Development Common Data Model for Adverse Drug Signal Detection based on multi-center EMR systems

Mi Jung, Rho
Graduate School of Healthcare Management and Policy,
The Catholic University of Korea
E-mail) rhomijung@gmail.com

Si Ra, Kim
Graduate School of Healthcare Management and Policy,
The Catholic University of Korea
E-mail) sira8805@naver.com

Seung Ho, Park
Graduate School of Information System,
The Hanyang University of Korea
E-mail) along0m0w@naver.com

Kwang Soo, Jang
Graduate School of Information System,
The Hanyang University of Korea
E-mail) jks8605@nate.com

Bum Joon, Park
Graduate School of Information System,
The Hanyang University of Korea
E-mail) blackss2@nate.com

In Young Choi (*Corresponding author*)
Graduate School of Healthcare Management and Policy,
The Catholic University of Korea
E-mail) iychoi@catholic.ac.kr

***Abstract*—The drug safety monitoring based on EMR system is able to collect more objective pharmacovigilance data and analyze adverse drug reaction (ADR) earlier than spontaneous ADR reporting. This study developed the Korea ADR common data model (K-ADR CDM) for early detection of adverse drug reaction which is feasible for Korean EMR systems. To do that, we analyzed previously studied data model from two prominent drug safety surveillance researches: Mini-Sentinel data model and Observational Medical Outcomes Partnership (OMOP) data model. The K-ADR CDM of eight tables which contain demographic table, drug table, visit table, procedure table, diagnosis table, death table, laboratory table and organization table. Each table consists of 5~12 fields. In addition, controlled terminology will be applied to integrate different EMR systems. To validate the data model of K-ADR, EMR data of S hospital was exported and mapped with the K-ADR. Further efforts for the standardization of procedure code and laboratory code will be needed for multi-institutional pharmacovigilance database system. The pharmacovigilance activity based EMR system will be cost-effective method to detect ADR signals.**

Keywords—Adverse Drug Signal Detection; Common Data Model; EMR; Pharmacovigilance

I. INTRODUCTION

Pharmacovigilance is process to define, assess, understand and prevent adverse effects of medicines [1]. Food and Drug Administration (FDA) has operated the Adverse Event Reporting system (AERS) which is the world's largest database for spontaneous reporting of adverse drug reactions since the 1960's [2]. In South Korea, spontaneous adverse drug reactions reporting system has been used since 1985. Spontaneous reporting system, however, has limitation to detect adverse drug reactions effectively [3]. Spontaneous reporting system need to allow more time and cost for collecting the data. In addition, because this system depends mainly on health care providers' voluntary report, it could postpone detecting the adverse drug reactions. Accordingly, drug safety monitoring based on EMR is getting more important for early detection of possible harmful adverse drug reactions.

There are drug safety surveillance projects using electronic health records (EHR). In the United States, the Mini-Sentinel project was initiated to perform active surveillance of the safety of marketed medical products, including drugs, biologics, and medical devices in 2009 [4]. The Mini-Sentinel, received the funding from the US FDA [5], is a collaborative projects including 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 EU-ADR project started to develop approach to the early detection of ADR from February 2008, funded by the European Commission [8]. Four European countries, which were 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 file from the distributed database.

In South Korea, many hospitals have heterogeneous EMR databases. There are a variety of data types and heterogeneous information. Therefore, the EMR date could not be analyzed consistently [10]. In addition to, there was an increased demand for development surveillance system based on EMR system.

In order to build a drug safety surveillance system, integration with different EMR systems was required. To incorporate multi-center EMR systems, common data model

978-1-4799-0604-8/13/$31.00 ©2013 IEEE

should be developed. The common data model (CDM) could integrate the data from the different EMR systems. Through the CDM, we could derive the reliable results for early detection of adverse drug reaction.

