Original Article

PI PROSTATE INTERNATIONAL

Prostate Int 2013;1(2):59-64 • http://dx.doi.org/10.12954/PI.12015

Development of prostate cancer research database with the clinical data warehouse technology for direct linkage with electronic medical record system

In Young Choi, Seungho Park[1], Bumjoon Park[1], Byung Ha Chung[2], Choung-Soo Kim[3], Hyun Moo Lee[4], Seok-Soo Byun[5], Ji Youl Lee[6]

Department of Medical Informatics, The Catholic University of Korea College of Medicine, Seoul, Korea
[1]Graduate School of Information System, Hanyang University, Seoul, Korea
[2]Department of Urology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea
[3]Department of Urology, Asan Medical Center, Seoul, Korea
[4]Department of Urology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea
[5]Department of Urology, Seoul National University Bundang Hospital, Seongnam, Korea
[6]Department of Urology, Seoul St. Mary's Hospital, The Catholic University of Korea College of Medicine, Seoul, Korea

Purpose: In spite of increased prostate cancer patients, little is known about the impact of treatments for prostate cancer patients and outcome of different treatments based on nationwide data. In order to obtain more comprehensive information for Korean prostate cancer patients, many professionals urged to have national system to monitor the quality of prostate cancer care. To gain its objective, the prostate cancer database system was planned and cautiously accommodated different views from various professions.
Methods: This prostate cancer research database system incorporates information about a prostate cancer research including demographics, medical history, operation information, laboratory, and quality of life surveys. And, this system includes three different ways of clinical data collection to produce a comprehensive data base; direct data extraction from electronic medical record (EMR) system, manual data entry after linking EMR documents like magnetic resonance imaging findings and paper-based data collection for survey from patients.
Results: We implemented clinical data warehouse technology to test direct EMR link method with St. Mary's Hospital system. Using this method, total number of eligible patients were 2,300 from 1997 until 2012. Among them, 538 patients conducted surgery and others have different treatments.
Conclusions: Our database system could provide the infrastructure for collecting error free data to support various retrospective and prospective studies.

Keywords: Prostate neoplasms, Hospital information systems, Information Storage and Retrieval, Retrospective studies

INTRODUCTION

The National Statistics Office in Korea exclaimed that the prevalence of prostate cancer quadrupled between 2002 and 2008 [1,2]. The incidence of prostate cancer in Korea increased up to 24.9 per 100,000 men in 2009 in comparison with 13 per 100,000 in 2008.

Environmental elements, western dietary habits, and the rise in average life expectancy are known as influential factors of the increased rate of prostate cancer patients. The Korean

Corresponding author: Ji Youl Lee
Department of Urology, Seoul St. Mary's Hospital, The Catholic University of Korea College of Medicine, 222 Banpo-daero, Seocho-gu, Seoul 137-701, Korea
E-mail: uroljy@catholic.ac.kr / Tel: +82-2-2258-6227 / Fax: +82-2-599-7839
Submitted: 20 December 2012 / Accepted after revision: 21 May 2013

Urologic Association launched a campaign of prostate-specific antigen (PSA) screening test to raise public awareness on the increase in prostate cancer in Korean patients. The rapid increase of prostate cancer patients in Korea requires intensive disease progression and management.

In spite of increased prostate cancer patients, little is known about the impact of treatments for prostate cancer patients and outcome of different treatments with nationwide data. In order to obtain more comprehensive information for Korean prostate cancer patients, multicenter longitudinal database had been proposed. There were similar projects in the United States and Japan. One of the most popular database for prostate cancer is the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE), which is the web-based database developed in 1995 for longitudinal observation of prostate cancer patients in natural settings in the United States. This project began with ten participating health care centers and increased to 26 centers in one year. Currently, the CaPSURE is one of the most powerful prospective study groups for prostate cancer in the world, composed of approximately 14,000 registered prostate cancer patients. The most recent mover in Asia, the Japan Study Group for Prostate Cancer (J-CaP), was developed in 2001. In the case of J-CaP, the database is comprised of 17,872 prostate cancer patients from prospective studies and research to improve patient care.

Against this background, many professionals urged to have national system to monitor the quality of prostate cancer care [3]. The requirement for multicenter observational prostate cancer database was also proposed in Korea. Observational databases are useful in evaluating large amounts of data in a timely manner, and evaluating clinical outcomes in real healthcare setting. The database is also used to improve the quality of clinical researches because data collection can assist investigators. To gain its objective, the database was carefully planned and cautiously accommodated different views from various professions.

Our first step was to develop research database structure including data elements for a successful observational database. We defined the important questions to which providers want answers, and data elements need to be captured.

