

Common data model for decision support system of adverse drug reaction to extract knowledge from multi-center database

Mi Jung Rho^{1,2} · Si Ra Kim² · Seung Ho Park³ · Kwang Soo Jang³ · Bum Joon Park³ · Joo Young Hong² · In Young Choi^{1,2}

Published online: 3 July 2015

© Springer Science+Business Media New York 2015

Abstract An adverse drug reaction (ADR) surveillance system integrated with various electronic medical record (EMR) systems has been suggested as an effective way to collect more data and analyze ADRs earlier than the spontaneous reporting of ADRs. Because Korean hospitals have heterogeneous EMR databases, a common data model (CDM) should first be defined to develop the multi-center EMR-based drug surveillance system. We investigated the data models from two prominent drug safety surveillance studies, the Mini-Sentinel program and the Observational Medical Outcomes Partnership, and developed an EMR-based ADR common data model (EADR CDM). The EADR CDM has eight tables, including a demographic table, drug table, visit table, procedure table, diagnosis table, death table, laboratory table and organization table. Each table consists of 5–12 fields. Among a total of 2,931,060 patients from January 2008 to December 2012 in clinical data warehouse of the S hospital, we extracted the data from 135,745 patients who were prescribed below drugs to determine whether the exported data were sufficient to detect ADRs of six drugs. After validation, we found that the transformed data based on EADR CDM is helpful to understand the prescription pattern and explore feasible medication list for adverse drug signal detection.

The collection of diverse data using the CDM is an effective method of early decision of ADRs. This study provides guidelines for developing the CDM and plans to develop the drug safety surveillance system based on multi-center EMR.

Keywords Adverse drug reactions · Common data model · EMR · Pharmacovigilance · Multi-center database

1 Introduction

Pharmacovigilance is the process of defining, assessing, understanding and preventing any adverse effects of medicines [1]. The Food and Drug Administration (FDA) has operated the Adverse Event Reporting system (AERS), which is the world's largest database for the spontaneous reporting of adverse drug reactions, since the 1960s [2]. In South Korea, a spontaneous reporting system for adverse drug reactions has been used since 1985. A spontaneous reporting system, however, is limited in its ability to detect adverse drug reactions effectively [3]. A spontaneous reporting system requires more time and cost to collect the data. In addition, because this system depends mainly on voluntary reports from health care providers, the detection of adverse drug reactions may be postponed. Accordingly, drug safety monitoring based on an EMR database is becoming more important for the early detection of possible harmful adverse drug reactions.

There are some representative drug safety surveillance projects using EMR system. In the United States, the Mini-Sentinel project was initiated in 2009 to perform active surveillance of the safety of marketed medical products, including drugs, biologics, and medical devices [4]. The Mini-Sentinel project, which received funding from the US

✉ In Young Choi
iychoi@catholic.ac.kr

¹ Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

² Catholic Institute for Healthcare Management and Graduate School of Healthcare Management and Policy, The Catholic University of Korea, Seoul, Republic of Korea

³ Graduate School of Information System, The Hanyang University of Korea, Seoul, Republic of Korea

FDA [5], is a collaborative project that includes 25 institutions [6]. In the fourth quarter of 2008, the Observational Medical Outcomes Partnership (OMOP) began to improve the monitoring of drugs for safety and effectiveness. The OMOP was a public–private partnership among the FDA, academia, data owners, and the pharmaceutical industry [7]. The OMOP was funded and managed through the Foundation for the National Institutes of Health. In Europe, the European Commission funded the EU-ADR project, which began to develop an approach to the early detection of ADRs in February 2008 [8]. Four European countries, Italy, the Netherlands, the United Kingdom and Denmark, participated in the EU-ADR project [9]. These drug safety surveillance models have chosen a common data model to create harmonized input files from the distributed databases.

Some studies focused on drug safety surveillance have been conducted [10–12]; a variety of studies were required for the development of a CDM for drug safety surveillance based on multi-center EMR systems. An approach to defining a common data model is the development of robust methods to overcome the possibility that different structures and coding schemes may limit the rapid evaluation of adverse drug reactions using multi-center EMR systems [13]. Using this common data model, which expresses the functional relationships regarding each database represented, has improved abstraction for accessing data through a common interface.

While there are advanced common data models, such as the OMOP data model and the Mini-Sentinel data

model, it is difficult to apply these models within South Korea due to different treatment practices and database structures. Additionally, most of the EMR systems in Korea were developed in-house with existing structures. Furthermore, hospitals have heterogeneous EMR database structures and different terminologies, except for diagnosis codes, and it is difficult to integrate information. This heterogeneity makes it difficult to analyze information consistently [14].

Accordingly, an approach utilizing a common data model is important in developing an ADR surveillance system. After developing the CDM for Korea's hospitals, if each hospital transforms the heterogeneous information utilizing the CDM with the EMR database, the EMR data could be analyzed consistently. In addition, it would be possible to detect an ADR early. Based on these considerations, we developed the EADR CDM for early decision of ADRs based on multi-center EMR systems. In addition, we validated the EADR CDM by exporting the actual EMR data of an academic hospital and comparing the EADR CDM with the EMR data table.

