JOURNAL OF SURGICAL RESEARCH 187 (2014) 466–470

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Electronic medical record: research tool for pancreatic cancer?

Edward J. Arous, MD,[a] *Theodore P. McDade, MD,*[a]
Jillian K. Smith, MD, MPH,[a] *Sing Chau Ng, MS,*[a] *Mary E. Sullivan, DNP,*[a]
Ralph J. Zottola, PhD,[b] *Paul J. Ranauro, BA,*[b] *Shimul A. Shah, MD,*[a,c]
Giles F. Whalen, MD,[a] *and Jennifer F. Tseng, MD, MPH*[a,d,*]

[a] *Surgical Outcomes Analysis & Research (SOAR), Department of Surgery, University of Massachusetts Medical School, Worcester, Massachusetts*
[b] *Office of Information Systems, Massachusetts Integrated Clinical Academic Research Database, University of Massachusetts Medical School, Worcester, Massachusetts*
[c] *Department of Surgery, University of Cincinnati School of Medicine, Cincinnati, Ohio*
[d] *Department of Surgery, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts*

ARTICLE INFO

Article history:
Received 22 July 2013
Received in revised form
16 October 2013
Accepted 18 October 2013
Available online 21 October 2013

Keywords:
Electronic medical record
Pancreatic cancer
Data warehouse
Database
EMR

ABSTRACT

Background: A novel data warehouse based on automated retrieval from an institutional health care information system (HIS) was made available to be compared with a traditional prospectively maintained surgical database.

Methods: A newly established institutional data warehouse at a single-institution academic medical center autopopulated by HIS was queried for International Classification of Diseases, 9th Revision, Clinical Modification (*ICD-9-CM*) diagnosis codes for pancreatic neoplasm. Patients with *ICD-9-CM* diagnosis codes for pancreatic neoplasm were captured. A parallel query was performed using a prospective database populated by manual entry. Duplicated patients and those unique to either data set were identified. All patients were manually reviewed to determine the accuracy of diagnosis.

Results: A total of 1107 patients were identified from the HIS-linked data set with pancreatic neoplasm from 1999–2009. Of these, 254 (22.9%) patients were also captured by the surgical database, whereas 853 (77.1%) patients were only in the HIS-linked data set. Manual review of the HIS-only group demonstrated that 45.0% of patients were without identifiable pancreatic pathology, suggesting erroneous capture, whereas 36.3% of patients were consistent with pancreatic neoplasm and 18.7% with other pancreatic pathology. Of the 394 patients identified by the surgical database, 254 (64.5%) patients were captured by HIS, whereas 140 (35.5%) patients were not. Manual review of patients only captured by the surgical database demonstrated 85.9% with pancreatic neoplasm and 14.1% with other pancreatic pathology. Finally, review of the 254 patient overlap demonstrated that 80.3% of patients had pancreatic neoplasm and 19.7% had other pancreatic pathology.

Conclusions: These results suggest that cautious interpretation of administrative data rely only on *ICD-9-CM* diagnosis codes and clinical correlation through previously validated mechanisms.

---

[*] *Corresponding author*. Department of Surgery, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Stoneman 9, Boston, MA 02215. Tel.: +1 617 667 3746; fax: +1 617 667 7756.

E-mail address: jennifer.tseng@mac.com (J.F. Tseng).

0022-4804/$ – see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jss.2013.10.036

1. Introduction

Electronic medical records are becoming progressively more important for both clinical care and research. In the current era, data warehouses and repositories have been established to serve the dual function of patient care and investigation. In theory, medical data warehouses may facilitate clinical research, including assessing and improving quality and outcomes.

Currently, outcomes and health services research can be based on a number of different data sources, including individual, department, or institutional databases; patient registries; administrative databases; and, more recently, data warehouses. All these have their own respective advantages and disadvantages. Large national databases, including the National Surgical Quality Improvement Program and the Nationwide Inpatient Sample, share some common ground with data warehouses as they are in part based on International Classification of Diseases, 9th Revision, Clinical Modification (*ICD-9-CM*) coding. Recently, at our institution, a newly established health care information system (HIS) data warehouse was established and made available to pilot investigators [1,2].