Although many studies have been conducted focused on drug safety surveillance, there were limited domestic researches related to drug safety surveillance system [11-13]. Previous domestic researches based on single EMR system. Therefore, a variety studies, related to development K-ADR CDM for drug safety surveillance system based on multi-center EMR systems, was required.

This study attempted to develop K-ADR CDM for multi-center EMR systems. In addition to, we validated the K-ADR CDM by exporting real EMR data of an academic hospital and comparing the K-ADR CDM with EMR data table

II. DRUG SAFETY SURVEILLANCE MODEL

To develop K-ADR CDM for multi-center EMR systems, we analyzed previously studied data model from two prominent drug safety surveillance researches: Mini-Sentinel data model and Observational Medical Outcomes Partnership (OMOP) data model.

A. Mini-Sentinel data model

Mini-Sentinel used a distributed database. Standardized individual level data remained the local database [14]. 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 dispensing [4]. In Mini-Sentinel project, data partners developed the Mini-Sentinel distributed data system such as Health Core, Inc.; the HMO Research Network; Humana; the Kaiser Permanente Center for Effectiveness and Safety Research; and Vanderbilt University [5].

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

The MSCDM V2.0 modified in January 2012. The MSCDM V2.1 consisted of ten tables that represent information for most of the key data elements needed for Mini-Sentinel activities. This study used the MSCDM V2.1 for development K-ADR CDM for multi-center EMR systems. Detail of the MSCDM V2.0 was provided in Table 1.

TABLE I. MINI-SENTINEL CDM V2.1 TABLE LIST

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, Hispanic, 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 Dt, DtImpute, 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, Teat 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, Tabacco status, Tobacco type, Diastolic blood pressure, Systolic blood pressure, Blood pressure type, Position

B. OMOP data model

The OMOP was initiated to conduct research activities about the governance, data access, technology, and methods for drug safety and benefit monitoring in 2008 [7]. The goal of OMOP was to refine the secondary use of multiple observational databases [15]. The OMOP methods were developed for distributed analysis, has applied an automatic mapping procedure using a standard vocabulary. 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 EMRs [16].

The OMOP CDM V3.0 consisted of eighteen tables that included person, drug exposure, drug era, condition occurrence, condition era, visit occurrence, procedure occurrence, observation, observation period, death, drug cost, procedure cost, location, provider, organization, care site, payer plan period, cohort [17]. Detail of the OMOP CDM V3.0 was provided in Table 2.

TABLE II. OMOP CDM V3.0 TABLE LIST

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, Ethinicity_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_occurance_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_occurance, Person_id, Condition_concept_id, Condition_start_date, Condition_end_date, Condition_type_concept_id, Stop_reason, Associatted_provider_id, Visit_occurance_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_occurance_count
6	Visit Occurrence	Visit_occurance_id, Person_id, Visit_start_date, Visit_end_date, Place_of_service_concept, Care_site_id, Visit_source_value
7	Procedure Occurrence	Procedure_occurance, Person_id, Procedure_concept_id, Procedure_date, Procedure_type_concept_id, Associatted_provider_id, Visit_occurance_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_occurance_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, Dispending_fee, Average_wholesale_price, Payer_plan_period_id
12	Procedure Cost	Procesure_cost_id, Procedure_occurance_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

III. K-ADR COMMON DATA MODEL

A. Development of K-ADR CDM and compare with EMR systems

K-ADR CDM consisted of eight tables including demographics, visit type, procedure, death, diagnosis, drug, laboratory, organization. Figure 1 is the entity relationship-diagram (ER-Diagram) for K-ADR CDM.