2 Drug safety surveillance model

The US NIH's OMOP and the US FDA's Mini-Sentinel were started in 2008 and comply with the HIPPA security rule for security. The OMOP and the Mini-Sentinel program have looked at opportunities associated with

Table 1 List of the Mini-Sentinel CDM V2.1 tables

| No. | Table name | Key data elements |
|-----|----------------|--|
| 1 | Enrollment | Patient ID, Enrollment_start, Enrollment_end, Medical care coverage, Drug coverage |
| 2 | Demographic | Patient ID, Birth date, Sex, Race |
| 3 | Dispensing | Patient ID, Dispensing data, NDC (national drug code), RxSup (Number of days that the medication supports), RxAmt (Number of units: pills, tablets, vials) |
| 4 | Encounter | Patient ID, Encounter ID, Encounter or admission data, Discharge data, Provider Code, Facility_location, Encounter type, Facility_code, Discharge_disposition, Discharge_status |
| 5 | Diagnosis | Patient ID, Encounter ID, Encounter or admission Data, Provider code, Enc type, Diagnosis code, Orig DX, Principal Diagnosis flag |
| 6 | Procedure | Patient ID, Encounter ID, Encounter or admission data, Provider code, Enc type, Procedure code, Original procedure code |
| 7 | Death | Patient ID, Death date, Date Impute, Source, Confidence |
| 8 | Cause of death | Patient ID, Diagnosis code, Code type, Cause type, Source, Confidence |
| 9 | Laboratory | Patient ID, MS_test_name, MS_test_sub_category, Specimen_source, LOINC code, LOINC_flag, Immediacy of test, Patients location, Location of test result, Local code related to lab test results, Battery_CD, Procedure code, Code type flag, Test data, Data specimen collected, Time specimen collected, Result data, Time of test result, Result_C (Test result as short string), Modifier, Result unit, Normal_low_C (Test result as short string), Modifier_low (Modifier for value is Result_C), Normal_high_C, Modifier_high, Abnormal result indicator, Provider code, Local code, Local facility code, Data partner-specific identifier |
| 10 | Vitals | Patient ID, Measure_date, Weight, Tobacco status, Tobacco type, Diastolic blood pressure, Systolic blood pressure, Blood pressure type, Position |

implementing and using a CDM. In addition, the OMOP and the Mini-Sentinel program have shared many resources in detail regarding CDMs. Therefore, we analyzed the characteristics of the OMOP and the Mini-Sentinel data models to develop the EADR CDM.

2.1 Mini-Sentinel data model

The Mini-Sentinel program has 19 partnerships and was funded by the US FDA. The Mini-Sentinel project used a distributed database for organizing and standardizing databases. The standardized data at the individual level remained

in the local database [15]. The Mini-Sentinel program used claims data and outpatient and inpatient EMR and registries. The distributed database included administrative and claims data from 2000 to 2011 for over 300 million person-years, 2.4 billion encounters, 38 million inpatient hospitalizations, and 2.9 billion dispensed drugs [4]. In the Mini-Sentinel project, the data partners, such as Health Core, Inc., the HMO Research Network, Humana, the Kaiser Permanente Center for Effectiveness and Safety Research and Vanderbilt University, developed the Mini-Sentinel distributed data system [5]. The Mini-Sentinel project provided a CDM for organizing and standardizing the databases.

