, muscle mass, immunology test, cardioactive drug level
Clinical studies other than laboratory tests (i.e., imaging studies, physical examinations)	Weight increase, joint range of motion decreased, intraocular pressure increased, electrocardiogram QT prolonged
Symptoms and clinical investigations	Rash, pain, nausea, skin discoloration, convulsion, back pain, dry mouth, enuresis, decreased libido
Internal or surgical procedures	Vasodilation procedure, assisted fertilization, muscle relaxant therapy, surgery, sensitization, hypoventilation
Disease that is easy to diagnose without laboratory tests	Eyelid ptosis, cerebral ischemia, cardiac arrest, erectile dysfunction, hypotension, dermatitis atopic
Too broad terms	Disability, infarction, acute abdomen, obstruction, fetal disorder, hypertrophy, angiopathy

physician. The KCD is a Korean version of the ICD with additional subcategories and some traditional medicine-related diagnostic terms. KCD-7 is currently used and is based on ICD-10. We downloaded ICD-10 from the WHO Classification [35] and KCD-7 from the Health Insurance Review and Assessment Service [36] websites. Laboratory test codes and names were matched to LOINC by three physicians. A general physician validated the EHR-based laboratory results. Some laboratory test codes overlapped because of the redundant use of the same laboratory tests in different departments, renewal of the laboratory test codes, or both. After validation, the two physicians in residency in the Department of Laboratory Medicine from Severance Hospital created a primary mapping table and the general physician participated in the final one-to-one mapping. We mapped the SNUH nursing records [37] (encoded by the ICNP [38]) to the WHO-ART terms. These mapping and manual curation processes for building the CVAD in the present study took over 2 years to complete and have been validated by six clinical experts.

We found that only 69 of the 608 ICD-10 terms suggested by Hohl et al. [34] were mapped to 83 MedDRA® PTs in SIDER during the disease classification mapping. ICD-10 diagnosis terms in daily practice cannot include most of the symptoms or signs experienced by patients. For example, the MedDRA® PTs such as dizziness, weight increase, nervousness, sepsis, and inflammation were not mapped to ICD-10. For LOINC laboratory test names, of the 4514 MedDRA® PTs in SIDER, 942 and 3572 MedDRA® PTs were mapped and excluded, respectively, according to the selection and exclusion criteria in Table 2.

Table 3 illustrates our two-stage evaluation of the clinical relevance of matching clinical narratives to MedDRA® PTs and its results. In the first evaluation, the general physician and pharmacist cross-checked each other's evaluation and the degree of agreement was calculated. If there was a disagreement, the two worked together to figure out the appropriate terms. In the next step, three experts (general physician, pharmacist, family medicine specialist) gathered and reviewed the results. If the mapping was found to not be appropriate and new mapping was suggested, they modified the match when agreement was reached. Compared with the first-stage mapping of 239 WHO-ART to 1235 MedDRA® PTs using 1439 SNSs, we successfully mapped 267 WHO-ART terms, chosen as nursing observation-related, to 1130

Table 3 Two-stage evaluation of the clinical relevance of matching 267 World Health Organization–Adverse Reaction Terminology (WHO-ART) terms to 1130 *Medical Dictionary for Regulatory Activities* (MedDRA®) Preferred Terms

	Evaluator B	
	Appropriate	Inappropriate
First evaluation (κ score = 0.34)		
Evaluator A		
Appropriate	824	32
Inappropriate	293	86
Second evaluation (κ score = 0.84)		
Evaluator A		
Appropriate	1130	6
Inappropriate	22	77

MedDRA® PTs using 1934 SNSs. This was evaluated by a general physician and pharmacist and reviewed by family medicine specialists, who exhibited a high degree of inter-observational agreement ($\kappa = 0.84$, $P < 0.001$).

3 Results

3.1 CVAD Statistics

Of the 4514 MedDRA® PTs in SIDER, 1762 (36.03%) were mapped to one of the three EHR data sources of CVAD (Fig. 2a). Clinical narratives, laboratory tests, and disease classifications covered 1130 (25.03%), 942 (20.86%), and 83 (1.83%) of the MedDRA® PTs, respectively. Only nine MedDRA® PTs were covered by all three EHR data sources (Fig. 2b). The number of mapped disease classification terms of the local EHR was the same as the controlled vocabularies used in CVAD. Laboratory tests in the two EHRs were matched to 277 LOINC terms—182 for SNUH and 183 for AUH—in CVAD. We mapped 1934 SNSs to 267 WHO-ART terms (Table 4).

CVAD contains 1345 drugs, 4251 ADRs, and 163,207 drug–ADR pairs (Electronic Supplementary Material 2a).