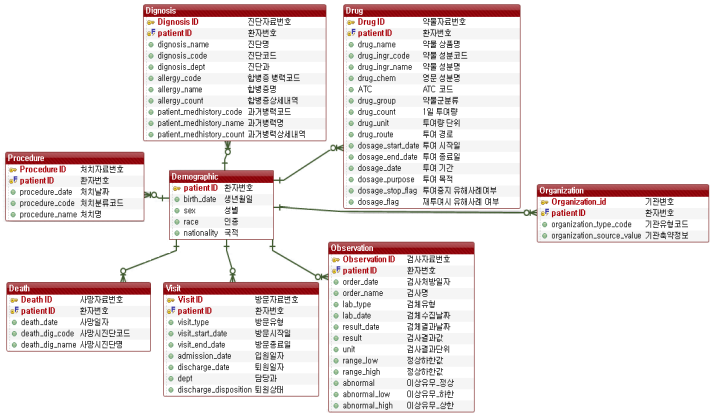

Fig. 1. ER-Diagram for K-ADR CDM

First, demographics table included patient ID, birth data, gender, race, and nationality. Second, visit type table consisted of eight columns: patient ID, visit-type, visit start data, visit-end date, admission-date, discharge-date, department, and discharge-disposition. Third, procedure table include patient ID, procedure-date, procedure-code, and procedure-name. Forth, death table consisted of patient ID, death-date, death-diagnosis-code, and death-diagnosis name. Fifth, diagnosis table included eleven columns: patient ID, diagnosis-date, diagnosis-name, diagnosis-code, diagnosis-dept, allergy-code, allergy-name, allergy-count, medical history code, medical history name, medical history count.

TABLE III. TABLES OF DEMOGRAPHICS, VISIT TYPE, PROCEDURE, DEATH AND DIAGNOSIS

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 ④ Others	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_dat	Visit start date	Y	

	e			
	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)	Y	
		E=Death		
		U=Unknown		
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	

Sixth, 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, dosage-stop flag, and dosage flag. Seventh, laboratory table were eleven columns: patient ID, order date, order name, lab type, lab data, result date, observation result, result unit, range low, range high, abnormal check. Eighth, organization table included three columns: organization ID, organization type code, and organization source value.

TABLE IV. TABLES OF DRUG, LABORATORY AND ORGANIZATION

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

	Range_low	Lower limit of normal	Y	
	Range_high	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	N	* Standard code required
		② Secondary medical institution		
		③ Tertiary medical institution		
		④ etc ( )		
	Organization_source_value	Abbreviated information about organization	N	

B. Data extraction of EMR system to map with K-ADR CDM

To validate the K-ADR CDM for adverse drug signal detection, EMR data of S hospital was exported and mapped with the K-ADR. First, in demographics table, there were three columns such as patient ID, birth data and gender in EMR system of S hospital. Race and nationality, however, were not included in EMR system of S hospital. Second, all columns of visit type table were EMR system of S hospital. In case of visit type, there were different code depending on admission, out-clinic, emergency medicine, and others. In column of discharge disposition, there were only data related to death in hospital. If patient died after discharging from hospital, it was difficult to trace of drugs. Third, all columns of procedure table were in EMR system of S hospital. The standard code for procedure-code would be essential to use K-ADR CDM. Forth, all columns of death table, which were patient ID, diagnosis-date, diagnosis-name and diagnosis-code, also were EMR system of S hospital. Reference code for death-dig-code was ICD-10. Fifth, among the eleven columns of diagnosis table, five columns such as patient ID, diagnosis-date, diagnosis-name, diagnosis-code, diagnosis-dept were in EMR system of S hospital. Reference code for diagnosis-code was ICD-10. Sixth, thirteen columns of drug table such as 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 were in EMR system of S hospital. Dosage-purpose, dosage-stop flag, dosage flag were not included in EMR system of S hospital. Although there were standard terms of drug, it was different code of drug name depending on each hospital. In case of drug group, there was no standard of classification. Seventh, although all columns of laboratory table were in EMR system of S hospital, standard code and unit were required. Finally, organization table was important to use data in multiple- center EMR systems. However, all columns of organization table were not in EMR system of S hospital.