Table 2 List of the OMOP CDM V3.0 tables

| No. | Table name | Key data elements |
|-----|----------------------|---|
| 1 | Person | Person_id, Gender_concept_id, Year_of_birth, Month_of_birth, Day_of_birth, Race_concept_id, Ethnicity_concept_id, Person_location_id, Provider_id, Case_site_id, Person_source_value, Gender_source_value, Race_source_value, Ethnicity_source_value |
| 2 | Drug exposure | Drug_exposure_id, Person_id, Drug_concept_id, Drug_exposure_start_date, Drug_exposure_end_date, Drug_type_concept_id, Stop_reason, Refills, Quantity, Days supply, Sig, Prescribing_provider_id, Visit_occurrence_id, Relevant_condition_concept, Drug_source_value |
| 3 | Drug era | Drug_era_id, Person_id, Drug_concept_id, Drug_era_start_date, Drug_era_end_date, Drug_type_concept_id, Drug_exposure_count |
| 4 | Condition occurrence | Condition_occurrence, Person_id, Condition_concept_id, Condition_start_date, Condition_end_date, Condition_type_concept_id, Stop_reason, Associatted_provider_id, Visit_occurrence_id, Condition_source_value |
| 5 | Condition era | Condition_era_id, Person_id, Condition_concept_id, Condition_era_start_date, Condition_era_end_date, Condition_type_concept_id, Condition_occurrence_count |
| 6 | Visit occurrence | Visit_occurrence_id, Person_id, Visit_start_date, Visit_end_date, Place_of_service_concept, Care_site_id, Visit_source_value |
| 7 | Procedure occurrence | Procedure_occurrence, Person_id, Procedure_concept_id, Procedure_date, Procedure_type_concept_id, Associatted_provider_id, Visit_occurrence_id, Relevant_condition_concept_id, Procedure_source_value |
| 8 | Observation | Observation_id, Person_id, Observation_concept_id, Observation_date, Observation_time, Value_as_number, Value_as_string, Value_as_concept_id, Unit_concept_id, Range_low, Range_high, Observation_type_concept_id, Associated_provider_id, Visit_occurrence_id, Relevant_condition_concept_id, Observation_source_value, Units_source_value |
| 9 | Observation period | Observation_period_id, Person_id, Observation_period_start_date, Observation_period_end_date |
| 10 | Death | Person_id, Death_date, Death_type_concept_id, Cause_of_death_concept_id, Cause_of_death_source_value |
| 11 | Drug cost | Drug_cost_id, Drug_exposure_id, Paid_copay, Paid_coinsurance, Paid_toward_deductible, Paid_by_payer, Paid_by_coordination_benefits, Total_out_of_pocket, Total_paid, Ingredient_cost, Dispensing_fee, Average_wholesale_price, Payer_plan_period_id |
| 12 | Procedure cost | Procedures_cost_id, Procedure_occurrence_id, Paid_copay, Paid_coinsurance, Paid_toward_deductible, Paid_by_payer, Paid_by_coordination_benefits, Total_out_of_pocket, Total_paid, Disease_class_concept_id, Revenue_code_concept_id, Payer_plan_period_id, Disease_class_source_value, Revenue_code_source_value |
| 13 | Location | Location_id, Address_1, Address_2, City, State, Zip, Country, Location_source_value |
| 14 | Provider | Provider_id, NPI, DEA, Speciality_concept_id, Care_site_id, Provider_source_value, Speciality_source_value |
| 15 | Organization | Organization_id, Organization_type_concept_id, Location_id, Organization_source_value, Organization_type_source_value |
| 16 | Care site | Care_site_id, Location_id, Organization_id, Place_of_service_concept_id, Care_site_source_value, Place_of_service_source_value |
| 17 | Payer plan period | Payer_plan_period_id, Person_id, Payer_plan_period_start_date, Payer_plan_period_end_date, Payer_source_value, Plan_source_value, Family_source_value |
| 18 | Cohort | Cohort_id, Cohort_concept_id, Cohort_start_date, Cohort_end_date, Person_id, Provider_id, Visit_occurrence_id, Stop_reason |

The Mini-Sentinel CDM (MSCDM) contained many of the data elements necessary for medical product safety evaluations [5]. The MSCDM V1.0 consisted of eight tables. The MSCDM V2.0 added new tables, including laboratory, vitals, and summary tables, which contained revised definitions of the encounter table variables such as discharge disposition and discharge status. The MSCDM V2.0 consisted of fourteen tables.

The MSCDM V2.0 was modified in January 2012. The MSCDM V2.1 consists of ten tables that represent information on most of the key data elements required for the Mini-Sentinel program activities. We used the MSCDM V2.1 to develop the EADR CDM for multi-center EMR systems. Details of the MSCDM V2.1 are provided in Table 1.

2.2 OMOP data model

The OMOP was initiated in 2008 to conduct research activities concerning the governance, data access, technology, and methods for drug safety and benefit monitoring [7]. The goal of the OMOP was to refine the secondary use of multiple observational databases [16]. The OMOP had 39 members and was funded by the US NIH. The OMOP methods were developed for distributed analysis, and an automatic mapping procedure using a standard vocabulary was applied. This project had a dictionary of coding

guidelines for health outcomes of interest (HOI) that could be applied to electronic data such as health insurer claims and EMR databases [17]. The OMOP used the data from administrative claims and EHR.

The OMOP CDM V3.0 consisted of eighteen tables that included person, drug exposure, drug era, condition occurrence, condition era, visit occurrence, procedure occurrence, observations, observation period, death, drug cost, procedure cost, location, provider, organization, care site, payer plan period, and cohort tables [18]. Details of the OMOP CDM V3.0 are provided in Table 2.

3 Methods

The following section describes study methods in detail. First, we analyzed the data models from two prominent drug safety surveillance studies, the Mini-Sentinel project and the OMOP, to develop the EADR CDM. Second, based on a common data model between the Mini-Sentinel data model and the OMOP data model, we developed the EADR CDM. Although the Mini-Sentinel and OMOP data models are advanced models, there are limitations in directly adopting a domestic system because Korean hospitals have heterogeneous EMR databases. Third, we exported the EMR data of the S

