Conversion of Diagnosis and Chemotherapy Data in Electronic Health Records to Episode-based Oncology Extension of OMOP-CDM

Hokyun Jeon[1], Seng Chan You[2], MD, Jimyung Park [1], Rae Woong Park[1,2], MD, Ph.D

[1]Department of Biomedical Science, Ajou University Graduate School of Medicine, Suwon, Gyeonggi-do, Republic of Korea; [2]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Gyeonggi-do, Republic of Korea

Is this the first time you have submitted your work to be displayed at any OHDSI Symposium?
Yes __√__ No_____

Abstract

Electronic Health Records (EHR) include the diagnosis and medications for patients with cancer, but there are still limitations to the utility of this information because most records are not machine-readable. Recently, an episode-based extension model has been proposed to provide further structured details of oncology data in the Observational Health Data Sciences and Informatics (OHDSI). In this study, we proposed a process curating unstructured oncology data from EHR and actual implementation of oncology extension model in the common data model.

Introduction

The primary merits of Electronic Health Records (EHR) embrace information about the patient's symptoms, clinical records, and actual treatment history, as well as potentially comprehensive and relatively timely clinical information using physicians' notes[1]. However, since EHRs were not designed for secondary use, it is not seamless in leveraging EHR data for outcomes and comparative effectiveness studies in oncology[2]. Although information of cancer outbreaks exists in the Observational Medical Outcomes Partnership (OMOP)-Common Data Model (CDM) as an observation or measurement records, it is challenging to obtain information related to specific natural history of cancer or subsequent processes of therapeutic courses. Hence, as a proof-of-concept study, we aimed to implement oncology extension model of the OMOP-CDM for the patients with colorectal cancer from EHR.

Method

We sampled a group of patients with colorectal cancer from Ajou University School Of Medicine (AUSOM) database in the format of OMOP-CDM version 5.3 as the proof-of-concept study. The pathology reports of these patients were curated and converted to JavaScript Object Notation (JSON) format. The ICD-O-3 OMOP concept IDs were assigned to each record based on the histology and topography information drawn by reviewing these JSON notes. Drug exposure dates of interest were extracted and the regimen types were classified according to the criteria that had previously been applied to SEER-Medicare[3]. Drugs of interest included fluorouracil, leucovorin, oxaliplatin, capecitabine, irinotecan, cetuximab, and bevacizumab, for derived treatment regimens such as FOLFOX, and FOLFIRI. Leucovorin, fluorouracil, and capecitabine are grouped as a "5FU/LV." When patients treated with oxaliplatin or irinotecan receive "5FU/LV," the initiation of 5FU/LV treatment is not considered to be a new line treatment. If the period of exposure to the specific drug is < 60 days, it is classified as belonging to the same line. Subsequent treatment begins with addition of new drugs to drugs of interest.

Result

The narrative pathology report was transformed into JSON form, which contains only information of

interest, such as the name of surgical procedure, histology, location of tumor, profile of biomarker, and regional lymph nodes, except for gross results as shown in Figure 1. In total, 112 patients diagnosed with carcinoma were assigned the ICD-O-3 code according to histology and location. Of those, 37 patients were classified as adenocarcinoma of sigmoid colon (CONCEPT ID: “44502464”), and other concept code patients were distributed to as shown below (Table 1).

Table 1. The distribution of ICD-O-3 diagnosis. (n = 112)

ICD-O-3 diagnosis	Concept code	concept ID	N	(%)
Adenocarcinoma of colon	8140/3-C18.9	44502464	12	10.7
Adenocarcinoma of hepatic flexure of colon	8140/3-C18.3	44501932	5	4.5
Adenocarcinoma in tubulovillous adenoma of ascending colon	8263/3-C18.2	44502946	1	0.9
Adenocarcinoma of transverse colon	8140/3-C18.4	44500927	7	6.3
Tubular adenocarcinoma of rectosigmoid junction	8211/3-C19.9	36526362	1	0.9
Adenocarcinoma of ascending colon	8140/3-C18.2	44502439	9	8.0
Adenocarcinoma of cecum	8140/3-C18.0	44504337	2	1.8
Tubular adenocarcinoma of colon	8211/3-C18.9	36530925	1	0.9
Adenocarcinoma of overlapping lesion of colon	8140/3-C18.8	36561605	4	3.6
Adenocarcinoma of rectum	8140/3-C20.9	44500130	16	14.3
Adenocarcinoma of rectosigmoid junction	8140/3-C19.9	44501075	12	10.7
Carcinoma of transverse colon	8010/3-C18.4	44504361	1	0.9
Adenocarcinoma of sigmoid colon	8140/3-C18.7	44504380	37	33.0
Adenocarcinoma of descending colon	8140/3-C18.6	44500497	4	3.6

We also derived the treatment regimen from drug exposure data. Figure 2 illustrates a fictitious example of tracking a patient’s treatment episode in the extended oncology CDM. Of the 122 AUMC patients, 28 (25%) received FOLFOX as a first-line treatment. Subsequently, 4 (14%) patients received FOLFIRI as a second-line therapy among them.

Conclusion

To our knowledge, this is the first attempt to convert oncology-specific data from EHR to an episode-based oncology extension model of OMOP-CDM in OHDSI network. By leveraging information from the structured pathology reports and the regimen-identifying algorithm, it was possible to populate oncology data in CDM. Further studies are required to apply this process to more scalable settings.

