VA-INPC: Linking Department of Veterans Affairs (VA) and Indiana Network for Patient Care (INPC) data to assess surveillance testing among veterans with colorectal cancer

David A. Haggstrom, MD, MAS[1-3], Marc Rosenman, MD[2,4], Laura J. Myers, PhD[1,3], Evgenia Teal, MS[2], Bradley N. Doebbeling, MD, MSc[1-3]

[1]VA HSR&D Center on Implementing Evidence-Based Practice, Roudebush VAMC, Indianapolis, IN; [2]Regenstrief Institute, Inc [3]Division of General Internal Medicine and Geriatrics, Department of Medicine, IU School of Medicine, Indianapolis, IN; [4]Department of Pediatrics, IU School of Medicine, Indianapolis, IN

Abstract

The goal of this project was to provide empiric evidence about the benefit to US veterans and the VA of capturing data from a citywide clinical informatics network (INPC) to assess care received outside the VA. We identified 468 veterans diagnosed with colorectal cancer from 2000-2007 in the Indianapolis VA cancer registry. Electronic VA healthcare data were linked with electronic health records from the regional health information organization (RHIO) INPC; 341 matches were found. Both the VA and INPC systems were queried regarding receipt of surveillance tests. The proportion with additional data from INPC varied by test: colonoscopy (3%), CT scan/abdomen (13%), CT scan/chest (79%), carcinoembryonic antigen test (8%), and other laboratory tests (25%-53%). An incremental benefit of linking VA and INPC data was present and may increase when expanded beyond patients with a single condition. New, important information about care outside the VA is obtained through RHIO data linkage.

Introduction

Data linkages of health care information have often been of significant value to medical researchers and clinicians. Creative linkages enable the connection of sets of data which taken together provide information that neither source of data can alone. In cancer research, cancer registry data have been linked with administrative claims data from both private[1] and public payors[2] to understand more about the health care services delivered to cancer patients. The most widely used such linkage is the Surveillance Epidemiology and End Results (SEER)-Medicare database. Data from the VA has also been linked with Medicare data to enable better ascertainment of care delivered to veterans outside the VA, given that many veterans are dually eligible for Medicare and VA services.[3] Limitations of such data include the fact that not all veterans meet age eligibility requirement for Medicare services.

As the largest integrated delivery system in the United States, the VA has been recognized as a leader in developing a more coordinated system of care. Beginning in the early 1990s, VA leadership instituted both an advanced electronic health record system and a quality improvement approach that holds regional managers accountable for multiple performance measures.

Meanwhile, the Regenstrief Institute, Inc., has developed the Indiana Network for Patient Care (INPC), a secure operational citywide RHIO organized by patient in order to enhance clinical care and foster quality of care improvement and research in Indianapolis.[4] The INPC maintains clinical data from public health departments, local laboratories, imaging centers, large group practices associated with hospitals, and more than 35 hospitals across Indiana. Clinical data about individual patients can be aggregated into a single virtual medical record. For this study, we used INPC data from the five major hospital systems in the Indianapolis metropolitan area. Together they generate 2.7 million outpatient visits per year. The Indianapolis VA is the only major healthcare system in Indianapolis which does not participate in the INPC.

The creation of a VA-INPC data link in order to more fully capture care delivered to veterans in the community is potentially very important for multiple reasons. First, obtaining information from outside the VA may improve the quality of care delivery for the individual, such as ensuring better communication between providers, enabling a system-wide view of quality, and reducing the risk of undergoing multiple imaging tests. Second, the identification of redundant testing services may improve efficiency and reduce the cost of care. In addition, including the measurement of care delivered in the community would inform both health services research, as well as

many clinical research studies. Finally, whereas all VA medical centers (VAMCs) are evaluated based upon quality of care performance measures, we have the potential to demonstrate the benefit to the VA of capturing previously unmeasured care that occurs in the community.

For example, gaps in evidence-based care have been documented for CRC survivors nationally both in the VA[5] and community practice.[6] Approximately 2,200 veterans are diagnosed with CRC each year.[7] Three recent high-quality meta-analyses reported a 20%-33% relative risk reduction of all-cause mortality (7% absolute risk reduction) for individuals who received more intensive follow-up after primary treatment for CRC.[8-10] These follow-up programs include periodic colonoscopy procedures, liver imaging (including CT scans), and blood carcinoembryonic antigen (CEA) testing. Clinical guidelines from the American Society of Clinical Oncology also changed in November 2005 to recommend annual CT of the abdomen/chest for the first time.[11] Furthermore, American Society of Clinical Oncology (ASCO) guidelines also outline tests that are not recommended for routine surveillance, including complete blood counts (CBCs) and liver function tests (LFTs). Of note, both guideline-concordant and non-guideline concordant surveillance tests may sometimes be done for diagnostic purposes among patients with signs or symptoms of cancer recurrence.