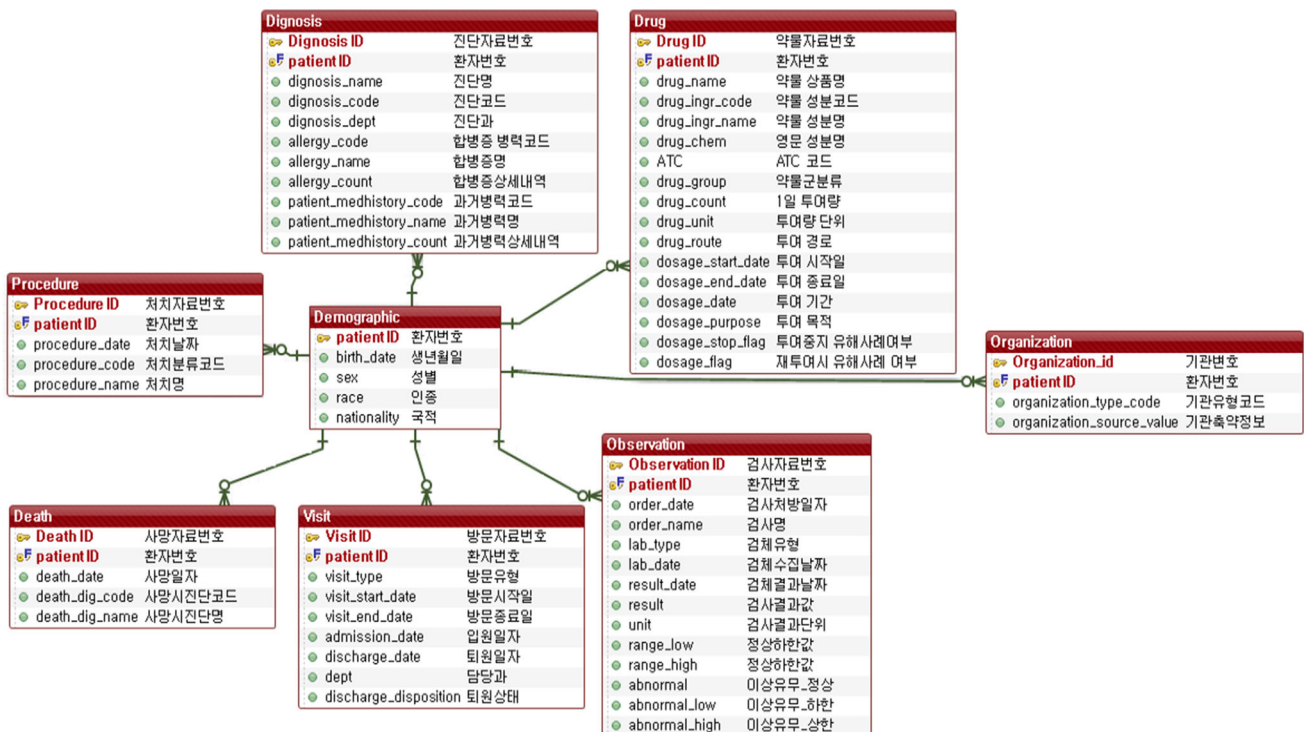


Fig. 1 ER-Diagram for the EADR CDM

hospital and converted this information based on the EADR CDM to evaluate the possibility of applying the EADR CDM to Korean hospitals. In addition, we exported data to detect ADRs according to research topic using the database that was converted based on the EADR CDM. For the data we used, the first prescription date was from January 2008 to December 2012. We counted the number of prescription based on patients' visit. The descriptive analysis was conducted using the

data extracted from 135,745 patient records. Age was calculated as 'the prescription date-birth date' in this study. We focused on six drugs that are the most popular medications in clinical practice: ranitidine, clopidogrel, rosuvastatin, ciprofloxacin, fluorouracil, and celecoxib. Through this process, we validated the EADR CDM for adverse drug reaction signal detection. In addition, we evaluated whether the exported data were sufficient to detect ADRs to the six drugs.

Table 3 Demographics, visit type, procedure, death and diagnosis tables

| Table | Column name | Values | EMR | Reference code |
|--------------------------|---|------------------------------|-----|------------------------|
| Demographics table | Patient ID | Number of patients | Y | |
| | Birth_date | YYYY/MM/DD | Y | |
| | Sex | Male/Female | Y | |
| | Race | ①White ②Asian ③Black ④Others | N | |
| | Nationality | Nationality | N | |
| Visit type table | Patient ID | Number of Patients | Y | |
| | Visit_type | Inpatient | Y | |
| | | Outpatient | | |
| | | Emergency visit | | |
| | | other visit types | | |
| | Visit_start_date | Visit start date | Y | |
| | Visit_end date | Visit end date | Y | |
| | Admission_date | Admission date | Y | |
| | Discharge_date | Discharge date | Y | |
| | Dept | Department | Y | |
| Discharge_Disposition | A = Alive (Discharge a patient from the hospital) E = Death U = Unknown | Y | | |
| Procedure table | Patient ID | Number of patients | Y | |
| | Procedure_date | Procedure date | Y | |
| | Procedure_code | Procedure code | Y | Standard code required |
| | Procedure_name | Procedure name | Y | |
| Death table | Patient ID | Number of patients | Y | |
| | Death_date | Death date | Y | |
| | Death_dig_code | Death diagnosis code | Y | ICD-10 |
| | Death_dig_name | Death diagnosis | Y | |
| Diagnosis table | Patient ID | Number of Patients | Y | |
| | Diagnosis_date | Diagnosis date | Y | |
| | Diagnosis_name | Diagnosis name | Y | |
| | Diagnosis_code | Diagnosis code | Y | ICD-10 |
| | Diagnosis_dept | Diagnosis department | Y | |
| | Allergy_code | Allergy code | N | |
| | Allergy_name | Allergy name | N | |
| | Allergy_count | Allergy details | N | |
| | Patient_medhistory_code | Patient medical history code | N | |
| | Patient_medhistory_name | Patient medical history name | N | |
| Patient_medhistory_count | Patient medical history count | N | | |