Acknowledgement

This work was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea [grant number : HI16C0992] and the Bio Industrial Strategic Technology Development Program (20003883) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Reference

1. Lau EC, Mowat FS, Kelsh MA, Legg JC, Engel-Nitz NM, Watson HN, et al. Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. Clinical epidemiology. 2011;3:259.

2. Manion FJ, Harris MR, Buyuktur AG, Clark PM, An LC, Hanauer DA. Leveraging EHR Data for outcomes and comparative effectiveness research in oncology. Current Oncology Reports. 2012;14(6):494-501.

3. Bikov KA, Mullins CD, Seal B, Onukwugha E, Hanna N. Algorithm for identifying chemotherapy/biological regimens for metastatic colon cancer in SEER-Medicare. Med Care. 2015;53(8):e58-64.

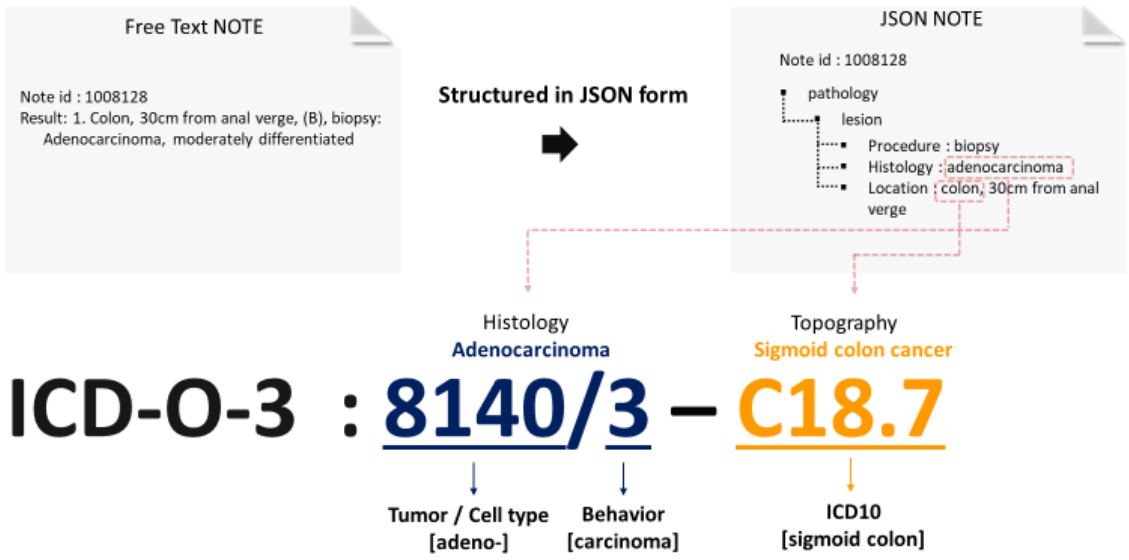

Figure 1. Structuralize pathology reports in JSON format, and extract ICD-O-3 diagnosis.

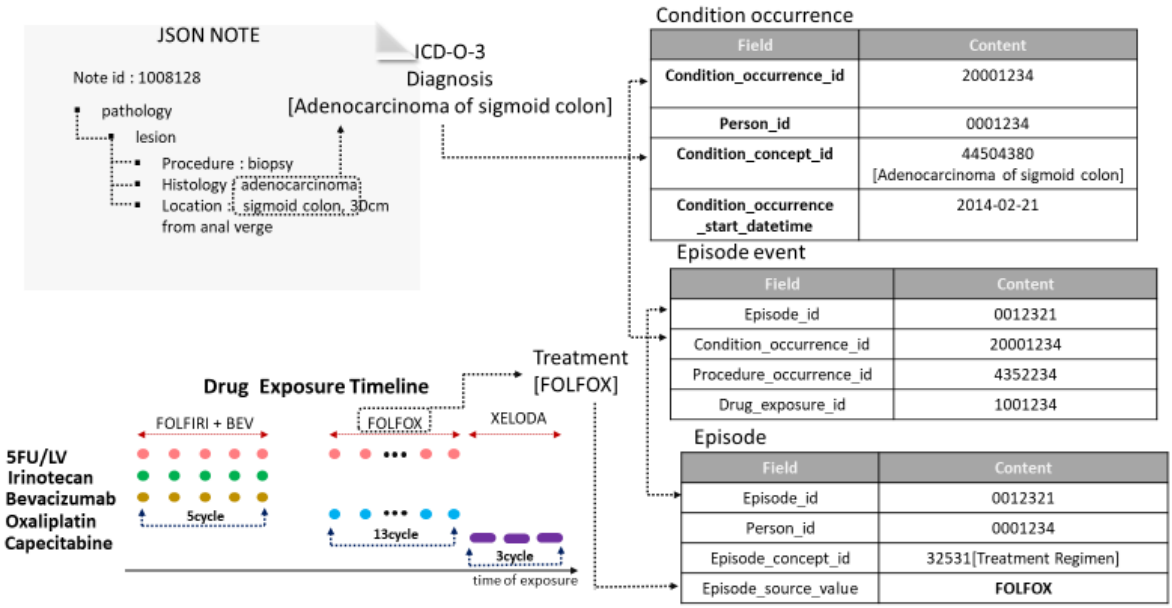

Figure 2. The tracking pathway for a treatment episode.

Abbreviations: JSON = JavaScript Object Notation; FOLFIRI = leucovorin, 5-fluorouracil, irinotecan and oxaliplatin; FOLFOX = folinic acid, fluorouracil, and oxaliplatin; BEV = Bevacizumab; 5FU/LV = Leucovorin, fluorouracil, and capecitabine; XELODA = capecitabine.