The overall goal of this project was to evaluate the feasibility and value of linking VA and RHIO data, in order to provide empiric evidence about the benefit of capturing data from veterans seen in community health care systems. The specific aim was to determine whether assessments of CRC surveillance received by US veterans cared for at the Indianapolis VA would change when considering care received outside the VA, in the greater Indianapolis area. To meet this aim, we created a VA-INPC dataset, linking a cohort of patients diagnosed with CRC at the Indianapolis VA to additional health care data from the INPC. CRC care delivered inside the VA was assessed using the Indianapolis VA cancer registry and electronic medical record (VISTA) data, while care delivered outside the VA was assessed using data from the INPC.

Methods

Study Population: We identified patients with incident colorectal cancer cared for at the Indianapolis VAMC and diagnosed from 2000-2007. A total of 468 individuals were identified in the Indianapolis VA cancer registry. Patients with pre-existing tumors were excluded.

Approval Process: We developed data security and data use agreements, reviewed and approved by review boards of both organizations. The study was approved by the Indiana University-Purdue University Indianapolis Institutional Review Board, the Indianapolis VA Medical Center Research Committee, and the INPC Management Committee.

Data linkage: We then linked identifiers from VA CRC patients to identifiers from individuals receiving care in the INPC.

The VA-INPC linkage was performed using a deterministic linkage, using an exact match of social security number (SSN) for the initial potential match: 335 matches were found (72% of the original VA sample). The VA-INPC linkage created no multiple matches (i.e., scenarios in which a single record from the VA matched multiple records from the INPC). Individuals were considered a match if there was complete agreement on 3 or more of the following (SSN, first name, last name, date/month/year of birth). For all of these initial putative matches, two investigators independently reviewed last name, first name, and date of birth to confirm the appropriateness of the match; during this process, 12 imperfect (probabilistic) matches that had a reasonable match with 3 of the above parameters (SSN, first name, last name, date/month/year of birth) were identified and one individual was excluded.

The VA-INPC linkage was then performed a second time using a combination of the same last name (first 6 positions), same first name (first 6 positions), and month of birth: 7 additional matches were found, for a total of 341 matches. After the second linkage, in order to be considered a match, one of the following conditions was required: agreement on 7 or 8 digits of the SSN or agreement on 2 or more of the following: year of birth, day of birth, or middle initial. The two investigators independently reviewed and confirmed the appropriateness of the additional potential matches. This data linkage method was unique to this study and does not represent the methodology used across actively participating INPC institutions.

Data sources: We queried the following data from the INPC: demographic file, procedures (colonoscopy), imaging (CT scans), and laboratory results. In the VA system, we queried the VA cancer registry and electronic medical record (VISTA) data, which contains information on demographics, diagnoses, procedures, and laboratory tests, from both inpatient and outpatient settings. VA surveillance data was collected within one year after diagnosis due

to feasibility. No time limit was set upon the interval over which INPC data was gathered.

Measures: We measured the receipt of the following types of guideline-concordant surveillance tests: colonoscopy, CEA tests, chest CT, and abdomen (with or without pelvis) CT. We also measured the following non-guideline concordant surveillance laboratory tests: CBCs (including at least serum hemoglobin) and LFTs (either alkaline phosphatase, SGOT, SGPT, or serum bilirubin).

Data Analysis: Demographic and clinical characteristics of the study population with CRC identified in the Indianapolis VA cancer registry are provided. Descriptive statistics of the VA-INPC linkage dataset are reported at both the patient and the test level; a single patient could receive more than one test. The number and proportion of (a) patients and (b) tests with surveillance data from VA data alone, both VA and INPC data, and INPC data alone are described. Finally, we calculated the *additional value of INPC* data by dividing the number of patients for whom INPC data was available (numerator) by the number of patients or tests for whom some type of electronic data (VA or INPC) was available.

Results

Study population: The mean age of the CRC patients was 67 years. Most were male (96%), and 11% were African-American. These sociodemographics were comparable to those receiving care in VAMCs in the Midwest. Most CRC patients had adenocarcinoma (86%). The pathologic cancer stage distribution of the study population was as follows: Stage I (23%), Stage II (16%), Stage III (22%), Stage IV (16%), and unknown (22%).

The receipt of different surveillance tests from different data sources varies by test (Table). The proportion of patients who received tests from a given data source(s) was usually similar to the proportion of tests from a given data source(s), although the distribution could differ. For example, across the denominator of all patients who received LFT lab tests, 47% of patients had their tests identified in VA data alone, 46% in INPC data alone, and 7% in both VA and INPC data. Across the denominator of all LFT lab tests collected, 21% of tests were identified in VA data alone, 71% in INPC data alone, and 8% in both VA and INPC data.