Table 4 Drug, laboratory and organization tables

| Table name | Column name | Values | EMR | Reference code | |
|---------------------------|--|--|------------------------|---|--|
| Drug table | Patient ID | Number of patients | Y | | |
| | Drug_name | Drug name | Y | | |
| | Drug_ingr_code | Drug ingredients code | Y | Standard code required | |
| | Drug_ingr_name | Drug ingredients name | Y | | |
| | Drug_chem | Drugs chemistry | Y | | |
| | ATC | ATC code | Y | | |
| | Drug_group | | ① Main category | Y | |
| | | | ② Middle category | | |
| | | | ③ Small category | | |
| | Drug_count | Amount of drug per 1 day | Y | | |
| | Drug_unit | Drug unit | Y | Standard unit required | |
| | Drug_route | Drug route | Y | Standard drug route required | |
| | Dosage_start_date | Dosage start date | Y | | |
| | Dosage_end_date | Dosage end date | Y | | |
| | Dosage_date | Medication period | Y | | |
| Dosage_purpose | Dosage purpose | N | | | |
| Dosage_stop_flag | Dosage stop flag | N | | | |
| Dosage_flag | Adverse events check for re-dosing | N | | | |
| Laboratory table | Patient ID | Number of patients | Y | | |
| | Order_date | Laboratory order data | Y | | |
| | Order_name | Laboratory order name | Y | Standard laboratory order name required | |
| | Lab_type | Laboratory type | Y | Standard code required | |
| | Lab_date | Laboratory data | Y | | |
| | Result_date | Result date | Y | | |
| | Result | Result value | Y | Standard result value required | |
| | Result unit | Result unit | Y | Standard result unit required | |
| | Low_range | Lower limit of normal | Y | | |
| | High_range | Upper limit of normal | Y | | |
| Abnormal | Abnormal check | Y | Standard code required | | |
| Organization table | Organization_ID | Number of organization | N | | |
| | Organization_type_code | ① First medical institution ② Secondary medical institution ③ Tertiary medical institution ④ etc. | N | Standard code required | |
| | | | | | |
| | | | | | |
| Organization_source_value | Abbreviated information about organization | N | | | |

4 EADR common data model

4.1 The EADR CDM

The EADR CDM consisted of eight tables, including demographics, visit type, procedure, death, diagnosis, drug, laboratory, and organization tables. Each table consisted of 5–12 variables. Figure 1 displays the entity relationship-diagram (ER-Diagram) for the EADR CDM.

The demographics table included the patient ID, birth data, gender, race, and nationality, as shown in Table 3. The visit-type table consisted of eight columns: patient ID, visit type, visit start date, visit end date, admission date, discharge date, department and discharge disposition. The procedure table included patient ID, procedure date, procedure code, and procedure name. The death table consisted of patient ID, death date, death diagnosis code and death diagnosis name. The diagnosis table included eleven

columns: patient ID, diagnosis date, diagnosis name, diagnosis code, diagnosis department, allergy code, allergy name, allergy count, medical history code, medical history name and medical history count.

As shown in Table 4, the drug table consisted of sixteen columns: patient ID, drug name, drug ingredients code, drug ingredients name, drug chemistry, ATC code, drug group, drug count, drug unit, drug route, dosage start date, dosage end date, dosage date, dosage purpose and dosage stop flag and dosage flag.

The laboratory table had eleven columns: patient ID, order date, order name, lab type, lab data, result date, observation result, result unit, low range, high range and abnormal check. Finally, the organization table included three columns: organization ID, organization type code and organization source value.

4.2 Data extraction from the EMR system to validate the EADR CDM

Among a total of 2,931,060 patients from January 2008 to December 2012 in clinical data warehouse of the S hospital, we extracted the data from 135,745 patients who were prescribed below drugs to determine whether the exported data were sufficient to detect ADRs of six drugs. The total number of prescription is 51,947,144 for the period. The total prescription of six drugs is 1,449,376. We focused on six drugs that are the most popular medications in clinical practice: ranitidine, clopidogrel, rosuvastatin, ciprofloxacin, fluorouracil, and celecoxib. Through this process, we validated the EADR CDM for adverse drug reaction signal detection.

There were issues to resolve when converting the data based on the EADR CDM.

In the demographics table, there were three columns that included patient ID, birth data and gender in the EMR system of the S hospital. All the columns of visit type table were in the EMR system of the S hospital. Regarding the visit type, there were different codes depending on admission, outpatient clinic, emergency medicine and others. In the discharge disposition column, there were only those data that related to death in the hospital. If the patient died after being discharged from the hospital, it was difficult to trace the effects of drugs. All the columns in the procedure table were in the EMR system of the S hospital. All the columns of the death table, which were patient ID, diagnosis date, diagnosis name and diagnosis code, were also in the EMR system of the S hospital. The reference code for the death and diagnosis code was ICD-10. Among the eleven columns of the diagnosis table, five columns, patient ID, diagnosis date, diagnosis name, diagnosis code, and diagnosis department, were in the EMR system of the S hospital. Thirteen columns of the drug table, patient ID, drug name, drug ingredients code, drug ingredients name, drug chemistry, ATC code, drug group, drug count, drug unit, drug route, dosage start date, dosage end date, and dosage date, were in the EMR system of the S hospital. The dosage purpose, dosage stop flag, and dosage flag were not included in the EMR system of the S hospital. Although there were standard terms for the drug, there were different codes for the drug name depending on each hospital. In the case of the drug group, there was no standard of classification. Finally, all the columns in the organization table were not in the EMR system of the S hospital.