Table 4 Controlled vocabulary-based adverse drug reaction signal dictionary (CVAD) terms mapped to clinical narratives, laboratory tests, and disease classifications

	Narrative	Laboratory test	Disease classification
No. of ADRs (MedDRA® PT)	1130	942	83
No. of SNS	1934		
No. of WHO-ART	267		
No. of LOINC		277	
No. of Laboratory tests in SNUH		182	
No. of Laboratory tests in AUH		183	
No. of ICD-10			69

ADR adverse drug reaction, *AUH* Ajou University Hospital, *ICD-10* International Classification of Diseases, 10th edition, *LOINC* Logical Observation Identifiers Names and Codes, *MedDRA® PT* Medical Dictionary for Regulatory Activities Preferred Term, *SNS* Standard Nursing Statement, *SNUH* Seoul National University Hospital, *WHO-ART* World Health Organization–Adverse Reaction Terminology

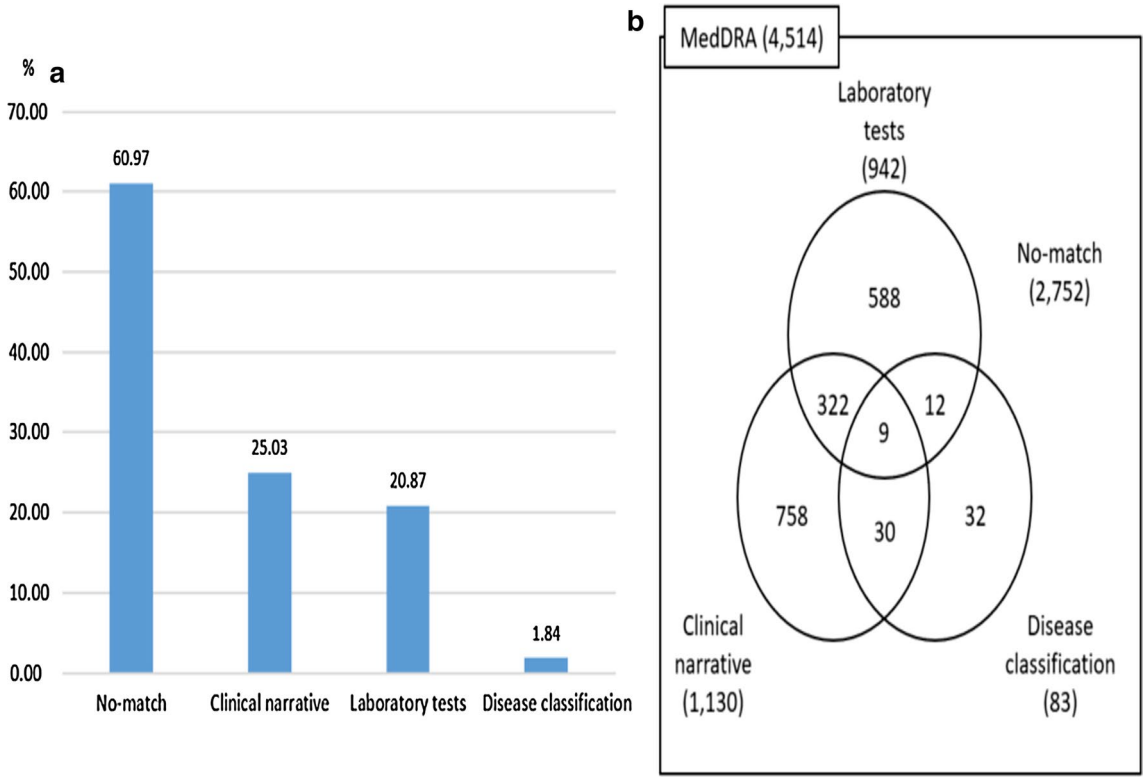

Fig. 2 Distributions of clinical narratives, laboratory tests, and disease classification in the controlled ADR vocabulary of CVAD. **a** Mapping percentage of clinical narrative, laboratory, and disease classification. **b** Venn diagram showing the number of mapped terms in the three data sources. *ADR* adverse drug reaction, *CVAD* controlled vocabulary-based ADR signal dictionary

3209 ADRs terms were mapped to UMLS CUIs and we assigned CVAD IDs to enable large-scale multicenter EHR data-based ADR studies (Electronic Supplementary Material 2b). Of the 4514 MedDRA® PTs that were utilized in SIDER, 1762 (39.04%) were mapped to one of the three EHR data sources: clinical narratives, laboratory tests, and disease classifications (Fig. 2a).

3.2 Examples and Use Cases of Application to Local Electronic Health Records Using CVAD

CVAD consists of (1) administrative classifications for the ICD codes mapped to MedDRA® PTs; (2) SNUH and AUH codes for laboratory tests mapped to the LOINC terms and MedDRA® PTs; and (3) both clinical narratives and ADRs encoded by SNS terms mapped to WHO-ART terms and MedDRA® PTs. We manually mapped the SNS terms (encoded by the ICNP) and MedDRA® PTs to WHO-ART [31] in CVAD, enabling systematic EHR-based ADR signal detection. Table 5 shows the mapping of CVAD for disease classifications, laboratory abnormalities, and clinical narratives that can be applied to pharmacovigilance studies using EHR data sources (Electronic Supplementary Material Table 2). CVAD enables multicenter EHR-based extensive ADR signal detection studies by providing a CVAD for different implementations. The contents are summarized as follows. If an issue of drug safety occurs, multiple institutional analyzes may be required. The terminology used by each institution differs when conducting multicenter analyses. In the case of the term thrombocytopenia in Fig. 3, the administrative code defined in CVAD can be retrieved from the EHRs using UMLS code C0040034 for Thrombocytopenia or ICD-10 code D69.6 for Thrombocytopenia, unspecified. Also, when searching by laboratory test, it is possible to search using 777-3, Platelets in Blood by Automated count in LOINC. Finally, in the case of clinical narrative, the study can be conducted by searching with "Decreased the level of hematocrit".