And, the second step was to compare different ways of data collection and suggest more efficient methods. Data capture with less human effort is important step to maintain the database for long period time.

The purpose of this study is to propose the multi center observational research database structure incorporating clinical factors and patient self reports and suggest effective ways of data collection linking with clinical information systems. Our objective was to develop research database system having easy data access, transparent scientific reproducibility, and interoperability between multiple centers.

MATERIALS AND METHODS

To develop database structure for multicenter prostate registry system, we analyzed previously developed database systems as follows.

1. Materials

1) CaPSURE database

According to cancer statistics in 2012, prostate cancer was the second leading cause of cancer-related death among men in the United States [4]. The CaPSURE was founded in 1995 as a disease registry of men with all stages of prostate cancer [5]. Currently, a group of 31 urological practice sites enroll patients in CaPSURE, and 40 sites including community based site, 4 Veterans Affairs medical centers are involved. CaPSURE collects approximately 1,000 clinical and patient reported variables. The clinical information includes history of prostate cancer diagnosis, biopsies, pathological findings, staging tests, primary and subsequent treatments, clinic procedures, Karnofsky performance status scores and medications. At each clinic visit the urologist completes a progress record, including current disease status, new prostate or unrelated diagnoses, disease signs and symptoms, and changes in medications. Results of imaging studies and laboratory tests are recorded when they are determined. In addition to the clinical data, the patient information is also collected. At enrollment each patient completes a questionnaire about sociodemographic parameters, comorbidities [6], and baseline health-related quality of life (HRQoL). Every 6 months thereafter patients are mailed a follow-up questionnaire, and HRQoL questionnaires including the Medical Outcomes Study Short Form-36 (SF-36) [7] for general HRQoL and the University of California-Los Angeles Prostate Cancer Index [8] for disease specific HRQoL are collected. Since 1999, the survey on patient satisfaction with care [9,10] and fear of cancer recurrence [11] are also included.

2) J-Cap database

In Japan, the J-CaP database was established in 2001 with financial support from Japan Kidney Foundation. And the Japanese Urological Association commenced a study to gather information about hormone therapy administered to Japanese patients and to analyze the outcomes of treatment. The purposes of this study group were to gather information

http://dx.doi.org/10.12954/PI.12015

about the hormone therapy administered to Japanese prostate cancer patients living in Japan and to analyze the outcomes of treatment in order to create a guideline for optimal hormone therapy. This study analyzes different forms primary androgen deprivation therapy (PADT), including combined androgen blockade therapy, for the treatment of prostate cancer within Japan. The J-CaP registry is a large, multicenter, population-based database of men newly starting PADT for prostate cancer. The following clinical information captured over the Internet; date of birth, family history, date of PSA reading, PSA value, PSA kit name, testosterone value, biopsy date, Gleason score, histological grade, clinical stage, case history, details of hormone therapy, whether or not there has been progress observation, whether or not surgery was carried out, date of surgery, operative procedure, whether or not radiotherapy is being conducted, irradiation method, irradiation date, progress.

The interim analysis of the registration status of the patients and their background variables was reported in 2003 [12], and treatment patterns with PADT have been reported along with an interim analysis of prognosis in 2007 [13]. As of 2005, J-CaP included data for 26,272 patients from 406 institutions comprising 77 university hospitals (67% of those in Japan), 267 general hospitals and 62 private hospitals. Around 50% of new prostate cancer patients treated with hormone therapy in Japan was registered with J-CaP at that time.

3) CPDR database

The Center for Prostate Disease Research (CPDR) was established in 1992, by the United States Congress-Public Law 102-172 [14]. The participating institutions of the CPDR are the Department of Defense, multisite program within the School of Medicine, Department of Surgery, at the Uniformed Services University of the Health Sciences, Bethesda, MD, with its primary clinical program located at Walter Reed Army Medical Center, Washington, D.C. and a scientific laboratory in Rockville, MD. For this program, urologists, cancer biologists, genitor-urinary pathologists, epidemiologists, biostaticians, medical- and bio-informaticians are involved. The goal of the CPDR-Clinical Center Program is to combine prostate screening, clinical diagnosis, data collection, education and counseling, and prostate disease clinical trial research. To address this goal, the National Multicenter Database had been developed. The National Multicenter Database of the CPDR is comprised of several military medical centers including the Army, Air Force, and Navy, and a civilian institution. The data base contains approximately 500 data fields in 48 tables that include registration, patient contact information, pretreatment diagnosis, cancer staging, treatment types, follow-up and recurrence, QoL issues, and many others. Data are collect for standard clinical care, which includes hormonal therapy, radiation therapy, chemotherapy, and surgery. Each major treatment type has a subset of questions asked. The majority of the information is collected during the treatment state. With an average follow-up of 8 years and over 1,200 prostate specimens processed, the CPDR database is rapidly becoming a national prostate cancer research resource. Their efforts had led to more than 300 peer-reviewed publications, numerous scientific presentations, more than 20 clinical trials ranging from disease prevention to QoL.