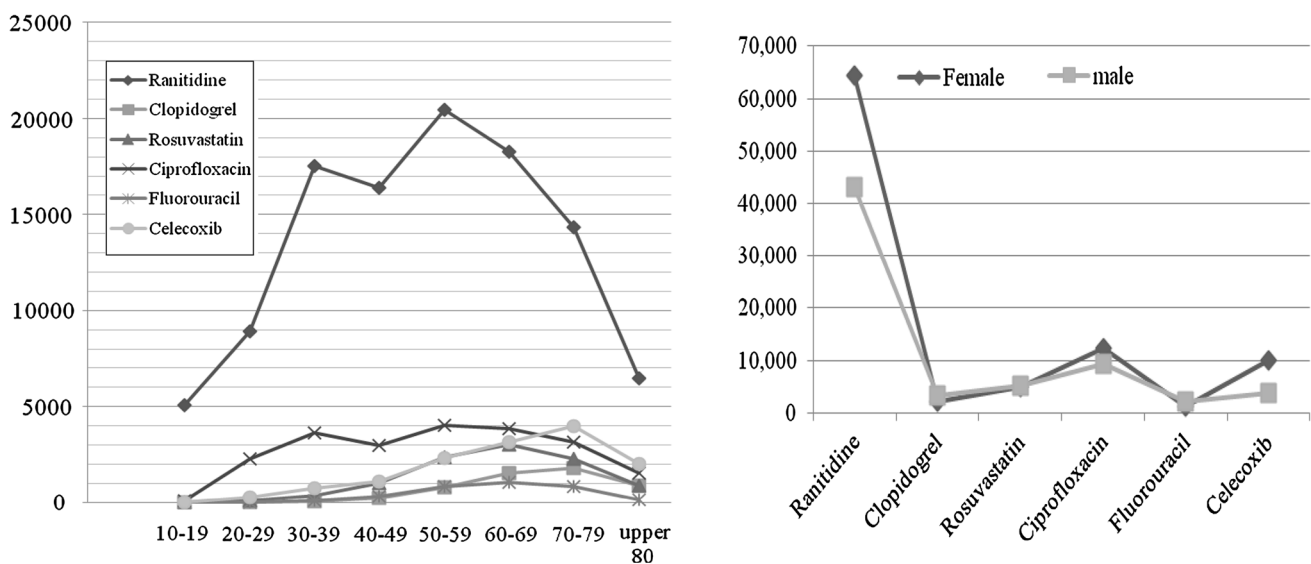


Fig. 2 Prescription status of six drugs according to age and gender

Table 5 Prescription distribution of six drugs according to age and gender

| Characteristic | ATC | | | | | |
|----------------|---------------|-------------|--------------|---------------|--------------|-------------|
| | Ranitidine | Clopidogrel | Rosuvastatin | Ciprofloxacin | Fluorouracil | Celecoxib |
| <i>Age</i> | | | | | | |
| 10–19 | 5065 (5 %) | 2 (0 %) | 8 (0 %) | 105 (0 %) | 7 (0 %) | 27 (0 %) |
| 20–29 | 8934 (8 %) | 6 (0 %) | 79 (1 %) | 2296 (11 %) | 19 (1 %) | 289 (2 %) |
| 30–39 | 17,530 (16 %) | 64 (1 %) | 375 (4 %) | 3618 (17 %) | 100 (3 %) | 756 (6 %) |
| 40–49 | 16,411 (15 %) | 226 (4 %) | 1028 (10 %) | 2997 (14 %) | 330 (10 %) | 1108 (8 %) |
| 50–59 | 20,461 (19 %) | 771 (15 %) | 2347 (23 %) | 4035 (19 %) | 845 (26 %) | 2328 (17 %) |
| 60–69 | 18,286 (17 %) | 1515 (29 %) | 3022 (30 %) | 3871 (18 %) | 1034 (31 %) | 3156 (23 %) |
| 70–79 | 14,332 (13 %) | 1795 (34 %) | 2268 (23 %) | 3164 (15 %) | 817 (25 %) | 3980 (29 %) |
| upper 80 | 6453 (6 %) | 891 (17 %) | 870 (9 %) | 1538 (7 %) | 144 (4 %) | 2030 (15 %) |
| Total | 107,472 | 5270 | 9997 | 21,624 | 3296 | 13,674 |
| <i>Gender</i> | | | | | | |
| Female | 64,410 (60 %) | 2068 (39 %) | 4922 (49 %) | 12,339 (57 %) | 1219 (37 %) | 9935 (73 %) |
| Male | 43,062 (40 %) | 3202 (61 %) | 5075 (51 %) | 9285 (43 %) | 2077 (63 %) | 3739 (27 %) |
| Total | 107,472 | 5270 | 9997 | 21,624 | 3296 | 13,674 |

Bracket percent

We obtained results regarding the characteristics of the actual exported EMR data of an academic hospital. We used the data from a total of 135,745 patient records. The most prescribed drug was ranitidine, as shown in Fig. 2.