This electronic medical record multicenter study allows post-market surveillance, pharmacovigilance, and drug repositioning to proceed. As shown in Fig. 3, CVAD can facilitate the identification of various types of phenotypes associated with specific ADRs of interest. For example, linezolid is an antimicrobial agent commonly used for the treatment of vancomycin-resistant enterococci and methicillin-resistant *Staphylococcus aureus* (MRSA) infections. However, the development of thrombocytopenia, one of the most frequent AEs of this antimicrobial agent, can lead to discontinuation of linezolid treatment [39]. As can be seen in Fig. 3, thrombocytopenia, as an ADR, could be indicated by numerous terms from different controlled vocabularies: "C0040034|Thrombocytopenia", "777-3|Platelets in Blood by Automated count", "Decreased the level of hematocrit", "Thrombocytopenia", and "D69.6|Thrombocytopenia, unspecified". Systematic exploration of ADRs in EHR observational data for thrombocytopenia could benefit from the use of CVAD. A serious and common complication of warfarin treatment is the risk of bleeding [40]. For any particular patient, the potential benefit of preventing thromboembolic disease needs to be balanced against the potential harm of induced hemorrhagic AEs [41]. As an ADR, this coagulopathy can be expressed via multiple terms from controlled vocabularies such as "C0005779|Disorder of hemostatic system", "14979-9|Activated partial thromboplastin time (aPTT) in Platelet-poor plasma by Coagulation assay", "Used inspirometer", "Coagulation disorder", and "D68.9|Coagulation defect, unspecified". The widespread use of more intensive regimens has been associated with an increase in vancomycin-induced nephrotoxicity reports [42]. Ramírez et al. [43] reported that renal function and vancomycin trough concentrations should be closely monitored from week 2 of treatment in adults, intensive care patients, and those who receive concurrent nephrotoxic agents using a pharmacovigilance program from laboratory signals at a hospital. Renal failure as an ADR could be composed of controlled vocabularies such as "C0035078|Renal failure", "2160-0|Creatinine [Mass/volume] in Serum or Plasma", "Urine osmolality below normal", "Renal function abnormal", and "N19|Unspecified renal failure" (Fig. 3).

4 Discussion

This study developed a CVAD for pharmacovigilance that encompasses disease classifications, laboratory tests, and clinical narratives for extensive EHR data exploration. CVAD aims to make this information available for ADR researchers to apply to heterogeneous EHR data resources for pharmacovigilance. Some of the greatest obstacles to ADR monitoring using real-world patient data are (1) how well the surrogate marker for the ADR signal reflects actual patient data and (2) how to isolate significant ADR signals from the volatility patterns in general medical situations. In addition, hospital-based ADR monitoring and reporting programs aim to identify and quantify the risks associated with the use of drugs administered in a hospital setting [44]. The success of pharmacovigilance depends on solving these problems. While these challenges could be addressed by the development of a variety of pharmacovigilance methods, the most important strategy is to ensure semantic interoperability by linking clinical data with ADR data [45].

The present study proposed a CVAD for pharmacovigilance that applies to extensive EHR data exploration to achieve the integration of different data sources. With CVAD, the same level of pharmacovigilance can be performed in different hospitals with different EHR data

Table 5 Example controlled vocabulary-based adverse drug reaction signal dictionary (CVAD) terms for (a) disease classification, (b) laboratory results, and (c) clinical narratives

MedDRA® in SIDER	Controlled vocabulary		EHR systems			
	ICD-10		SNUH (ICD-10)		AUH (KCD-7)	
Name	ID	Name	ID	Name	ID	Name
(a) CVAD disease						
Hypersensitivity	T78.4	Allergy, unspecified	T78.4	Allergy, unspecified	T784	Allergy, unspecified
Hematuria	R31	Unspecified hematuria	R31	Unspecified hematuria	R31	Unspecified hematuria
Pharyngeal haemorrhage	R04.1	Hemorrhage from throat	R04.1	Hemorrhage from throat	R041	Hemorrhage from throat

MedDRA® in SIDER	Controlled vocabulary				EHR systems			
	LOINC				SNUH		AUH	
Name	ID	Name	Unit	High/low/other	Code	Name	Code	Name
(b) CVAD Lab								
Cardiac failure	30934-4	Natriuretic peptide B [Mass/volume] in serum or plasma	pg/mL	H	L8191	B-type Natriuretic Peptide (BNP) (em)	227	POCT BNP
	10839-9	Troponin I. cardiac [Mass/volume] in serum or plasma	ng/mL	H	L3546	Troponin I	215	Troponin I
	6598_07_01	Troponin T. cardiac [Mass/volume] in serum or plasma	µg/L	H	L3553	Troponin T	132	Troponin T hs
Hepatitis	1742-6	Alanine aminotransferase [Enzymatic activity/volume] in serum or plasma	U/L	H	L3015	Alanine aminotransferase (ALT)	4	Alanine aminotransferase (ALT)
	1920_8	Aspartate aminotransferase [Enzymatic activity/volume] in serum or plasma	U/L	H	L3014	Aspartate aminotransferase (AST)	3	Aspartate aminotransferase (AST)
Rheumatoid arthritis	11572_5	Rheumatoid factor [Units/volume] in serum	IU/mL	H	L5103	Rheumatoid factor	102	Rheumatoid factor
	30341-2	Erythrocyte sedimentation rate	mm/h	H	L7201	Erythrocyte sedimentation rate (ESR)	76	Erythrocyte sedimentation rate (ESR)
	1988_5	C reactive protein [Mass/volume] in serum or plasma	mg/dL	H	L5102	C-reactive protein (CRP)	74	C-reactive protein (CRP)
	33935_8	Cyclic citrullinated peptide IgG Ab [Units/volume] in serum		H	L7348	Anti cyclic citrullinated peptide (CCP) antibody	320	Anti-CCP Ab IgG
	8061_4	Nuclear Ab [Presence] in serum		O	L5128	Antinuclear Ab(FANA)	150	ANA Screening

