# Development of Clinical Data Mart of HMG-CoA Reductase Inhibitor for Varied Clinical Research

Hun-Sung Kim[1,2,]*, Hyunah Kim[3,]*, Yoo Jin Jeong[1], Tong Min Kim[1], So Jung Yang[1], Sun Jung Baik[1], Seung-Hwan Lee[2], Jae Hyoung Cho[2], In Young Choi[1], Kun-Ho Yoon[1,2]

[1]Department of Medical Informatics, College of Medicine, The Catholic University of Korea; [2]Division of Endocrinology and Metabolism, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea; [3]College of Pharmacy, Sookmyung Women's University, Seoul, Korea

**Background:** The increasing use of electronic medical record (EMR) systems for documenting clinical medical data has led to EMR data being increasingly accessed for clinical trials. In this study, a database of patients who were prescribed statins for the first time was developed using EMR data. A clinical data mart (CDM) was developed for cohort study researchers.

**Methods:** Seoul St. Mary's Hospital implemented a clinical data warehouse (CDW) of data for ~2.8 million patients, 47 million prescription events, and laboratory results for 150 million cases. We developed a research database from a subset of the data on the basis of a study protocol. Data for patients who were prescribed a statin for the first time (between the period from January 1, 2009 to December 31, 2015), including personal data, laboratory data, diagnoses, and medications, were extracted.

**Results:** We extracted initial clinical data of statin from a CDW that was established to support clinical studies; the data was refined through a data quality management process. Data for 21,368 patients who were prescribed statins for the first time were extracted. We extracted data every 3 months for a period of 1 year. A total of 17 different statins were extracted. It was found that statins were first prescribed by the endocrinology department in most cases (69%, 14,865/21,368).

**Conclusion:** Study researchers can use our CDM for statins. Our EMR data for statins is useful for investigating the effectiveness of treatments and exploring new information on statins. Using EMR is advantageous for compiling an adequate study cohort in a short period.

**Keywords:** Electronic health records; Clinical data mart; Hydroxymethylglutaryl-CoA reductase inhibitors; Clinical data warehouse

## INTRODUCTION

Recently, there has been a rapid increase in the use of electronic medical records (EMRs). These computerized database (DB) system records eliminate the need for hand writing medical information, such as symptoms, diagnoses, and clinical results, on paper [1]. Concerns exist about the use of EMR for clinical research owing to a lack of random sampling and subsequent generalizability [2]. However, several recent reports have presented the advantages of EMR [3-6]. EMR data are systematically managed and easily accessed, enabling the ability to collect information about a patient's current medical, past medical, family, and therapeutic histories. EMR data also include interdisciplinary clinical treatments and prescriptions from other departments within the same clinical center. With the increasing use of EMR systems for documenting clinical medical data, clinical trials have also increasingly accessed EMR data [7-9].

Randomized controlled trials (RCTs), in which medical interventions are conducted in a targeted patient population, are preferred in clinical trials when testing the efficacy and effectiveness of a drug or treatment, or when documenting the progression of a disease. Strict standards are generally applied to participant selection for RCTs to ensure the validity of the results. However, RCTs are expensive and time-intensive; consequently, study enrollment is often limited. EMR-based studies are similar in structure to a cohort study; however, the use of EMR data in studies allows for quick and simple extraction of large amounts of data that were collected over long periods of time. This capability will become a greater advantage as more EMR data are accumulated over time [10-12]. Therefore, EMR-based clinical research requires a standardized clinical data mart (CDM) that various researchers can employ to readily extract necessary data depending on their diverse research objectives.

In this study, 3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitor (statin) was selected. It is the predominantly prescribed statin for preventing cardiovascular disorders [13]. Statins are used to prevent cardiovascular disease by lowering low density lipoprotein cholesterol (LDL-C) and triglyceride (TG) levels and increasing high density lipoprotein cholesterol (HDL-C) levels [14,15]. Several types of statins exist on the market with differing efficacies, and numerous clinical trials have compared their respective effects [14,15]. The primary purpose of this study was to develop a clinical statin data mart to address various purposes, such as assessments of drug efficacy and safety. We intended to aggregate a large amount of data on patients who were prescribed a statin for the first time to en-

able researchers to conduct their relevant studies. By establishing a CDM that includes patient personal data, medical history, medication history, and other patient information relating to statins, we strived to facilitate rapid and effective access to diverse patient information and data. We furthermore strived to open a CDM for authorized researchers and other users to share and enhance study results through open technology.