Additional value of INPC data: The additional value of the linked INPC data varied by type of test. The proportion of the study population with additional data from the INPC by test were the following: colonoscopy (3%), CT scan/abdomen (13%), CT scan/chest (79%), CEA test (8%), CBCs (25%), and LFTs (53%). At the test level, the proportion of all tests that came from the INPC were the following: colonoscopy (3%), CT scan/abdomen (13%), CT scan/chest (86%), CEA test (6%), CBCs (28%), LFTs (79%).

Discussion

Our study demonstrates that data linkage of VA patient records with their records from outside the VA at health care organizations represented in a citywide RHIO is both feasible and informative. New clinical and performance information is obtained about care received by veterans that cannot be obtained from VA administrative, EHR and electronic registry data alone. An incremental benefit of linking

Surveillance test categories	# of patients (% of pts receiving test from data source)			Total patients (% of eligible)	# of surveillance tests (% of tests received from data source)			Total tests
	VA data alone	both VA and INPC data	INPC data alone		VA data alone	both VA and INPC data	INPC data alone	
Procedure and imaging tests								
Colonoscopy	150 (97%)	4 (3%)	0	154	190 (97%)	5 (3%)	0	195
CT scan (abdomen)	202 (87%)	17 (7%)	12 (5%)	231	369 (87%)	40 (9%)	16 (4%)	425
CT scan (chest)	4 (21%)	0	15 (79%)	19	4 (14%)	0	25 (86%)	29
Lab tests								
Carcinoembryonic antigen (CEA) test	196 (92%)	10 (5%)	6 (3%)	212	524 (95%)	23 (4%)	12 (2%)	559
Complete blood count (CBC) test	223 (75%)	49 (17%)	24 (8%)	296	1823 (72%)	416 (16%)	286 (11%)	2525
Liver function test (LFT) test	45 (47%)	7 (7%)	44 (46%)	96	62 (21%)	24 (8%)	206 (71%)	292

VA and INPC data was present, varied by the test of interest, and may increase when expanded beyond patients with a single condition and as INPC data capture increases.

The additional value of INPC (RHIO) data varied by type of surveillance test. Differences in care measured outside the VA may vary either because of the nature of the way clinical data are collected or because of the actual health care received. The INPC made the greatest relative contribution to data regarding laboratory tests, perhaps because these data were captured directly from central laboratories serving multiple hospitals, thus resulting in more complete data collection. Furthermore, an inpatient hospitalization outside the VA,[12] involving daily routine laboratory tests (CBCs and LFTs), may increase the proportion of laboratory tests collected outside the VA substantially. Colonoscopy tests may be underrepresented in INPC due to their delivery in freestanding endoscopy facilities that may not share data with hospital information systems integrated into the INPC. The one-year time interval over which VA data was collected may also have depressed the proportion of eligible patients who received colonoscopy among our cohort (33%). In either case, veterans likely receive care not captured by our methods given the higher receipt colonoscopy among colorectal cancer survivors in previous studies (55-64%, over varying time intervals).[13-15]

According to a previous study of veterans,[3] for outpatient service use, 18% were VA-only users, 36% were Medicare-only users, and 46% were both VA and Medicare users. Among veterans with inpatient use, 24% were VA only, 69% were Medicare only, and 6% were both VA and Medicare users. With the VA-INPC data linkage, we discovered less use of surveillance care outside the VA. One key explanation for this may be selection bias, that is, our initial cohort of patients was identified in the VA cancer registry, and patients identified in this manner may be more likely to receive all or most of their care in the VA.

The value of additional RHIO data may vary by health condition and will be greater, in absolute terms, when more patients are included in the linkage. For example a data linkage involving more types of cancer or a higher prevalence disease, such as diabetes, may provide a greater increase in additional information. We did not examine the value of obtaining data from INPC specifically for those with an Indianapolis or Marion county address, although limiting the analyses to those residing in the city of Indianapolis would likely have shown higher rates of additional information. On the other hand, the differing time frames over which data was collected from INPC (no limit) and the VA (one year) may have raised the relative proportion of additional information found outside the VA.

From the perspective of VA operational leadership, any additional data capture of care received from INPC institutions outside the VA is greater operational knowledge than was present before, and therefore, an improvement. Given the VA's commitment to "providing the best care anywhere", this study may furnish evidence of the value of data sharing with RHIOs to both deliver more comprehensive care in the community and the VA, as well as more complete estimates of performance across both settings. From both the INPC and VA perspective, striking the right balance between community data exchange – for clinical care, research and public health – and patient and institutional privacy considerations will be of ongoing importance. The approach outlined in this study used human review; in order to scale this project to substantially increase the amount of data available, more advanced approaches would be necessary, but are certainly available, to perform high volume record matching.[16]

Future directions in the use of this type of data linkage include further characterization of how the use of data from multiple health care systems influences the operational functions of participating health care organizations. Notably, the addition of RHIO data may influence assessment of the quality of care delivered to veterans. Pooled results of the economic impact across patients and tests may also allow the VA and outside organizations to reduce redundancy and realize potential cost savings for both the health care system and patient. Including data on tests obtained within the community may correctly account for those individuals who refuse to be tested again within the VA, when the test was already obtained and documented in the community. In addition, the incorporation of INPC data may identify instances where redundant testing takes place, another practical example of the value of health information exchange.