MedDRA® in SIDER	Controlled vocabulary				EHR systems
	WHO-ART		SNS		ICNP
Name	ID	Name	ID	Statement	Terms
(c) CVAD Clinical					
Ascites	715	Ascites	9232	Ascites present	Ascites, Yes
			9233	No ascites	Ascites, No

Table 5 (continued)

MedDRA® in SIDER	Controlled vocabulary				EHR systems
	WHO-ART		SNS		ICNP
Name	ID	Name	ID	Statement	Terms
Abnormal faces	1537	Faecal abnormality nos	322	Changed stool aspect	stool characteristics, Bowel Elimination, Altered
			325	Changed stool color	stool characteristics, color, Bowel Elimination, Altered
			1608	Normal defecation	stool characteristics, Yes, Bowel Elimination defecation amount, defecation amount
			3455	No mucoid stool	Bowel Elimination , No
			3456	Mucoid stool present	color, Yes, Bowel Elimination, defecation amount

AUH Ajou University Hospital, *CVAD* comprehensive controlled vocabulary-based adverse drug reaction signal dictionary, *EHR* electronic health record, *ICD-10* International Classification of Diseases, 10th edition, *ICNP* International Classification for Nursing Practice, *KCD-7* Korean Standard Classification of Diseases-7, *LOINC* Logical Observation Identifiers Names and Codes, *MedDRA® PT* Medical Dictionary for Regulatory Activities Preferred Term, *SIDER* Side Effect Resources, *SNS* Standard Nursing Statement, *SNUH* Seoul National University Hospital, *WHO-ART* World Health Organization–Adverse Reaction Terminology