## METHODS

### Clinical data warehouse

Directly extracting research data via a query from the EMR system can impact the performance of the system for routine hospital use. A clinical data warehouse (CDW) provides numerous benefits to researchers for quality data collection and decision-making capabilities through quick and efficient access to patient information and links to multiple operational data sources. Furthermore, it can be used to discover disease relationships and drugs in addition to repositioning drugs by combining different data sources and validating the consistency of information. Therefore, we established the study dataset to include diagnosis and laboratory data based on a CDW system [16]. The CDW system of Seoul St. Mary's Hospital is currently comprised of 30 tables that include all clinical, prescription, laboratory, radiology, pathology, and other information of all patients at the hospital since 1997. Currently, data exist in the system for approximately 2.8 million patients, 47 million medication prescription events, and laboratory results for 150 million cases. Of the 2.2 billion total records, we developed a research DB from a subset of the data based on a study protocol. It additionally provides comprehensive views of clinical data for specific purposes [17,18].

### Extraction of the study sample

Data for patients who were prescribed a statin for the first time at Seoul St. Mary's Hospital between January 1, 2009 and December 31, 2015 were extracted from the CDW. Cases were identified as those who did not have a statin prescription for at least six months before a statin was initially prescribed. The date on which the initial statin prescription occurred was defined as visit 0 (index date, baseline) (Fig. 1). Visit 1 (an average of 3 months later) was defined as the next occurrence of a laboratory test and subsequent renewal of the statin prescription within 45 to 135 days of the baseline. Visit 2 was defined as the subsequent visit that occurred within 136 to 225 days after the baseline (an average of 6 months later). Visit 3 was defined as
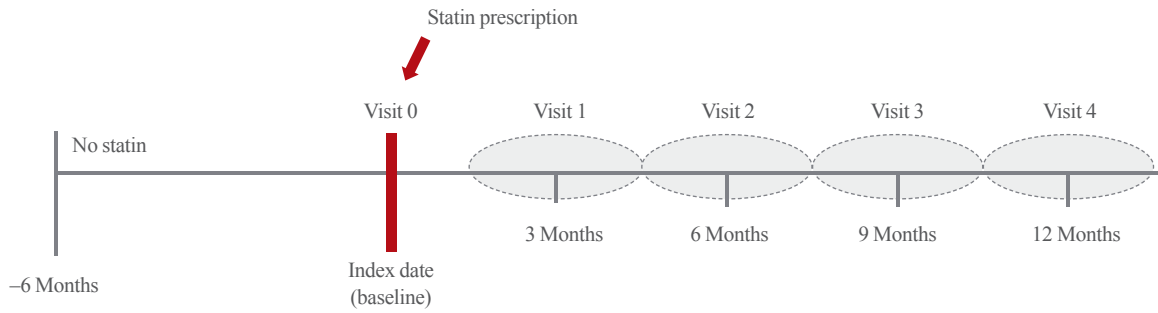
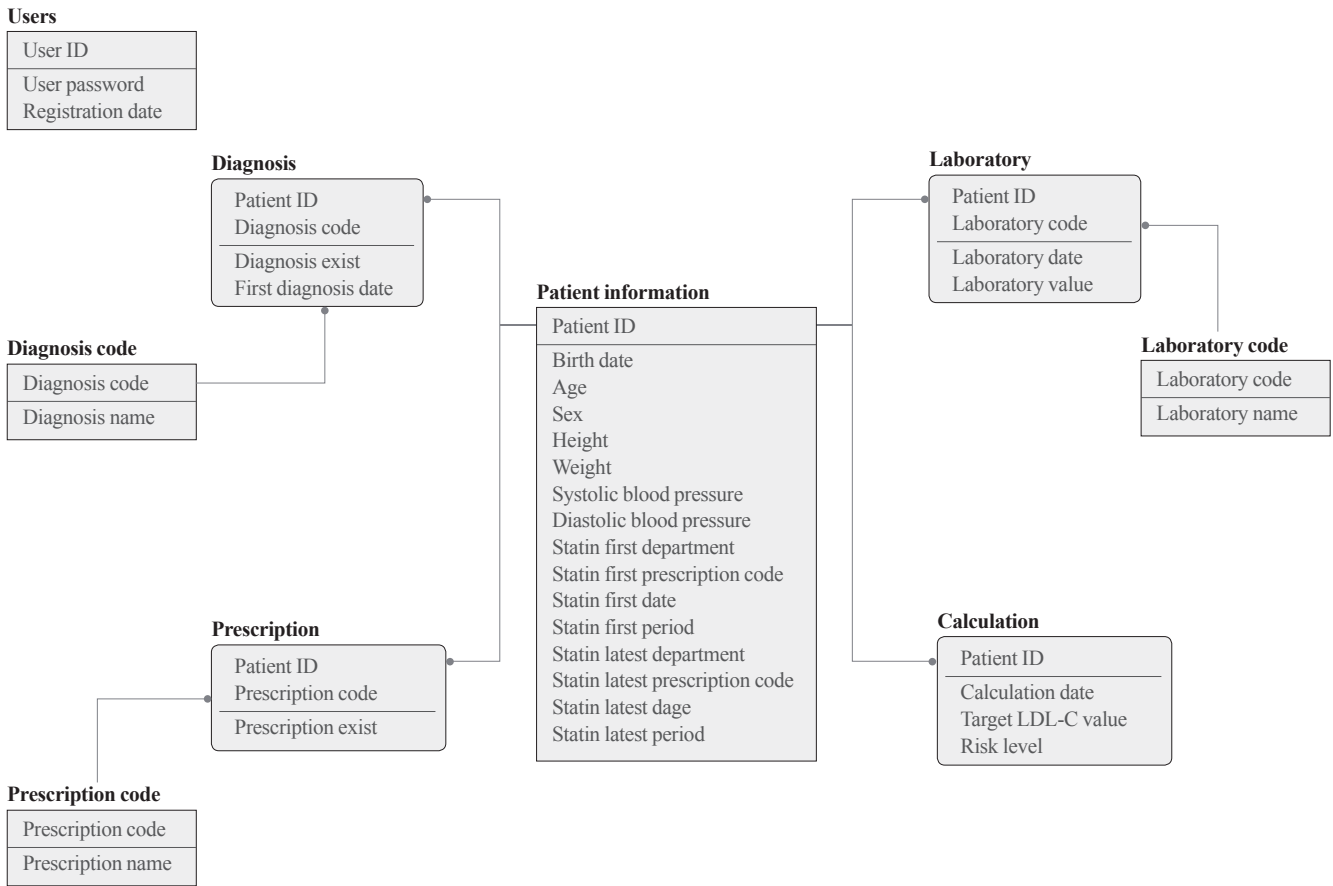**Fig. 1.** Design of clinical data mart for clinical trials on statins.