By age, ranitidine was prescribed more to patients in their fifties ($n = 20,461$, 19 %). Clopidogrel was prescribed more to those in their seventies ($n = 1795$, 34 %). In the case of clopidogrel, this drug is prescribed increasingly for the elderly because of it is an antithrombotic drug [19]. In our result, clopidogrel was also prescribed increasingly for the elderly and especially to males. These results have features similar to the general prescription features of clopidogrel. Rosuvastatin was prescribed more to patients in their sixties ($n = 3022$, 30 %). Ciprofloxacin was prescribed more to those in their fifties ($n = 4035$, 19 %). Fluorouracil was prescribed more to patients in their sixties ($n = 1034$, 31 %). Celecoxib was prescribed more to those in their seventies ($n = 3980$, 29 %).

According to gender, ranitidine was prescribed more to females (64,410, 60 %). Clopidogrel was prescribed more to males (3202, 61 %). Rosuvastatin was prescribed with similar frequencies to females and males. Ciprofloxacin was prescribed more to females ($n = 12,339$, 57 %). Fluorouracil was prescribed more to males (2077, 63 %). Finally, celecoxib was prescribed more to females (9935, 73 %) as shown in Table 5.

5 Conclusions

We developed the EADR CDM for the early decision of ADRs, and it can be extended to multiple EMR systems. The EADR CDM included eight tables that contain

demographic, drug, visit, procedure, diagnosis, death, laboratory and organization information. The EADR CDM was implemented using an authentic EMR system within the S hospital and provided the following concerns and issues for further study.

Since race and nationality were not included in the EMR of the S hospital, the current EMR system cannot detect different effects or side effects according to race or nationality. If race and nationality are important factors, then this information should be captured at the point of care in the hospital for a straightforward analysis.

Standard codes were found to be major concerns in integrating multiple EMR databases. For example, in the visit type table, there is no national standard code for visit type. Because the S hospital used internal codes, we transformed the visit data to our standard codes, such as inpatient, outpatient, emergency visit and other visit type. The OMOP project proposed using SNOMED-CT for the procurement data. However, many Korean hospitals do not use SNOMED code for procedures. Therefore, standard procedure codes are essential when using the EADR CDM.

Regarding the drug table, there were many issues when the data related to the drug code were extracted because hospitals had different codes for the drug name, drug ingredients, drug ingredients name, and drug chemistry. Some drugs did not map onto the ATC code. The drug code should be a standardized code and map onto the ATC code.

In the case of the laboratory table, because each hospital used different codes and units for order name, laboratory type, result, result unit and abnormal check, standardization would be required. Free text is another major concern. The medical history was written in a free text format, and thus,

the medical history should be re-coded for data analysis. To detect ADR signals in multi-center EMR systems, the development of standardized guidelines regarding laboratory codes is necessary for a multi-institutional pharmacovigilance database system. In addition, there were character type data in many columns when the data related to the observation result were extracted. Accordingly, data cleaning was absolutely necessary. For this purpose, an effective data cleaning tool should be developed.

The quality of the data in the EMR system is not satisfactory from a research perspective. When we extracted the data from the EMR database, we found that there were many missing values. Most of the medical history was missing or incomplete. To resolve these problems, guidelines for medical record should be prepared with the cooperation of a pharmacovigilance team.

Although the reasons for stopping a medication or changing the dose are important information, that information was not collected from the EMR system of the hospital. Accordingly, systems for the dosage stop flag and dosage flag must be developed to indicate the occurrence of adverse drug events.

In our results, all the columns of the organization table were not in the EMR system of the hospital. Accordingly, standard organization type codes are essential to use the EADR CDM. In addition, there were some problems when we extracted the data concerning hospitalization history because the hospitals used different names for medical departments. In addition, a single code sometimes mapped onto several medical departments. Therefore, a standard organization type code is required to use the EADR CDM based on multi-center EMR systems.

Controlled terminology will be applied to integrate different EMR systems. Further efforts for the standardization of procedure and laboratory codes will be required to develop a multi-institutional pharmacovigilance database system.

We extracted the data from 135,745 patient records to evaluate whether the exported data were sufficient to detect ADRs to the six most popular drugs. After exporting the data, we found that a diverse database-based multi-center EMR system is required, although certain drugs such as clopidogrel have relevant data to detect ADRs within one system. More research into the characteristics of ADRs and the causes of side effects in elderly patients is required [20]. Drug safety monitoring based on an EMR database provides information to detect the characteristics of ADRs and the causes of side effects in elderly patients. In addition, there is little information on drug safety for pediatric patients and inadequate scientific evidence to prove safety. An ADR surveillance system based on multi-center EMR databases using the EADR CDM will aid in finding the scientific evidence necessary for adequate proof.

Accordingly, the collection of diverse data based on multi-center EMR systems using the EADR CDM is a method to effectively detect ADRs. Drug safety monitoring based on EMR databases is increasingly important for the early detection of possible harmful adverse drug reactions.

This study had some limitations. We compared the EADR CDM with the EMR data of the S hospital, i.e., we extracted the data from the EMR system to validate the EADR CDM. However, the EADR common data model is intended for adverse drug signal detection based on multi-center EMR systems. Therefore, future studies should be conducted focused on multi-center EMR databases, such as in Overhang et al. [18] in which ten different datasets were converted into the OMOP CDM to validate the OMOP CDM. Despite these limitations, we provide guidelines for developing the CDM for early decision of ADRs and make plans to develop the drug safety surveillance system based on multi-center EMR systems.