2. Methods

1) Staged approach of database structure development

Developing research data base structure for multi center prostate cancer research is a complex undertaking. We benchmarked previous research project by searching PubMed database for prostate cancer registry system. We entered the keyword as "prostate cancer registry system", "prostate cancer database system", and "prostate cancer retrospective research". Then, large number of articles used CaPSURE database system. Then, we reviewed CaPSURE database system and explore the similarity and difference of both systems and develop basic category for database system. Later, we finalized data elements by working together with physicians of the related departments and refine the set of data element. We had a weekly meeting with doctors from urology, pathology, radiology and radiology oncology for two months to explore database design structure. Total numbers of participants were around 20 people. During the meeting, we asked the participants to gain consensus regarding important items to which patients and providers want answers to understand data elements to be captured. Finally, our research database structure includes major outcome results for prostate cancer such as Table 1. The database incorporates all information about a prostate cancer research; demographic data, medical history, clinical information, laboratory, survey, and follow-up data. The final database results to include approximately 222 clinical and patient-reported items. Our database structure has flexibility to add new measurements when appropriate and to ensure variables to compare outcomes across other healthcare organizations.

2) Data collection method

After determining the data elements for multicenter research, we then develop the strategy of data collection within timelines. Most of the previous research database is completed by manual entry of physicians or clinical research coordinators or data entry staffs. This data entry method is very labor

Table 1. Data elements of prostate cancer database system

1. Demographics	Age, sex, height, weight, body mass index
2. Medical history	Operation history, preexisting comorbidities, postdiagnosis comorbidities
3. Cancer stage	Clinical TNM stage, pathological stage
4. Laboratory	Prostate-specific antigen, hemoglobin, UDS finding
5. Patient self-reported survey	IPSS, IIEF
6. Radiology	MRI (initial, follow-up), CT (initial, follow-up), TRUS
7. Pathology	Preoperative Bx, postoperative Bx
8. Treatment type	Active surveillance, surgery, hormonal therapy, brachytherapy, cryotherapy
9. Medication	Medication category, duration, route

UDS, urodynamic study; IPSS, International Prostate Symptom Score; IIEF, International Index of Erectile Function; MRI, magnetic resonance imaging; CT, computed tomography; TRUS, transrectal ultrasonography; Bx, biopsy.

Table 2. Data collection methods of prostate cancer database system

Data collection methods	Items	Percent
Direct electronic medical record extraction	59	26.6
Indirect electronic medical record link (unstructured)	155	69.8
Patient reports (paper or mobile link in future)	8	3.6
Total	222	100

intensive and cumbersome, so it usually fails to capture relevant data at early times.

We concluded that high quality data collection strategy is fundamental to achieve relevant information at early times. Our system includes three different ways of clinical data collection to produce a comprehensive data base; direct data extraction from electronic medical record (EMR) system, manual data entry after linking EMR documents like magnetic resonance imaging findings and paper-based data collection for survey from patients. We combine various data collection methods for different types of data elements. For example, preoperative PSA value can be collected from EMR system. Table 2 shows different data collection methods to include research data into the database system. 27% of total data elements can be collected through direct EMR link, and 70% can be completed through indirect EMR link.

Our integrated data collection and data management will contribute to prevent redundant entry of the same information such as direct linking with hospital information systems using clinical data warehousing technique.

3) Clinical data warehouse for direct link with EMR system

Implementing the direct extraction program may decrease the performance of EMR system, and thus it is very difficult to add programs into the hospital's operating system. We suggest using clinical data warehouse (CDW) technology to access and extract research information with less effort. The CDW is the method to develop clinical database that is optimized for distribution, mass storage and complex query processing [15]. It also provides comprehensive views of clinical data for specific purpose. A CDW can provide numerous benefits to researchers with quality data collection, and decision support capability by quick and efficient access to patient information and linkage to multiple operational data sources.