**Fig. 2.** Entity relationship diagram. ID, identification; LDL-C, low density lipoprotein cholesterol.

226 to 315 days after the baseline (an average of 9 months later), and visit 4 was defined as 316 to 405 days after the baseline (an average of 1 year later). Cases were checked when the prescription changed to a different statin type or the statin prescription was suspended during the study period. When the patients visited the hospital or had a blood test performed more than once within one period of a visit, the test results of the dates closest to the 91th, 182th, 273th, and 365th day were retrieved.

This study covered all statin types that are prescribed at Seoul

St. Mary's Hospital. According to American College of Cardiology/American Heart Association guidelines [15], we classified statins based on intensity, type, and dose. The statin types and dosages are as follows: atorvastatin (10, 20, and 40 mg), fluvastatin (40 and 80 mg), pitavastatin (2 and 4 mg), pravastatin (10, 20, and 40 mg), rosuvastatin (5, 10, and 20 mg), simvastatin (20 and 40 mg), and simvastatin plus ezetimibe (10/10 and 20/10 mg). The combination statins and other drugs with other effects were excluded from the study (atorvastatin plus amlodipine or

pravastatin plus fenofibrate).

Patient data were extracted, including date of birth, age (when first prescribed statins), sex, department in which the statins were first prescribed, first statin prescription date, days of prescription, etc. Besides total cholesterol, TG, HDL-C, LDL-C, which are indicators relevant to hyperlipidemia, blood tests were also covered in this study, such as blood urea nitrogen, creatinine, aspartate aminotransferase/alanine aminotransferase (AST/ALT), hemoglobin/hematocrit, glycated hemoglobin, alkaline phosphatase, high-sensitivity C-reactive protein, γ-glutamyl transpeptidase, thyroid function test (thyroid stimulating hormone, free thyroxine), and others. In modeling the DB, we structuralized relations between objects of unique features comprising it (Fig. 2).