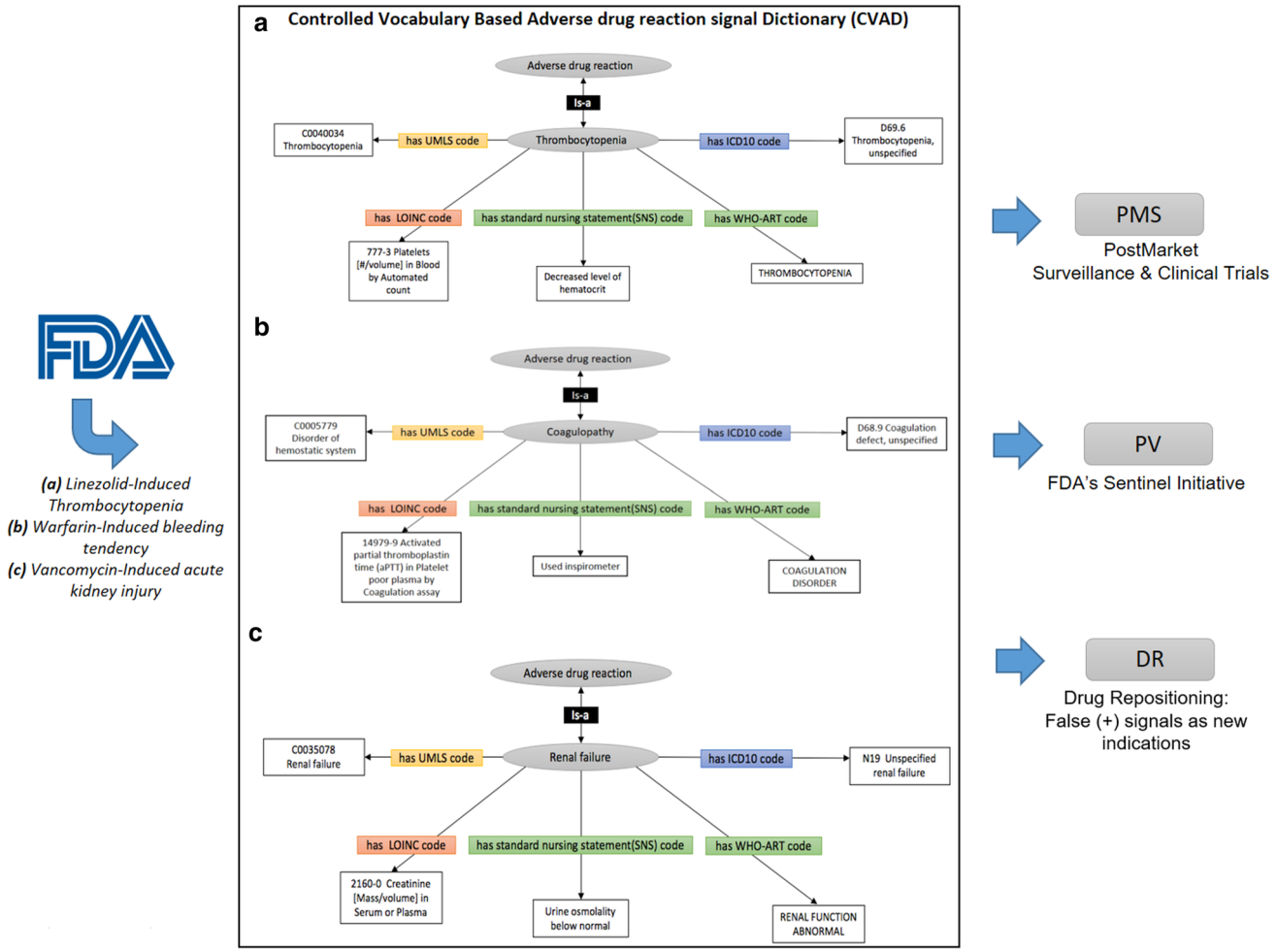

Fig. 3 Use case of CVAD for PMS and clinical trials, US FDA's sentinel initiative (PV), and DR. **a** LZD-induced thrombocytopenia, **b** warfarin-induced bleeding tendency, and **c** vancomycin-induced acute kidney injury cases. *ADR* adverse drug reaction, *CVAD* controlled vocabulary-based ADR signal dictionary, *DR* drug repositioning, *FDA* Food and Drug Administration, *LZD* linezolid, *PMS* post-market surveillance, *PV* pharmacovigilance

structures, and the pharmacovigilance data can be extended to a larger ADR reference standard. Therefore, CVAD could be an extremely useful pharmacovigilance tool. CVAD consists of the MedDRA® PTs contained in SIDER, covering 996 drugs and 3209 MedDRA® PTs. SIDER has been used as a benchmarking set in ADR studies [46, 47] and as a source of other ADR databases since its launch in 2010 [48, 49]. Although drug–ADR pair databases such as SIDER are not considered to contain pairs with considerable evidence, it is reasonable to use as a core terminology because SIDER was developed based on data from regulatory bodies, which are well-organized among ADR databases.

CVAD uses MedDRA® PTs in SIDER that covers a wide range of drug–ADR pairs for systematic integration of many EHR ADR data sources. Current ADR study results are fragmented and tend to focus on a few drugs, ADRs, or both. The use of one data source to conduct an ADR signal detection study would make it hard to detect the signal because of its rarity as an ADR, or because its ADR characteristics are limited to certain data. Thus, an increasing number of ADR studies are attempting to use large datasets across different EHR sources [26]. Researchers can also integrate study results with a few drug–ADR pairs using the ADR signal dictionary. We hope CVAD will enable large-scale multicenter EHR data-based ADR studies to overcome the limited sensitivity of previous ADR studies that rely on restricted data sources such as literature reviews or regulatory databases [50]. Diagnosis in daily practice alone, for example, cannot include most of the symptoms or signs that patients experience. This leads to low sensitivity of the signal detection when using diagnosis data in ADR signal detection. Integrating various EHR data sources such as laboratory tests or clinical narratives addresses this problem by allowing the mutual complementarity of different data sources. Symptoms such as a headache or stomachache and abnormal physical examination findings are not detectable with laboratory tests [51]. Abnormal liver and kidney function or abnormal kidney function can be found in clinical narratives through NLP methods. These methods are also effective for searching for these symptoms through laboratory test results. For this reason, it is important to use CVAD to find discoverable drug AEs.

CVAD structure can be linked with existing OMOP vocabulary to increase versatility of both models. For example, the condition concept structure of OMOP vocabulary employed the SNOMED-CT (Systematized Nomenclature of Medicine Clinical Terms) as a standard vocabulary and ICD-10, ICD-10, Clinical Modification (ICD-10-CM), and other terminologies as source codes. CVAD is also applicable to part of the OMOP vocabulary. For example, the SNSs of SNUH statements (encoded by the ICNP) could include OMOP vocabulary source codes such as ICD-10, SNOMED-CT, MeSH (Medical Subject Headings), etc. We also considered maintenance of CVAD to be useful in a variety of research. Implementation of CVAD consists of extracting the resources from each hospital and mapping them into standard terminology. CVAD was developed with (1) Extraction–Transformation–Loading of EHR data into CVAD data structure; and (2) EHR data sources, i.e., drugs, disease classifications, laboratory results, and clinical narratives are mapped to Search Tool for Interactions of Chemicals (STITCH), ICD, LOINC, and WHO-ART, respectively. The example shown in Fig. 1a does not need to be continually updated because all of the EHRs for the two hospitals have been mapped to the full range of terms available in pharmacovigilance studies. However, the overall mapping needs to be reviewed and updated by additional expert groups. Also, the strategies given in Fig. 1b and c will periodically be updated with reference to the SNOMED-CT terminology update. We also referred to the guidelines outlined in the OMOP Common Vocabulary Model and considered the OMOP vocabulary maintenance concept such as that built on the shoulders of National Library of Medicine UMLS which requires community participation and support in the OHDSI network [52].