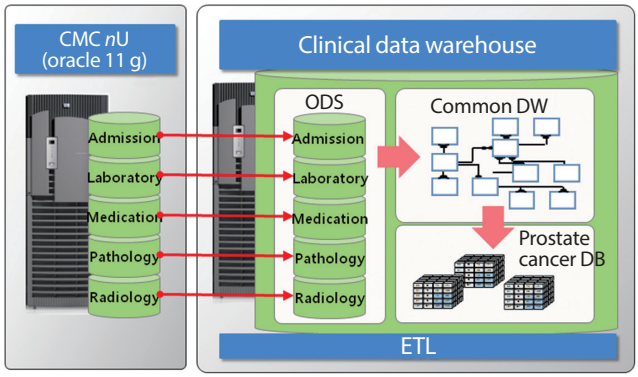

Fig. 1. Clinical data warehouse for electronic medical record link. CMC, catholic medical center; ODS, operational data store; DW, data warehouse; DB, database; ETL, extraction transaction loading .

Using CDW methodology, accurate and high quality prostate cancer patients' data can be collected from EMR system and feeds them to central prostate cancer registry system. All eligible patients with newly diagnosed prostate cancer can be electronically transferred into the prostate cancer registry database from EMR system. And all registered patient information will be periodically updated (Fig. 1).

RESULTS

Our prostate cancer research database system is developed with Microsoft SQL server running on the Microsoft NT servers and is programmed with Java for user interface develop-

http://dx.doi.org/10.12954/PI.12015

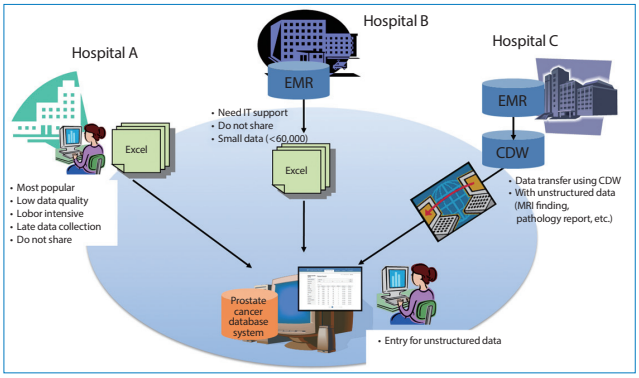

Fig. 2. Integrated prostate cancer database system. EMR, electronic medical record; CDW, clinical data warehouse; MRI, magnetic resonance imaging.

ment. All data and research records are maintained in Internet Data Center which is a facility having telecommunications and storage systems, backup power supplies and redundant data communications connections.

This database system provides three different ways of data collection depending on the information technology infrastructure; direct manual entry; Excel upload, and direct EMR link using CDW technology (Fig. 2). For example, a hospital A already has affluent human resource for prostate cancer registry, and prefers manual entry. Then they can access web browser and enter each data field. A hospital prefers to enter data into Excel file and later upload Excel file into the web system. Other hospital wants to download research data automatically and import the extracted file into the research data base system.

We implemented CDW technology to test direct EMR link method with St. Mary's Hospital system. We defined the EMR variables mapped with the research data elements. Relevant data was transferred into the EMR system of St. Mary's Hospital using CDW technology and extracted data into the prostate cancer database system. To validate CDW technology, we selected sample data and compared with EMR system. Using this method, total number of eligible patients were 2,300 from 2008 until 2012. Among them, 538 patients conducted surgery and others have different treatments.

After completion of the data control for registered patients, analytic reports can be prepared for prospective studies and research for improved prostate cancer patient care. These data summaries evaluate clinical occurrences, patient QoL, economic impact, and oncology outcomes, as well as compare types of treatment by stage and practice, among patients with prostate cancer in other institutions of Korea.

Our system can provide visualization integrating valuable information from different data sources. Researchers interpret

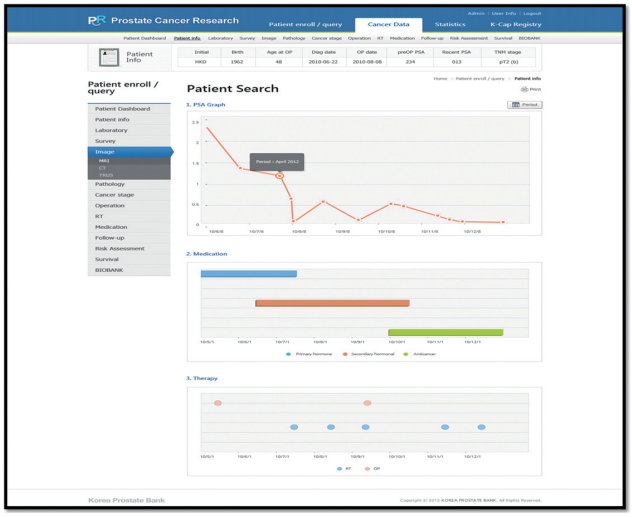

Fig. 3. Patient characteristics by treatment.

clinical effectiveness in place and can be the turning point in uncovering new insights and knowledge about a patient or a disease (Fig. 3).