In addition to hyperlipidemia, various diseases, including cardiovascular disorders, diabetes, and others, have been recently reported as relevant to statins; thus, they were included. Moreover, to study statin side effects, consideration was given to whether a patient used fenofibrate, omega-3 fatty acids, propranolol, thyroxine, warfarin, nicotinic acid, etc., which are known to affect hepatotoxicity.

**Privacy protection**

This study was a retrospective cohort study using EMR data retained by one hospital. All files of extracted data were encoded to prevent personal identification of patients during the extraction process. Only one managing researcher was allowed to access the data; observers and analysts received the data with all personally identifying patient information deleted. Thus, they were unable to identify the actual patient numbers. Data utilized in this study did not include patient's personal information, and there was no risk of physical or psychological damage to the patient subject. Because the data were encoded and anonymized, and because the study was a retrospective cohort type, this study did not affect the patient subjects' rights and welfare. Therefore, informed consent was not required. This study was approved by the Institutional Review Board of the Catholic University of Korea.

## RESULTS

**Establishing the statin data mart to support clinical studies**

We employed initial clinical data of statin obtained from a CDW that was established to support clinical studies. Through a data quality management (DQM) process, the initial clinical

data were refined. Because data was extracted from the EMRs of patients, there were many duplicates and errors. For example, the most frequent case was the inclusion of letters or inequality in the extracted clinical laboratory scores field. We performed pre-processing work and DQM operations on these data successively. As part of the DQM, abnormal data (redundant data, out-of-range data, meaningless data, null space, etc.) were identified using clinical judgments as well as statistical methods. The abnormal data were re-confirmed by direct chart review. This enabled further extraction of patient data in accordance with the objective of each clinical study and the use of an optimized DB. In this study, clinical information of patients who were prescribed statins for the first time was extracted from the CDW. A data mart for analysis appropriate to the study objective was established, and relevant DB information was obtained. Various structured or encoded clinical information could be automatically incorporated into the DB. However, unstructured free text content, such as patient height and weight, was manually inputted by direct chart review. To enhance the reliability of the data, problematic values were reviewed and compared with the original data. A data description table was developed (Fig. 3). Personal identification information (name, social security number, etc.) was not included when extracting the requisite clinical data. Patient numbers were collected anonymously during the data collection.

**Statin data mart composition**

The data included patient personal information, such as height, weight, age, sex, etc. Patient laboratory results included glucose levels, AST/ALT values, and others, as shown in Table 1 for the baseline, visit 1 (an average of 3 months later), visit 2 (an average of 6 months later), visit 3 (an average of 9 months later), and visit 4 (an average of 12 months later) (Fig. 1). The presence of a diagnosis, such as hypertension, diabetes mellitus, etc., and the date when the diagnosis was first provided were together extracted. Accordingly, it was possible to distinguish whether certain diseases occurred before or after the first statin prescription and how soon the diseases occurred after the first statin prescription. Additionally extracted were various medications that were prescribed besides statins or those that could interact with statins.

**Baseline characteristics of statin data mart**

Data were extracted of a total of 21,368 patients who were prescribed statins for the first time at the hospital over 7 years, from January 2009 to December 2015 (Fig. 4). The percentage of

**Table Specification**

| No | Definition | Column | Domain | Type of data | Default | NOT NULL | PK | FK | Definition of Column |
|---|---|---|---|---|---|---|---|---|---|
| | Subject | Project of prof.Kim | Entity | patients | | Table | | patients | |
| | Definition of table | Patient information | | | | | | | |
| 1 | Patient ID | PID | Patient ID | INT | | Y | PK | | |
| 2 | Birth | BRTHDD | Birth | NVARCHAR | | | | | |
| 3 | Age | AGE | Age | INT | | | | | |
| 4 | Sex | SEX | Sex | NVARCHAR | | | | | |
| 5 | Height | HEIGHT | Height | NVARCHAR | | | | | |
| 6 | Weight | WEIGHT | Weight | NVARCHAR | | | | | |
| 7 | Systolic blood pressure | SBP | Systolic blood pressure | NVARCHAR | | | | | |
| 8 | Diastolic blood pressure | DBP | Diastolic blood pressure | NVARCHAR | | | | | |
| 9 | The initial prescription department of statin | INITDEPT | The initial prescription department of statin | NVARCHAR | | | | | |
| 10 | The initial prescription code of statin | INITPRCPCD | The initial prescription code of statin | NVARCHAR | | | | | |
| 11 | The initial prescription date of statin | INITDT | The initial prescription date of statin | NVARCHAR | | | | | |
| 12 | The initial prescription period of statin | INITDAYNO | The initial prescription period of statin | NVARCHAR | | | | | |
| 13 | The latest prescription department of statin | LASTDEPT | The latest prescription department of statin | NVARCHAR | | | | | |
| 14 | The latest prescription code of statin | LASTPRCPCD | The latest prescription code of statin | NVARCHAR | | | | | |
| 15 | The latest prescription date of statin | LASTDT | The latest prescription date of statin | NVARCHAR | | | | | |
| 16 | The latest prescription period of statin | LASTDAYNO | The latest prescription period of statin | NVARCHAR | | | | | |