C. Data result from EMR system extraction

We selected six drugs which is most popular medication in clinical practice; ranitidine, clopidogrel, rosuvastatin, ciprofloxacin, fluorouracil, and celecoxibm. Relevant data were extracted from EMR system of S hospital. The descriptive analysis was conducted using extracted drugs data.

1) Age

The total number of prescription was 672,753. First, among total prescription, 60% (total 402,659) was prescribed for the fifties -seventies. Ranitidine was prescribed more for the fifties (total 87,908, 20 %). Clopidogrel was prescribed more for the seventies (total 15,983, 35%). Rosuvastatin was prescribed more for the seventies (total 5,691, 30%). Ciprofloxacin was prescribed more for the fifties (total 24,827, 24%). Fluorouracil was prescribed more for the sixties (total 16,125, 33%). Finally, celecoxib was prescribed more for the seventies (total 5,350, 32%)

TABLE V. THE NUMBER OF PRESCRIPTION DEPENDING ON AGE

ATC / Age	Ranitidine	Clopidogrel	Rosuvastatin	Ciprofloxacin	Fluorouracil	Celecoxib	Total
10-19	34,389	62	17	801	147	8	36,281
20-29	24,834	95	469	10,396	272	215	35,424
30-39	41,416	470	644	13,849	1,441	734	60,125
40-49	51,681	1,816	1,901	18,136	4,757	1,304	58,554
50-59	1) 87,908	6,125	3,021	1) 24,827	12,624	2,707	79,595
60-69	84,260	11,819	4,423	20,066	1) 16,125	3,437	125,172
70-79	75,304	1) 15,983	1) 5,691	10,389	12,455	1)5,350	140,130
Upper 80	37,384	9,636	3,009	5,212	1,755	3,129	137,212
Total	437,176	46,006	19,175	103,676	49,576	16,884	672,493

2) Gender

Among total prescription, 49 % (total 328,086) was prescribed for female, 51% (total 344,667) the rest was prescribed for male. Ranitidine and ciprofloxacin were prescribed similarly for female and male. Clopidogrel was prescribed more for male (total 28,960, 63%). Rosuvastatin was prescribed more for male (total 11,033, 58%). Fluorouracil was prescribed more for male (total 32,530, 66%). Finally, celecoxib was prescribed more for female (total 10,994, 65%).

TABLE VI. THE NUMBER OF PRESCRIPTION DEPENDING ON GENDER

ATC / Gender	Ranitidine	Clopidogrel	Rosuvastatin	Ciprofloxacin	Fluorouracil	Celecoxib	Total
Female	223,398	7,046	8,142	51,352	17,046	10,994	317,978
male	213,778	28,960	11,033	52,324	32,530	5,890	344,515
Total	437,176	46,006	19,175	103,676	49,576	16,884	662,493

3) Department

The most prescribed drug was ranitidine. The number of ranitidine prescription was 437,176(65%). There were results about drug prescriptions depending on diagnostic department. The total number of drug prescriptions depending on diagnostic department was 2,872,382. The top five diagnostic departments were general surgery, hemato-oncology, emergency medicine,

gynecology, and blood & marrow transplantation. The total number of prescription that belonged to the top five diagnostic departments was 1,959,596 (68%). In hemato-oncology, ranitidine was prescribed more (total 285,032, 18%), celecoxib was prescribed more in hemato-oncology (total 24,870, 32%). Clopidogrel was prescribed more in cardiology (total 59,227, 41 %). Rosuvastatin was prescribed more in cardiology (total 17,403, 25%). Ciprofloxacin was prescribed more in emergency medicine (total 113,209, 29%). In general surgery, fluorouracil was prescribed more (total 368,393, 60%).