DISCUSSION

Our suggested database structure is applicable for any hospital which wants to link their EMR system directly with our research system and can be a representative database to understand prostate cancer patients and treatment patters in Korea.

Our system can provide complete treatment histories and patient information and can allow for comparison of different outcomes. As the number of enrolled patients is increased, the system will contribute to compare primary indicator for prostate cancer patients with other institutions in Korea and other databases such as CaPSURE from United States and J-CaP from Japan. For example, UCSF-CAPRA (University of California, San Francisco Cancer of the Prostate Risk Assessment) is a risk assessment tool developed from a cohort of radical prostatectomy patients (n=1,439) in the CaPSURE database. The Japanese tool named as 'J-CAPRA' is developed for patients undergoing PADT and is applied for those with both localized and advanced disease [16].

The longitudinal observation database is an important source to investigate therapeutic efficacy and patient outcomes in the real clinical settings and therefore can be an invaluable complement of randomized clinical trials. Our database system could provide the infrastructure for collecting data on the quality of prostate cancer care. Our large database system, like J-CaP and CaPSURE, can provide valuable real-world information and would help advance clinical management of prostate

cancer patients in the future.

CONFLICT OF INTEREST

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

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. Ministry for Health & Welfare. Annual report of cancer incidence (2006) and survival (1993–2006) in Korea. Seoul: Ministry for Health & Welfare; 2009.
2. Akaza H, Carroll P, Cooperberg MR, Hinotsu S. Fifth Joint Meeting of J-CaP and CaPSURE: advancing the global understanding of prostate cancer and its management. Jpn J Clin Oncol 2012;42:226-36.
3. Malin JL, Kahn KL, Adams J, Kwan L, Laouri M, Ganz PA. Validity of cancer registry data for measuring the quality of breast cancer care. J Natl Cancer Inst 2002;94:835-44.
4. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin 2012;62:10-29.
5. Cooperberg MR, Broering JM, Litwin MS, Lubeck DP, Mehta SS, Henning JM, et al. The contemporary management of prostate cancer in the United States: lessons from the cancer of the prostate strategic urologic research endeavor (CapSURE), a national disease registry. J Urol 2004;171:1393-401.
6. Stier DM, Greenfield S, Lubeck DP, Dukes KA, Flanders SC, Henning JM, et al. Quantifying comorbidity in a disease-specific cohort: adaptation of the total illness burden index to prostate cancer. Urology 1999;54:424-9.
7. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
8. Litwin MS, Hays RD, Fink A, Ganz PA, Leake B, Brook RH. The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure. Med Care 1998;36:1002-12.
9. Hall JA, Feldstein M, Fretwell MD, Rowe JW, Epstein AM. Older patients' health status and satisfaction with medical care in an HMO population. Med Care 1990;28:261-70.
10. Lubeck DP, Litwin MS, Henning JM, Mathias SD, Bloor L, Carroll PR. An instrument to measure patient satisfaction with healthcare in an observational database: results of a validation study using data from CaPSURE. Am J Manag Care 2000;6:70-6.
11. Kornblith AB, Herr HW, Ofman US, Scher HI, Holland JC. Quality of life of patients with prostate cancer and their spouses. The value of a data base in clinical care. Cancer 1994;73:2791-802.
12. Akaza H, Usami M, Hinotsu S, Ogawa O, Kagawa S, Kitamura T, et al. Characteristics of patients with prostate cancer who have initially been treated by hormone therapy in Japan: J-CaP surveillance. Jpn J Clin Oncol 2004;34:329-36.
13. Hinotsu S, Akaza H, Usami M, Ogawa O, Kagawa S, Kitamura T, et al. Current status of endocrine therapy for prostate cancer in Japan analysis of primary androgen deprivation therapy on the basis of data collected by J-CaP. Jpn J Clin Oncol 2007;37:775-81.
14. Brassell SA, Dobi A, Petrovics G, Srivastava S, McLeod D. The Center for Prostate Disease Research (CPDR): a multidisciplinary approach to translational research. Urol Oncol 2009;27:562-9.
15. Harrison JH Jr. Introduction to the mining of clinical data. Clin Lab Med 2008;28:1-7.
16. Cooperberg MR, Hinotsu S, Namiki M, Ito K, Broering J, Carroll PR, et al. Risk assessment among prostate cancer patients receiving primary androgen deprivation therapy. J Clin Oncol 2009;27:4306-13.

http://dx.doi.org/10.12954/PI.12015