| No | Index | Index column | Type of index | UNIQUE | Remarks |
|---|---|---|---|---|---|
| 1 | User ID | PID | PK | Y | |

Revision history / List of tables / User / Patient information / laboratory / laboratory code / diagnosis / diagnosis code / prescription / prescription code / calculation

**Fig. 3.** Example of data table specification.

**Table 1.** Extraction of Data Description Table of Clinical Research Data

| Patient information | Laboratory information | Drug information | Diagnosis information |
|---|---|---|---|
| ID number/birth date/age/sex/ height/weight/BMI/systolic/ diastolic BP | WBC count/ANC/RBC count/platelet count/Hb/Hct/prothrombin time (INR)/activated PTT/ESR/hs-CRP/ glucose/HbA1c/BUN/creatinine MDRD GFR/total bilirubin | Atorvasatin/fluvastatin/ pitavastatin/pravastatin/ rosuvastatin/simvastatin/ simvastatin plus ezetimibe complex | Hypertensive diseases/ischemic heart diseases/cerebrovascular diseases/ aneurysm dissection |
| Department of first prescription | AST/ALT/ALP/γ-GTP/LDH | Fenofibrate/gemfibrozil/niacin/ omega-3 fatty acid | Abnormal/GTT/hyperglycemia/IFG/ IGT/prediabetes/diabetes mellitus |
| Date of first prescription/last prescription | Total cholesterol/triglyceride/HDL-C/ LDL-C | Propranolol/thyroxine/warfarin/ bisphosphonate/metformin | Hypothyroidism/osteoporosis/fracture |
| No. of days of prescription in every visit | Sodium/potassium/CPK/PSA/free PSA/TSH/free T4 | | ARF/ARF d/t RHABDO/nontraumatic RHABDO |
| | | | Renal stone/calculus of kidney |
| | | | Erectile dysfunction/necrotizing myopathy/prostate cancer/cancer |

ID, identification; BMI, body mass index; BP, blood pressure; WBC, white blood cell; ANC, absolute neutrophil count; RBC, red blood cell; Hb, hemoglobin; Hct, hematocrit; INR, international normalized ratio; PTT, partial thromboplastin time; ESR, erythrocyte sedimentation rate; hs-CRP, high-sensitivity C-reactive protein; HbA1c, glycated hemoglobin; BUN, blood urea nitrogen; MDRD GFR, modification of diet in renal disease glomerular filtration rate; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; γ-GTP, γ-glutamyl transpeptidase; LDH, lactate dehydrogenase; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; CPK, creatine phosphokinase; PSA, prostate-specific antigen; TSH, thyroid stimulating hormone; free T4, free thyroxine; GTT, glucose tolerance test; IFG, impaired fasting glucose; IGT, impaired glucose tolerance; ARF, acute renal failure; d/t, due to; RHABDO, rhabdomyolysis.

males was 44.2% (9,439/21,368); the percentage of females was 55.8% (11,929/21,368). The mean age was $63 \pm 12$ years. A total of 17 different statins were extracted. Atorvastatin (10 mg; 21.0%, 4,490/21,368) and rosuvastatin (10 mg; 20.4%, 4,364/