The present study has some limitations. CVAD covers 1762 MedDRA® PTs, which consists of 36.03% of the 4251 total MedDRA® PTs in SIDER 4.1. The unmappable 1447 MedDRA® PTs that were not used in the ADR dictionary may affect the use of the ADR signal dictionary. Since the controlled vocabulary of disease classification was matched to only 83 MedDRA® PTs, we reviewed the laboratory test and clinical narrative as part of the ADR dictionary. We reviewed the proportion of the MedDRA® PTs that fall into the exclusion criteria of the mapping of a laboratory test. Also, "Symptoms and clinical investigations", "Clinical studies other than laboratory tests", and "Disease which is easy to diagnose without laboratory tests" were covered the most by clinical narrative terms. The MedDRA® PTs excluded using the exclusion criteria "Symptoms and clinical investigations" were highly covered by clinical narrative because nursing records are the major source of clinical investigation and symptoms. The exclusion criteria "Laboratory test terms without abnormalities", "Internal or surgical procedures", "Too broad terms", and "Inherited diseases" are the terms that are not useful in ADR studies.

Currently, the clinical data used in CVAD are diagnosis, laboratory test, and nursing records. To better reflect actual patient data, various data resources, such as image inspection, physical examinations, doctor's notes, and so on, need to be added to the CVAD. The gap between clinical data and ADR standard terms was also much larger than we expected. This may be because the definitions of most ADRs have accumulated through voluntary AE reports. Since real-time drug safety monitoring using real patient data can be used as a major strategy to monitor drug AEs, new areas of AEs

are expected to emerge [54]. Therefore, the ADR terminology system will gradually evolve to reflect actual clinical situations, and CVAD is expected to improve this further.

CVAD will be available online to enable researchers to explore it. We are currently developing a web page with user-friendly browsing and searching functions which will integrate and release the results of pharmacovigilance studies that employed CVAD. We believe that it will contribute to pharmacovigilance research in the future.

Future pharmacovigilance is expected to begin at the drug development stage and include the entire life cycle of the drug. As the clinical and research gaps narrow and unconfirmed AEs become more common, patient health improvements and healthcare system efficiencies will increase [53]. Therefore, it is clear that CVAD is a very useful and appropriate pharmacovigilance support tool.

5 Conclusion

Large-sale EHR-based computational approaches are an important part of pharmacovigilance. As the use of EHRs expands, research on ADR signal detection using real patient records is increasing. CVAD has a high coverage of ADRs and integrates standard controlled vocabularies to the EHR data sources, and researchers can take advantage of these features for EHR observational data-based extensive pharmacovigilance studies to improve sensitivity and specificity.

Acknowledgements The authors would like to acknowledge the National Research Foundation of Korea (NRF) and Korean Health Technology R&D Project, Ministry of Health and Welfare. The authors thank the anonymous reviewers for their helpful feedback.

Compliance with Ethical Standards

Funding This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science ICT and Future Planning (MSIP) (2018R1D1A1B07049155) and by a grant from the Korean Health Technology R&D Project, Ministry of Health and Welfare (HI13C2164, HI16C11280000). This research was supported by a grant (16183MFDS541) from Ministry of Food and Drug Safety in 2018.

Conflict of interest Ju Han Kim, Suehyun Lee, Jongsoo Han, Rae Woong Park, Grace Juyun Kim, John Hoon Rim, Jooyoung Cho, Kye Hwa Lee, Jisan Lee, and Sujeong Kim have no conflicts of interest directly relevant to the content of this study. The results of this study do not reflect the views of the National Research Foundation of Korea (NRF) or Ministry of Health and Welfare.

References

1. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet. 2000;356(9237):1255–9.
2. FDA. FAERS reporting by patient outcomes by year. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070461.htm. Accessed 1 Nov 2016.
3. Koutkias VG, Jaulent MC. Computational approaches for pharmacovigilance signal detection: toward integrated and semantically-enriched frameworks. Drug Saf. 2015;38(3):219–32. https://doi.org/10.1007/s40264-015-0278-8.
4. Hauben M, Madigan D, Gerrits CM, Walsh L, Van Puijenbroek EP. The role of data mining in pharmacovigilance. Expert Opin Drug Saf. 2005;4(5):929–48.
5. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther. 2012;91(6):1010–21. https://doi.org/10.1038/clpt.2012.50.
6. Koutkias V, Jaulent M-C. Leveraging post-marketing drug safety research through semantic technologies. In: The PharmacoVigilance signal detectors ontology, SWAT4LS workshop, 10 Dec 2014, Berlin; 2014.
7. Declerck G, Hussain S, Daniel C, Yuksel M, Laleci GB, Twagirumukiza M, et al. Bridging data models and terminologies to support adverse drug event reporting using EHR data. Methods Inf Med. 2015;54(1):24–31. https://doi.org/10.3414/ME13-02-0025.
8. Lee S, Choi J, Kim HS, Kim GJ, Lee KH, Park CH, et al. Standard-based comprehensive detection of adverse drug reaction signals from nursing statements and laboratory results in electronic health records. J Am Med Inform Assoc. 2017;24(4):697–708. https://doi.org/10.1093/jamia/ocw168.
9. Backstrom M, Mjorndal T, Dahlqvist R. Spontaneous reporting of adverse drug reactions by nurses. Pharmacoepidemiol Drug Saf. 2002;118:647–50.
10. Ranganathan SS, Houghton JE, Davies DP, Routledge PA. The involvement of nurses in reporting suspected adverse drug reactions: experience with the meningococcal vaccination scheme. Br J Clin Pharmacol. 2003;566:658–63.
11. Ahn HJ, Park HA. Adverse-drug-event surveillance using narrative nursing records in electronic nursing records. Comput Inform Nurs. 2013;311:45–51.
12. Conforti A, Opri S, D'Incau P, et al. Adverse drug reaction reporting by nurses: analysis of Italian pharmacovigilance database. Pharmacoepidemiol Drug Saf. 2012;216:597–602.
13. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform. 2015;216:574–8.
14. WHO. ICD-10: international statistical classification of diseases and health related problems: tenth revision. 2nd ed. Geneva: World Health Organization; 2004.
15. Park MY, Yoon D, Lee K, Kang SY, Park I, Lee SH, et al. A novel algorithm for detection of adverse drug reaction signals using a hospital electronic medical record database. Pharmacoepidemiol Drug Saf. 2011;20(6):598–607. https://doi.org/10.1002/pds.2139.
16. Liu M, McPeek Hinz ER, Matheny ME, Denny JC, Schildcrout JS, Miller RA, et al. Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records. J Am Med Inform Assoc. 2013;20(3):420–6. https://doi.org/10.1136/amiajnl-2012-001119.
17. Ji Y, Ying H, Dews P, Mansour A, Tran J, Miller RE, et al. A potential causal association mining algorithm for screening adverse drug reactions in postmarketing surveillance. IEEE Trans Inf Technol Biomed. 2011;15(3):428–37. https://doi.org/10.1109/TITB.2011.2131669.
18. Yoon D, Park MY, Choi NK, Park BJ, Kim JH, Park RW. Detection of adverse drug reaction signals using an electronic health records database: comparison of the Laboratory Extreme