TABLE VII. THE NUMBER OF PRESCRIPTION DEPENDING ON DIAGNOSTIC DEPARTMENT

No.	ATC / Diagnostic department	Ranitidine	Clopidogrel	Rosuvastatin	Ciprofloxacin	Fluorouracil	Celecoxib	Total
1	General surgery	151,101	7,219	4,801	36,605	368,393	3,804	571,923
2	Hemato-oncology	1) 285,032	9,848	5,067	41,125	186,426	2) 24,870	552,368
3	Emergency medicine	257,124	14,693	11,682	1) 113,209	1) 20,351	9,205	426,264
4	Gynecology	205,263	323	157	11,097	3,172	1,926	221,938
5	Blood & marrow transplantation	82,397	351	1,066	101,151	104	2,034	187,103
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
38	Nuclear medicine	6	-	-	-	-	-	6
Total		1,562,668	144,306	70,296	397,426	618,987	78,899	2,872,382

4) Laboratory test

Finally, the seven drugs were checked by 41 laboratory tests. The total number of prescription was 10,688,268. The top four laboratory tests were hemoglobin (total 569,502, 5.3%), hemotocrit (total 569,396, 5.3%), platelet (total 566,056, 5.3%), white blood cell (total 565,099, 5.3%) and eosinophil (total 544,548, 5.1%).

TABLE VIII. THE NUMBER OF PRESCRIPTION DEPENDING ON LABORATORY TEST

No.	Code	Test name	Total
1	LHR102	Hemoglobin	569,502
2	LHR103	Hemotocrit	569,396
3	LHR104	Platelet	566,056
4	LHR100	White blood cell	565,099
5	LHR10505	Eosinophil	544,561
. . .	. . .	. . .	. . .
41	LEE101	Myogloin	199
Total			10,688,268

IV. CONCLUSION

We developed the K-ADR CDM for early detection of adverse drug reaction which was feasible for Korean EMR systems. The K-ADR CDM included eight tables which contain demographic table, drug table, visit table, procedure table, diagnosis table, death table, laboratory table and organization table. Each table consisted of 5~12 fields. In addition, terminology standard such as ICD-10 and WHO-ART would be provided to integrate multiple EMR systems.

After comparing the K-ADR common data model with EMR data of S hospital, it should be concluded as follows.

First, in demographics table, race and nationality were not included in EMR of S hospital. In case of drug, effect or side effect may be different depending on race and nationality. In addition, as many foreigners immigrated to South Korea, South Korea became multicultural society. Therefore, race and nationality should be collected from EMR system in hospital. The K-ADR CDM has developed to detect adverse drug signal based on multi-institutional pharmacovigilance database system. For multi-institutional pharmacovigilance database system, the reference code for race and nationality will be required.

Second, in case of visit type, there are different code depending on admission, out-clinic, emergency medicine, others. Therefore, different visit type code must be unified. In case of patients' death after discharging from hospital, method is needed to trace of drugs.

Third, although all four columns of procedure table were included in EMR of S hospital, the standard code for procedure-code will be essential to use K-ADR CDM. Accordingly, further efforts for development of the standardized guidelines about procedure code would be needed for multi-institutional pharmacovigilance database system.

Fourth, among the diagnosis table, although there were items for allergy-code and allergy-name, it is difficult to collect data. Therefore, allergy-code and allergy-name must be collected. The patient_medhistory_name and patient_medhistory_count have been collected in the form of free text, it should be re-coded to analysis. Accordingly, these columns were needed coding work to standardize. In addition, when we extracted the data from EMR, we found that there was column without diagnosis name during the data preprocessing for inpatients' first diagnosis name. Column without diagnosis name should be reconfirmed by principal diagnosis code. In addition to, there were multiple diagnosis codes or principal diagnosis to the one patient. To solve these problems, guideline for these problems would be prepared.