Large computerized health care system databases have been touted as the research tools of the future [1,3,4]. However, manually maintained departmental databases still exist. Our research group, Surgical Outcomes Analysis & Research, has previously analyzed different types of data resources [5,6]. Manasanch *et al*. [6] demonstrated that a significant number of pancreatic neuroendocrine tumors are missed via tumor registry's use of *ICD-9-CM* diagnosis codes [7]. Other previous studies have analyzed the effectiveness of data collection systems, such as tumor registries with varying results [8–11]. The aim of the present study was to determine whether clinical research could be facilitated using a novel data warehouse, using pancreatic cancer as a model.

2. Methods

2.1. HIS-linked database

Encrypted records from a newly developed HIS-linked institutional clinical data warehouse, capturing inpatient and outpatient clinical and billing information from a pool of 2.05 million patients evaluated at an academic medical institution and its affiliates since 1995, were queried. This database, still currently under development, allows investigators to use a web-based query tool to determine the aggregate total number of patients at network hospitals who meet a given set of inclusion and exclusion criteria (including demographics, diagnoses, medications, and laboratory values) [1,2]. After institutional review board approval, demographic and clinical data for 1198 patients were captured with *ICD-9-CM* diagnosis codes of pancreatic cancer and/or neoplasm (Table 1). To maintain a contemporary focus, patients outside of the years 1999–2009 were excluded.

2.2. Surgery database

To serve as a valid comparison for the novel data warehouse, we used the pancreatic subset of an institutional review board–approved longitudinal prospectively maintained gastrointestinal tumors database. The pancreatic surgery database (SURG) includes all patients referred to and evaluated in the surgical oncology/pancreatic surgery clinic with suspected pancreatic cancer since 1999. For the purposes of the present study, patients with pancreatic diagnoses evaluated by surgery from 1999–2009 were identified.

Table 1 – *ICD-9-CM* diagnosis codes used in this study.

ICD-9-CM diagnosis code	Diagnosis
157.0	Malignant neoplasm of head of pancreas
157.1	Malignant neoplasm of body of pancreas
157.2	Malignant neoplasm of tail of pancreas
157.3	Malignant neoplasm of pancreatic duct
157.4	Malignant neoplasm of islets of langerhans
157.8	Malignant neoplasm of other unspecified sites of pancreas
157.9	Malignant neoplasm of pancreas, part unspecified
211.6	Benign neoplasm of pancreas, except islets of langerhans
211.7	Benign neoplasm of islets of langerhans

2.3. Data collection

Both the HIS and the SURG databases were queried for nine primary and secondary *ICD-9-CM* discharge diagnosis codes for pancreatic neoplasm (Table 1). Patients were subsequently limited to the years 1999–2009, inclusive. All patients were identified as being unique to either data set or both.

2.4. Statistical analysis

First, for validation purposes, all patients were manually reviewed to determine the diagnosis and assess whether pancreatic tumor or pancreatic pathology were present. The manual review processes analyzed medical records including clinic notes, admission/discharge notes, diagnostic imaging, and pathology reports.

3. Results

From an initial HIS-linked database of 2.05 million patients and 121 million data points, 1198 were identified using our search query for pancreatic neoplasm. This number was subsequently winnowed to 1107 patients on limiting to the years 1999–2009. Along the same period, 444 patients were identified in the prospectively maintained surgical database. This was subsequently narrowed to 394 patients with pancreatic neoplasm over the years 1999–2009 (Fig. 1).

Of the 1107 HIS-linked patients, 254 (22.9%) patients were also captured by the SURG database, whereas 853 (77.1%) patients were only in the HIS-linked data set. Patients identified in either database or found in both databases had similar ages, sex, and race (Table 2). The HIS-only database validation underwent complete manual review demonstrating 45.0% of patients without identifiable pancreatic pathology, whereas