Abnormality Ratio (CLEAR) algorithm. Clin Pharmacol Ther. 2012;91(3):467–74. https://doi.org/10.1038/clpt.2011.248.
19. LePendu P, Iyer SV, Bauer-Mehren A, Harpaz R, Mortensen JM, Podchiyska T, et al. Pharmacovigilance using clinical notes. Clin Pharmacol Ther. 2013;93(6):547–55. https://doi.org/10.1038/clpt.2013.47.
20. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54–60. https://doi.org/10.1136/amiajnl-2011-000376.
21. Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. EU-ADR Consortium. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf. 2011;20(1):1–11. https://doi.org/10.1002/pds.2053.
22. Eriksson R, Jensen PB, Frankild S, Jensen LJ, Brunak S. Dictionary construction and identification of possible adverse drug events in Danish clinical narrative text. J Am Med Inform Assoc. 2013;20(5):947–53. https://doi.org/10.1136/amiajnl-2013-001708.
23. Stausberg J. International prevalence of adverse drug events in hospitals: an analysis of routine data from England, Germany, and the USA. BMC Health Serv Res. 2014;13(14):125. https://doi.org/10.1186/1472-6963-14-125.
24. Neubert A, Dormann H, Prokosch HU, Bürkle T, Rascher W, Sojer R, et al. E-pharmacovigilance: development and implementation of a computable knowledge base to identify adverse drug reactions. Br J Clin Pharmacol. 2013;76(Suppl 1):69–77. https://doi.org/10.1111/bcp.12127.
25. Patel VN, Kaelber DC. Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: a case study of azathioprine. J Biomed Inform. 2014;52:36–42. https://doi.org/10.1016/j.jbi.2013.10.009.
26. Haerian K, Varn D, Vaidya S, Ena L, Chase HS, Friedman C. Detection of pharmacovigilance-related adverse events using electronic health records and automated methods. Clin Pharmacol Ther. 2012;92(2):228–34. https://doi.org/10.1038/clpt.2012.54.
27. Li Y, Ryan PB, Wei Y, Friedman C. A method to combine signals from spontaneous reporting systems and observational healthcare data to detect adverse drug reactions. Drug Saf. 2015;38(10):895–908. https://doi.org/10.1007/s40264-015-0314-8.
28. Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. J Am Med Inform Assoc. 2014;21(2):308–14. https://doi.org/10.1136/amiajnl-2013-001718.
29. Reich C, Ryan PB, Stang PE, Rocca M. Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. J Biomed Inform. 2012;45(4):689–96. https://doi.org/10.1016/j.jbi.2012.05.002.
30. Reisinger SJ, Ryan PB, O'Hara DJ, Powell GE, Painter JL, Pattishall EN, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. J Am Med Inform Assoc. 2010;17(6):652–62. https://doi.org/10.1136/jamia.2009.002477.
31. Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012;31(30):4401–15. https://doi.org/10.1002/sim.5620.
32. Wang L, Rastegar-Mojarad M, Ji Z, Liu S, Liu K, Moon S, et al. Detecting pharmacovigilance signals combining electronic medical records with spontaneous reports: a case study of conventional disease-modifying antirheumatic drugs for rheumatoid arthritis. Front Pharmacol. 2018;7(9):875. https://doi.org/10.3389/fphar.2018.00875.
33. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol. 2010;6:343. https://doi.org/10.1038/msb.2009.98.
34. Hohl CM, Karpov A, Reddekopp L, Doyle-Waters M, Stausberg J. ICD-10 codes used to identify adverse drug events in administrative data: a systematic review. J Am Med Inform Assoc. 2014;21(3):547–57. https://doi.org/10.1136/amiajnl-2013-002116.
35. Classification of Disease (ICD). https://www.who.int/classifications/icd/icdonlineversions/en/. Accessed 15 May 2016.
36. Korean Standard Classification of Diseases (KCD). https://kssc.kostat.go.kr:8443/ksscNew_web/kssc/main/main.do?gubun=1. Accessed 12 Dec 2018.
37. Yu OS, Park IS, Joo YH, Woo KS, Shin HJ, Ahn TS, et al. Classification of nursing statements based on the ICNP, the HHCC, and the nursing process for use in electronic nursing records. Stud Health Technol Inform. 2006;122:718–21.
38. Park IS, Shin HJ, Kim EM, Park HA, Kim YA, Jo EM. Mapping nursing statements with the ICNP and its practical use in electronic nursing records. Stud Health Technol Inform. 2006;122:989–90.
39. Tajima M, Kato Y, Matsumoto J, Hirosawa I, Suzuki M, Takashio Y, et al. Linezolid-induced thrombocytopenia is caused by suppression of platelet production via phosphorylation of myosin light chain 2. Biol Pharm Bull. 2016;39(11):1846–51.
40. Shoeb M, Fang MC. Assessing bleeding risk in patients taking anticoagulants. J Thromb Thrombolysis. 2013;35(3):312–9. https://doi.org/10.1007/s11239-013-0899-7.
41. Fitzmaurice DA, Blann AD, Lip GY. Bleeding risks of antithrombotic therapy. BMJ. 2002;325(7368):828–31.
42. van Hal SJ, Paterson DL, Lodise TP. Systematic review and meta-analysis of vancomycin-induced nephrotoxicity associated with dosing schedules that maintain troughs between 15 and 20 milligrams per liter. Antimicrob Agents Chemother. 2013;57(2):734–44. https://doi.org/10.1128/AAC.01568-12.
43. Ramírez E, Jiménez C, Borobia AM, Tong HY, Medrano N, Krauel-Bidwell L, et al. Vancomycin-induced acute kidney injury detected by a prospective pharmacovigilance program from laboratory signals. Ther Drug Monit. 2013;35(3):360–6. https://doi.org/10.1097/FTD.0b013e318286eb86.
44. Lobo MG, Pinheiro SM, Castro JG, Momenté VG, Pranchevicius MC. Adverse drug reaction monitoring: support for pharmacovigilance at a tertiary care hospital in Northern Brazil. BMC Pharmacol Toxicol. 2013;14:5. https://doi.org/10.1186/2050-6511-14-5.
45. Härmark L, van Grootheest AC. Pharmacovigilance: methods, recent developments and future perspectives. Eur J Clin Pharmacol. 2008;64(8):743–52. https://doi.org/10.1007/s00228-008-0475-9.
46. Xu R, Wang Q. Automatic construction of a large-scale and accurate drug-side-effect association knowledge base from biomedical literature. J Biomed Inform. 2014;51:191–9. https://doi.org/10.1016/j.jbi.2014.05.013.
47. Gurulingappa H, Mateen-Rajput A, Toldo L. Extraction of potential adverse drug events from medical case reports. J Biomed Semant. 2012;3(1):15. https://doi.org/10.1186/2041-1480-3-15.
48. Cai MC, Xu Q, Pan YJ, Pan W, Ji N, Li YB, et al. ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. Nucleic Acids Res. 2015;43(Database issue):D907–13. https://doi.org/10.1093/nar/gku1066.
49. Juan-Blanco T, Duran-Frigola M, Aloy P. IntSide: a web server for the chemical and biological examination of drug side effects. Bioinformatics. 2015;31(4):612–3. https://doi.org/10.1093/bioinformatics/btu688.
50. Khan LM, Al-Harthi SE, Alkreathy HM, Osman A-MM, Ali AS. Detection of adverse drug reactions by medication antidote