21,368) were the most commonly prescribed, followed by simvastatin (20 mg; 11.7%, 2,493/21,368) and pitavastatin (2 mg; 11.1%, 2,364/21,368). Statins were most often first prescribed by the endocrinology department (69.6%, 14,865/21,368), fol-
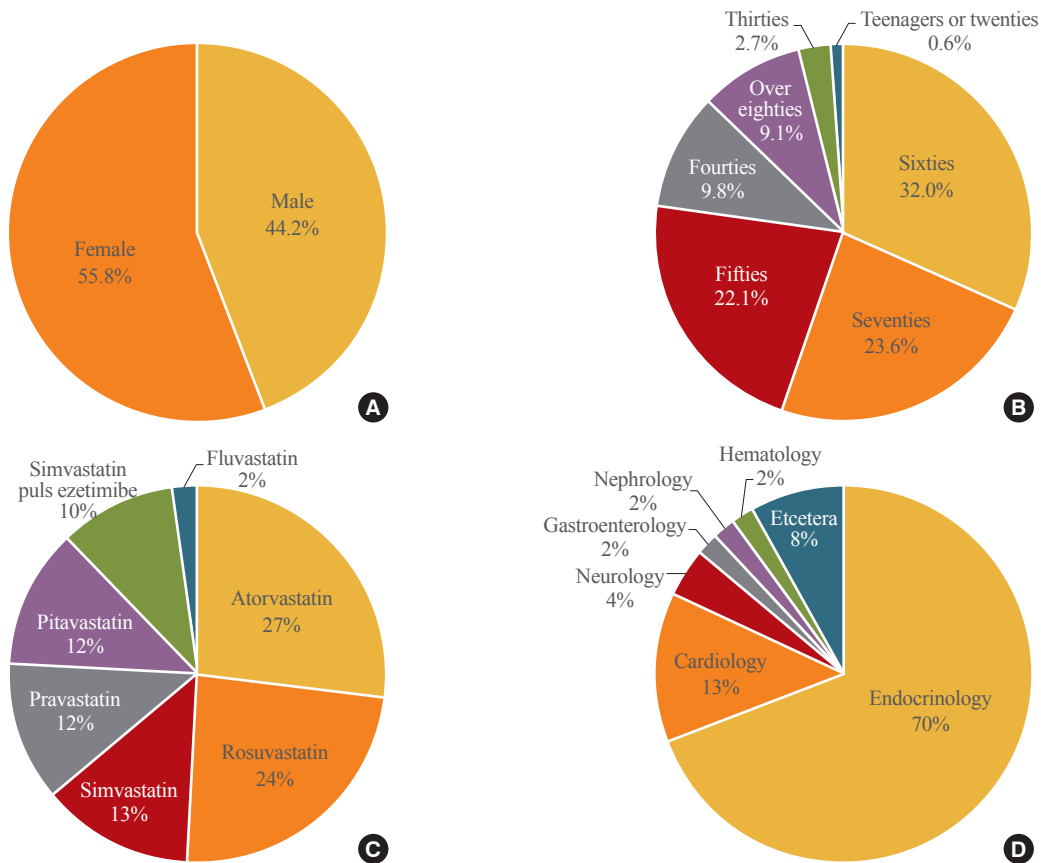
Fig. 4. Areas of focus in the 21,368 patients examined: (A) sex, (B) age, (C) type of statin, and (D) department.

lowed by cardiology (13.0%, 2,770/21,368) and neurology (3.9%, 832/21,368).

## DISCUSSION

We selected statins that were initially prescribed when addressing hyperlipidemia. We established a clinical DB using the data of patients who were prescribed statins for the first time at Seoul St. Mary's Hospital. Subjects of EMR clinical data extraction for the statin study were defined; thus, this CDM included the patient personal information, diagnosis information, etc. Clinical researchers can conduct diverse, appropriate, and optimal EMR-based large-scaled retrospective cohort studies using this data depending on their study purposes.

Through a year of laboratory follow-up testing, researchers can analyze effects of statins, such as LDL-C lowering effects, and they can conduct basic research about guidelines for each statin. By including various illnesses and diseases, a risk model of occurrences [19] of cardiovascular disorders [20] can be developed. Moreover, research on recent issues, such as correla-

tions between statins and occurrences of diabetes [21] or cancers [22], can be quickly and easily conducted. Researchers can perform assessments and analyses of economic efficiency [23] regarding statins and comparison analyses of diverse adverse drug effects [24]. It is expected that analysis of effects of statins and the causes of disease occurrences will be possible by using a Bayesian network. Cost effectiveness, which depends on the different patient ages, risk types, and risks of side effects, can also be analyzed. Effects of statins that can be evaluated as outcomes include the rate of LDL-C lowering and contraction of cardiovascular disorders [13].