Fifth, in drug table, there were many issues. Although there were standard terms of drug, it was different code of drug name depending on each hospital. Accordingly each hospital must change from drug name to drug-ingredients code. The standard code for drug-ingredients code, drug-unit, drug-route and drug-group would be necessary to use K-ADR CDM. Especially, when drug-route was well recorded, it was easy to map the data. To standard code for drug-ingredients code, drug-unit, drug-route and drug-group, regulation would be required. In addition to, although dosage stop flag and dosage flag were important columns, there were not in EMR system of S hospital. In case of Japan, EMR system provided pop-up window service to notice dosage stop flag. Accordingly, systems for dosage stop flag and dosage flag must be developed. While we extracted the data from EMR, it was difficult to identify dosage data exactly, because dosage data were different depending on prescription data and nursing records data. From the advice of doctors, we could confirm that nursing records data was more reasonable. In addition, there were some problems when data related to drug code was

extracted, because hospitals had a different drug-name, drug-ingredients code, drug-ingredients name, drug chemistry. Some drugs did not map on the ATC code. Therefore, drug code should be standardized code, map on the ATC code.

Sixth, in case of laboratory table, because each hospital has used different code and unit such as order name, laboratory type, result, result unit and abnormal check would be required. Although the standard for limit of normal was very important, each hospital have used different standard. Therefore, to detect ADR signals on multiple-center EMR systems, central laboratory should manage different standard. Development of the standardized guidelines about laboratory code would be needed for multi-institutional pharmacovigilance database system. In addition, there were character type data in many columns when the data related to observation result was extracted. Accordingly, data cleaning was positively necessary. To do that, effective data cleaning tool should be developed.

Seventh, Organization table was important to use data in multiple-center EMR systems. However, all columns of organization table were not in EMR system of S hospital. Accordingly, the standard code for organization type code would be essential to use K-ADR CDM. In addition, there were some problems when we extracted the data about hospitalization history, because hospitals had a different name of medical departments. In addition to, the one code sometimes has mapped on several medical departments. Therefore, the standard code for organization type code would be required to use K-ADR CDM based on multi-center EMR systems.

This study had some limitation. Overhang et al. (2012) converted ten different dataset into the OMOP CDM, to validate the OMOP CDM [17]. We compared the K-ADR CDM with EMR data of S hospital, extracted data from EMR system to validate the K-ADR CDM. However, the K-ADR common data model was for adverse drug signal detection based on multi-center EMR systems. Therefore, future study should be conducted focused on multi-center EMRs such as Overhang et al. (2012).

This study would provide guideline to develop the method in order to detect ADR signals. The pharmacovigilance activity based EMR would be cost-effective method to detect ADR signals.

REFERENCES

[1] World Health Organization, “The Importance of Pharmacovigilance ? Safety Monitoring of Medicinal Products.” Geneva: World HealthOrganization; 2002. Available from: http://apps.who.int/medicinedocs/pdf/s4893e/s4893e.pdf. Accessed March 6, 2013.

[2] E.M. Rodriguez, J.A. Staffa, D.J. Graham, “The role of databases in drug postmarketing surveillance,” Pharmacoepidemiol Drug Saf., vol. 10, no. 5, pp. 407-410, 2001.

[3] A.K. Jha, G. J. Kuperman, J. M. Teich, L. Leape, B. Shea, R. Eve, E. Burdick, D. L. Seger, M. V. Vliet, and D. W. Bates, “Identifying Adverse Drug Events Development of a Computer-based Monitor and Comparison with Chart Review and Stimulated Voluntary Report,” J Am Med Inform Assoc, vol. 5, pp. 305-314, 1998.

[4] R. Platt, R. M. Carnahan, J. S. Brown, E. Chrischilles, L. H. Curtis, S. Hennessy, J. C. Nelson, J. A. Racoosin, M. Robb, S. Schneeweiss, S. Toh, and M. G. Weiner, “The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction,” Pharmacoepidemiol Drug Saf, vol. 21, pp. 1-8, 2012.