Acknowledgments This study was supported by a grant of the Korea Health technology R&D Project, Ministry of Health & Welfare, Republic of Korea (A112022).

References

1. World Health Organization (2002) The importance of pharmacovigilance? Safety monitoring of medicinal products, World Health Organization, Geneva. <http://apps.who.int/medicinedocs/pdf/s4893e/s4893e.pdf>. Accessed 6 Mar 2013
2. Rodriguez EM, Staffa JA, Graham DJ (2001) The role of databases in drug postmarketing surveillance. *Pharmacoepidemiol Drug Saf* 10(5):407–410
3. Jha AK, Kuperman GJ, Teich JM, Leape L, Shea B, Eve R, Burdick E, Seger DL, Vliet MV, Bates DW (1998) Identifying adverse drug events development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *J Am Med Inform Assoc* 5:305–314
4. Platt R, Carnahan RM, Brown JS, Chrischilles E, Curtis LH, Hennessy S, Nelson JC, Racoosin JA, Robb M, Schneeweiss S, Toh S, Weiner MG (2012) The U.S. food and drug administration's mini-sentinel program: status and direction. *Pharmacoepidemiol Drug Saf* 21:1–8
5. Curtis LH, Weiner MG, Boudreau DM, Cooper WO, Daniel GW, Nair VP, Raebel MA, Beaulieu NU, Rosofsky R, Woodworth TS, Brown JS (2012) Design considerations, architecture, and use of the mini-sentinel distributed data system. *Pharmacoepidemiol Drug Saf* 21:23–31
6. Forrow S, Campion DM, Herrinton LJ, Nair VP, Robb MA, Wilson M, Platt R (2012) The organizational structure and governing principles of the Food and Drug Administration's Mini-Sentinel pilot program. *Pharmacoepidemiol Drug Saf* 21:12–17
7. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, Welebob E, Scarnecchia T, Woodcock J (2010) Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med* 153:600–606
8. Trifiro G, Fourier-Reglat A, Sturkenboom MC, Díaz Acedo C, Lei JVD (2009) The EU-ADR project: preliminary results and perspective. *Stud Health Technol Inform* 148:43–49

9. Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, Mazzaglia G, Giaquinto C, Corrao G, Pedersen L, Van Der Lei J, Sturkenboom M (2011) Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf* 20:1–11
10. Hwang SH, Kim EY, Lee YS, Jung SY, Lee YM, Son KH, Choi KU, Lee SH, Kim Y (2005) Implementation and evaluation of the computerized surveillance system to identify adverse drug events: pilot study. *J Korean Soc Health Syst Pharm* 22(2):118–136
11. Lee YH, Yoon YM, Lee BM, Hwang HJ, Kang UG (2009) Development of mining model through reproducibility assessment in adverse drug event surveillance system. *Korean Soc Comput Inf* 1(3):183–192
12. Lee YH, Kang UG, Park RW (2009) Development of adverse drug event surveillance system using BI Technology. *J Korean Contents* 9(2):106–114
13. Le HV, Beach KJ, Powell G, Pattishall ED, Ryan P, Mera RM (2013) Performance of a semi-automated approach for risk estimation using a common data model for longitudinal healthcare databases. *Stat Methods Med Res* 22(1):97–112
14. Kim HS, Cho H, Lee IK (2011) Design and development of an EHR platform based on medical informatics standards. *J Fuzzy Log Intell Syst* 21:456–462
15. Cook AJ, Tiwari RC, Wellman RD, Heckbert SR, Li L, Heagerty P, Marsh T, Nelson JC (2012) Statistical approaches to group sequential monitoring of postmarket safety surveillance data: current state of the art for use in the Mini-Sentinel pilot. *Pharmacoepidemiol Drug Saf* 21:72–81
16. Stang PE, Ryan PB, Dusetzina SB, Hartzema AG, Reich C, Overhage JM, Racoosin JA (2012) Health outcomes of interest in observational data: issues in identifying definitions in the literature. *Health Outcomes Res Med* 3(1):e37–e44
17. Murphy SN, Castro V, Colecchi J, Dubey A, Gainer V, Herrick C, Sordo M (2011) Partners HealthCare OMOP Study Report. Foundation for the National Institutes of Health Observational Medical Outcomes Partnership Partners HealthCare Jan 10
18. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE (2012) Validation of a common data model for active safety surveillance research. *Am Med Inform Assoc* 19:54–60
19. Sung ES, Kim TH, Cho SH, Kim KR, Park CW, Jeong JH (2012) Epistaxis in patients taking oral anticoagulant and antiplatelet medication. *Korean J Otorhinolaryngol Head Neck Surg* 55(5):290–294. doi:[10.3342/kjorl-hns.2012.55.5.290](https://doi.org/10.3342/kjorl-hns.2012.55.5.290)
20. Lee MW, Lee JS, Han OY, Choi IY, Jeong SH, Yim HW, Lee DG, La HO, Park YM (2014) Study for association between adverse drug reactions and causative drugs in the elderly using data-mining analysis. *Korean J Clin Pharm* 24(1):39–44