Various studies using the CDM have been conducted [25]. Prescription rates of each department have been analyzed along with prescription patterns [26]. In a previous study, cardiologists were determined to be the most frequent prescribers of statins [26]. However, our study showed that the initial prescription of a statin was most commonly provided by an endocrinologist. This may be because the endocrinologist prescribes statins for preventing cardiovascular disease. The validity of statin studies using EMR data has been proved [13]. Various studies can thus

be conducted, such as those on different side effects that depend on different types/volumes of statins, actual statin prescription examples and problems, development of programs to predict clinical aspects, research on various cases that cannot be conducted with RCT, and rapid studies on rare side effects.

CDM development has become possible because of the dissemination of EMR. Accordingly, clinical studies that use a CDM have various advantages [2,27,28]. EMR data can provide information that is not available with traditional paper medical records, including information about the various treatments for each patient. Extensive quantities of medical data can be easily and quickly extracted by EMR [2]. The use of EMR addresses the disadvantages of both cohort studies (follow-up costs, long study periods, and maintenance of consistency during the study period) and RCTs (loss of follow-up, changes in treatment, long study periods, and expenses). The investigator has direct access to the EMR system and can quickly verify a hypothesis by sampling variables from the system that are based on the hypothesis. As additional varied data are accumulated, studies regarding cause-and-effect relationships of rare factors become possible.

Furthermore, the additional advantage exists of increasing the generalizability of the results. In this study, we believed it was necessary to organize such diverse data by standardizing them. We thus strived to establish a CDM under a systematic plan from the beginning. Moreover, it is possible to develop a multi-center integration data mart to enable researchers of other hospitals to add their statin data to our data mart. We therefore employed a standard protocol. Consequently, clinical research using our CDM has many benefits. It can save time and labor compared with conventional clinical studies, which is the most important aspect. This is because it can help preview results before conventional long-term studies and can thus benefit clinical researchers. A large amount of clinical data can be collected within a short period. Unlike conventional clinical studies that often extrapolate results of the entire population on the basis of a small sample size, in EMR analysis, the amount of data can be significantly larger, thereby offsetting other limitations.

Previous research on the use of EMR data focused primarily on its convenience and ease of access [2,27,28]. However, with the recent accumulation of extensive data in EMR systems, these data can be used in clinical studies. EMR data can be used to determine the efficacy of medications that are currently available. Owing to the ability of inexpensively extracting large quantities of data in a short period, EMR systems will become more valuable for research in the future. Moreover, previously

unknown information that clinical researchers did not predict in the study planning process may come to light. Therefore, it seems that EMR systems may lead to new knowledge and theories.

Despite these strengths of the EMR system, EMR-based clinical research has certain limitations. First, some possible cofactors and confounders cannot be accessed from the DB, including compliance with medication prescription, and the severity of potential diseases. However, an EMR-based trial is a reflection of actual practice relating to statin use. A researcher has a standardized study plan that is identified early to minimize confounding factors, which can have a significant impact on the results. Second, our data mart was conducted over a short period of 12 months in a single center. With an increase in the amount of clinical or biochemical data and its intensified variation, additional biochemical variables can become available that would identify patients who respond well to a statin compared to those who do not. To accomplish this, additional data in the EMR and a longer period study are necessary.

As EMR have been widely implemented in hospitals in Korea, we expect a drastic increase in the analysis of accumulated EMR data. To obtain clinically significant data from the data analyzed from large hospitals, a CDW can be an important tool. Standardization of statin prescription can be possible by effectively verifying effects of each statin through various studies using CDM, and collecting domestic data will contribute to ground-based prescription and clinical and health research. Moreover, establishing data and developing algorithms can become foundational to Korean guidelines of dyslipidemia. Clearly, CDM cannot replace RCT. Nonetheless, CDM research in which rapid data extraction is possible can help establish directions of RCT research. New directions of RCT can be roughly established through CDM research prior to RCT. With the advent of technology, we expect that the use of EMR can potentially lead to various types of clinical research.

## CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

## ACKNOWLEDGMENTS

and Welfare, Republic of Korea (grant number: HC15C1362).