[5] L. H. Curtis, M. G. Weiner, D. M. Boudreau, W. O. Cooper, G. W. Daniel, V. P. Nair, M. A. Raebel, N. U. Beaulieu, R. Rosofsky, T. S. Woodworth, and J. S. Brown, “Design considerations, architecture, and use of the Mini-Sentinel distributed data system,” Pharmacoepidemiol Drug Saf, vol. 21, pp. 23–31, 2012.

[6] S. Forrow, D.M. Campion, L. J. Herrinton, V. P. Nair, M. A. Robb, M. Wilson, and R. Platt, “The organizational structure and governing principles of the Food and Drug Administration's Mini-Sentinel pilot program,” Pharmacoepidemiol Drug Saf, vol. 21, pp. 12–17, 2012.

[7] P. E. Stang, P. B. Ryan, J. A. Racoosin, J. M. Overhage, A. G. Hartzema, C. Reich, E. Welebob, T. Scarneccia, and J. Woodcock, “Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership,” Ann Intern Med, vol. 153, pp. 600-606, 2010.

[8] G. Trifiro, A. Fourrier-Reglat, M. C. Sturkenboom, C. Díaz Acedo, and J. Van Der Lei, “The EU-ADR project: preliminary results and perspective,” Stud Health Technol Inform, vol. 148, pp. 43-9, 2009.

[9] P. M. Coloma, M. J. Schuemie, G. Trifirò, R. Gini, R. Herings, J. Hippisley-Cox, G. Mazzaglia, C. Giaquinto, G. Corrao, L. Pedersen, J. Van Der Lei, and M. Sturkenboom, “Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project,” Pharmacoepidemiol Drug Saf, vol. 20, pp.1–11, 2011.

[10] H. S. Kim, H. Cho, and I. K. Lee, “Design and Development of an EHR Platform Based on Medical Informatics Standards,” Journal of Fuzzy Logic and Intelligent Systems, vol. 21, pp. 456-462, 2011.

[11] S.H. Hwang, E.Y. Kim, Y. S. Lee, S. Y. Jung, Y. M. Lee, K. H. Son, K. U. Choi, S. H. Lee, and Y. Kim, “Implementation and Evaluation of the Computerized Surveillance System to Identify Adverse Drug Events: pilot study,” J. Kor. Soc. Health-Syst. Pharm., vol. 22, no. 2, pp. 118-136, 2005.

[12] Y. H. Lee, Y. M. Yoon, B. M. Lee, H. J. Hwang, and U.G. Kang, “Development of Mining model through reproducibility assessment in Adverse drug event surveillance system,” The Korea Society of Computer and Information, vol. 1, no. 3, pp. 183-192, March 2009a.

[13] Y. H. Lee, U.G. Kang, and R. W. Park, “Development of Adverse Drug Event Surveillance System using BI Technology,” Journal of Korea Contents, vol. 9, no. 2, pp. 106-114, 2009b.

[14] A.J. Cook, R. C. Tiwari, R. D. Wellman, S. R. Heckbert, L. Li, P. Heagerty, T. Marsh, and J. C. Nelson, “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, vol. 21, pp. 72-81, 2012.

[15] P.E. Stang, P.B. Ryan,S.B. Dusetzina, A.G. Hartzema, C. Reich, J. M. Overhage,J. A. Racoosin, "Health Outcomes of Interest in Observational Data: Issues in Identifying Definitions in the Literature," Health Outcomes Research in Medicine, vol. 3, Iss. 1, pp. e37–e44, 2012.

[16] S.N. Murphy, V. Castro, J. Colecchi, A. Dubey, V. Gainer, C. Herrick, and M. Sordo,"Partners HealthCare OMOP Study Report," Foundation for the National Institutes of Health Observational Medical Outcomes Partnership Partners HealthCare, Jan. 10, 2011.

[17] J. M. Overhage, P. B. Ryan, C. G. Reich, A. G. Hartzema, and P. E. Stang, “Validation of a common data model for active safety surveillance research,” Am Med Inform Assoc, vol. 19, pp. 54-60, 2012.