Conclusions

Electronic clinical data from outside the VA informs a more complete understanding of both the quality and cost of care delivered to patients in the VA. Our local project experience with data linkage and information exchange mirrors the experience nationally. Currently, there is no regular mechanism for information exchange of real-time data between

the VA and outside organizations. Demonstration projects such as ours lay the groundwork for establishing the processes, feasibility, and value of such data linkages. Pursuing opportunities for live health information exchange between the VA (or Department of Defense) and RHIOs would be promising future activity.

Acknowledgements

This research was supported by a VA Young Investigator Award from the Indiana Medical Institute and Department of Defense grant W81XWH-08-1-0065. Dr. Haggstrom is the recipient of VA HSR&D Career Development Award CD207016-2 and part of the VA/Robert Wood Johnson Foundation Physician Faculty Scholars Program. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

References

1. Doebbeling BN, Wyant DK, McCoy KD, et al. Linked insurance-tumor registry database for health services research. *Med Care.* Nov 1999;37(11):1105-1115.

2. Bradley CJ, Given CW, Luo Z, Roberts C, Copeland G, Virnig BA. Medicaid, Medicare, and the Michigan Tumor Registry: a linkage strategy. *Med Decis Making.* Jul-Aug 2007;27(4):352-363.

3. Hynes DM, Koelling K, Stroupe K, et al. Veterans' access to and use of Medicare and Veterans Affairs health care. *Med Care.* Mar 2007;45(3):214-223.

4. McDonald CJ, Overhage JM, Barnes M, et al. The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. *Health Aff (Millwood).* Sep-Oct 2005;24(5):1214-1220.

5. El-Serag HB, Petersen L, Hampel H, Richardson P, Cooper G. The use of screening colonoscopy for patients cared for by the Department of Veterans Affairs. *Arch Intern Med.* Nov 13 2006;166(20):2202-2208.

6. Cooper GS, Koroukian SM. Geographic variation among Medicare beneficiaries in the use of colorectal carcinoma screening procedures. *Am J Gastroenterol.* Aug 2004;99(8):1544-1550.

7. Veterans Affairs Central Cancer Registry. November 20, 2006; http://www1.va.gov/cancer/page.cfm?pg=17. Accessed December 1, 2006.

8. Figueredo A, Rumble RB, Maroun J, et al. Follow-up of patients with curatively resected colorectal cancer: a practice guideline. *BMC Cancer.* Oct 6 2003;3:26.

9. Renehan AG, Egger M, Saunders MP, O'Dwyer ST. Impact on survival of intensive follow up after curative resection for colorectal cancer: systematic review and meta-analysis of randomised trials. *Bmj.* Apr 6 2002;324(7341):813.

10. Jeffery GM, Hickey BE, Hider P. Follow-up strategies for patients treated for non-metastatic colorectal cancer. *Cochrane Database Syst Rev.* 2002(1):CD002200.

11. Desch CE, Benson AB, 3rd, Somerfield MR, et al. Colorectal cancer surveillance: 2005 update of an American Society of Clinical Oncology practice guideline. *J Clin Oncol.* Nov 20 2005;23(33):8512-8519.

12. West AN, Weeks WB, Wright SM, Wallace AE, Fisher ES. When VA patients have non-VA hospitalizations, who pays for what services, and what are the research implications? A New York case study. *Med Care.* Aug 2008;46(8):872-877.

13. Cooper GS, Schultz L, Simpkins J, Lafata JE. The utility of administrative data for measuring adherence to cancer surveillance care guidelines. *Med Care.* Jan 2007;45(1):66-72.

14. Fisher DA, Jeffreys A, Grambow SC, Provenzale D. Mortality and follow-up colonoscopy after colorectal cancer. *Am J Gastroenterol.* Apr 2003;98(4):901-906.

15. Knopf KB, Warren JL, Feuer EJ, Brown ML. Bowel surveillance patterns after a diagnosis of colorectal cancer in Medicare beneficiaries. *Gastrointest Endosc.* Nov 2001;54(5):563-571.

16. Zhu VJ, Overhage MJ, Egg J, Downs SM, Grannis SJ. An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. *J Am Med Inform Assoc.* Sep-Oct 2009;16(5):738-745.