## ORCID

Hun-Sung Kim *http://orcid.org/0000-0002-7002-7300*

## REFERENCES

1. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010;363:501-4.
2. Aaronson JW, Murphy-Cullen CL, Chop WM, Frey RD. Electronic medical records: the family practice resident perspective. Fam Med 2001;33:128-32.
3. Makoul G, Curry RH, Tang PC. The use of electronic medical records: communication patterns in outpatient encounters. J Am Med Inform Assoc 2001;8:610-5.
4. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care. Health Aff (Millwood) 2009;28:323-33.
5. Christensen T, Grimsmo A. Expectations for the next generation of electronic patient records in primary care: a triangulated study. Inform Prim Care 2008;16:21-8.
6. Callen JL, Bevis M, McIntosh JH. Patients' perceptions of general practitioners using computers during the patient-doctor consultation. HIM J 2005;34:8-12.
7. 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;91:467-74.
8. 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;20:598-607.
9. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol Drug Saf 2010;19:858-68.
10. Waller PC, Coulson RA, Wood SM. Regulatory pharmacovigilance in the United Kingdom: current principles and practice. Pharmacoepidemiol Drug Saf 1996;5:363-75.
11. Hauben M, Patadia V, Gerrits C, Walsh L, Reich L. Data mining in pharmacovigilance: the need for a balanced perspective. Drug Saf 2005;28:835-42.
12. Waller PC, Evans SJ. A model for the future conduct of pharmacovigilance. Pharmacoepidemiol Drug Saf 2003;12:17-29.
13. Kim HS, Lee H, Park B, Park S, Kim H, Lee SH, et al. Comparative analysis of the efficacy of low- and moderate-intensity statins in Korea. Int J Clin Pharmacol Ther 2016;54:864-71.
14. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation 2002;106:3143-421.
15. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014;129:S1-45.
16. Harrison JH Jr. Introduction to the mining of clinical data. Clin Lab Med 2008;28:1-7.
17. Choi IY, Park S, Park B, Chung BH, Kim CS, Lee HM, et al. Development of prostate cancer research database with the clinical data warehouse technology for direct linkage with electronic medical record system. Prostate Int 2013;1:59-64.
18. de Mul M, Alons P, van der Velde P, Konings I, Bakker J, Hazelzet J. Development of a clinical data warehouse from an intensive care clinical information system. Comput Methods Programs Biomed 2012;105:22-30.
19. Lee YH, Bang H, Kim DJ. How to establish clinical prediction models. Endocrinol Metab (Seoul) 2016;31:38-44.
20. Ryan RP, McManus RJ, Mant J, Macleod JA, Hobbs FD. Statins in heart failure: retrospective cohort study using routine primary care data. Ann Med 2009;41:490-6.
21. Cho Y, Choe E, Lee YH, Seo JW, Choi Y, Yun Y, et al. Risk of diabetes in patients treated with HMG-CoA reductase inhibitors. Metabolism 2015;64:482-8.
22. Marelli C, Gunnarsson C, Ross S, Haas S, Stroup DF, Cload P, et al. Statins and risk of cancer: a retrospective cohort analysis of 45,857 matched pairs from an electronic medical records database of 11 million adult Americans. J Am Coll Cardiol 2011;58:530-7.
23. Ohsfeldt RL, Gandhi SK, Fox KM, McKenney JM. Statin cost-effectiveness comparisons using real-world effectiveness data: formulary implications. Value Health 2008;11:

1061-9.

24. Kim HS, Lee SH, Kim H, Lee SH, Cho JH, Lee H, et al. Statin-related aminotransferase elevation according to baseline aminotransferases level in real practice in Korea. J Clin Pharm Ther 2016;41:266-72.

25. Jones PH, Nair R, Thakker KM. Prevalence of dyslipidemia and lipid goal attainment in statin-treated subjects from 3 data sources: a retrospective analysis. J Am Heart Assoc 2012;1:e001800.

26. Kim HS, Kim H, Lee H, Park B, Park S, Lee SH, et al. Analysis and comparison of statin prescription patterns and outcomes according to clinical department. J Clin Pharm Ther 2016;41:70-7.

27. Hippisley-Cox J, Pringle M, Cater R, Wynn A, Hammersley V, Coupland C, et al. The electronic patient record in primary care: regression or progression? A cross sectional study. BMJ 2003;326:1439-43.

28. El-Kareh R, Gandhi TK, Poon EG, Newmark LP, Ungar J, Lipsitz S, et al. Trends in primary care clinician perceptions of a new electronic health record. J Gen Intern Med 2009; 24:464-8.