signals and comparison of their sensitivity with common methods of ADR detection. Saudi Pharm J. 2015;23(5):515–22. https://doi.org/10.1016/j.jsps.2014.10.003.
51. Hui C, Vaillancourt R, Bair L, Wong E, King JW. Accuracy of adverse drug reaction documentation upon implementation of an ambulatory electronic health record system. Drugs Real World Outcomes. 2016;3(2):231–8. https://doi.org/10.1007/s40801-016-0071-8.
52. Belenkaya R, Natarajan K, Velez M, Voss E. OMOP common data model (CDM) & extract-transform-load (ETL) tutorial. 24 Sep 2016. https://www.ohdsi.org/wp-content/uploads/2016/09/MAIN-OHDSI-Symposium-2016-Common-Data-Model-and-Extract-Transform-Load-Tutorial.pptx.pdf. Accessed 4 Dec 2018.
53. Santoro A, Genov G, Spooner A, Raine J, Arlett P. Promoting and protecting public health: how the European Union pharmacovigilance system works. Drug Saf. 2017;40(10):855–69. https://doi.org/10.1007/s40264-017-0572-8.
54. Wise L, Parkinson J, Raine J, Breckenridge A. New approaches to drug safety: a pharmacovigilance tool kit. Nat Rev Drug Discov. 2009;8(10):779–82. https://doi.org/10.1038/nrd3002.