Defining Colorectal Cancer Recurrence with Structured-Health Data Algorithms: A Targeted Literature Review

Spencer Cheng, PharmD, Postdoctoral Fellow[1]; Aasthaa Bansal, PhD[1]; David Veenstra, PharmD, PhD[1]
1. CHOICE Institute, School of Pharmacy, University of Washington, Seattle, WA.

THE CHOICE INSTITUTE
School of Pharmacy

BACKGROUND & RATIONALE

> **Recurrent cancer** is defined as the reoccurrence of disease after a period of undetectable disease[1]

> About **50%** of patients with **colorectal cancer** (CRC) develop recurrent disease after surgical resection within 2 years[2]

> Population-level structured-health data (e.g., administrative claims, electronic medical records, registries) often **lacks recurrence status** making it difficult to adequately identify this population[3]

> Several algorithms have been developed to detect recurrent cancer in structured-health data, but methods and validation **vary**, thus potentially **misclassifying** these patients and treatment outcomes

> Reviewing the components and validation results of published CRC algorithms may help guide the **standardization** and **accuracy** of CRC detection in **real-world data**

OBJECTIVE

> **Identify the current approaches to developing and validating structured-health data algorithms used to define disease recurrence in patients with CRC**

Conclusion & Implications

> Recent algorithms for estimating CRC recurrence using population-level health data are **limited** in number and **heterogenous**

> All validated algorithms had **high specificity**, but **sensitivities** were **low and variable** (Figure 2)

> Low sensitivities could result in an **underestimation** of CRC recurrence impacting disease burden measurements and outcomes studies

> There is a need for standardization of detection algorithms with higher sensitivities to accurately identify this patient population in structured-health data

Abbreviations: CRC = colorectal cancer, EMR = electronic medical record, EHR = electronic health record, ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification, DRG = Diagnosis-Related Groups, CPT= Current Procedure Terminology , HCPCS = Healthcare Common Procedure Coding System, NDC = national drug code

METHODS

Targeted literature review (PubMed)

Key inclusion criteria:

> Observational or validation study (original source of algorithm)

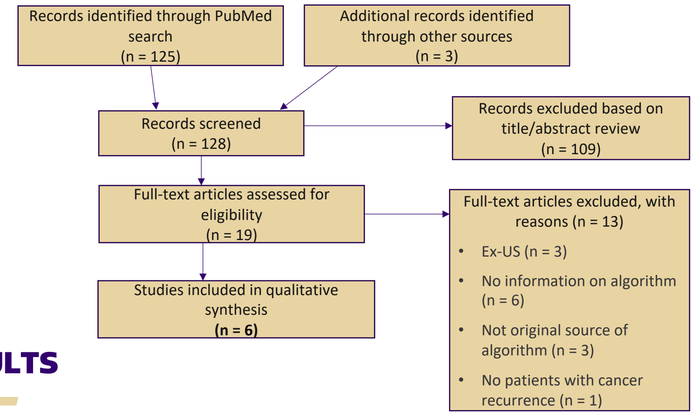
> Study population: US patients with CRC recurrence

> Data source(s): administrative claims, EMR/EHR, cancer registries

> Algorithm defines or estimates recurrence status (not timing)

> Published within past 10 years

Figure 1. PRISMA Flow Diagram



RESULTS

Table 1. Study Characteristics

Reference	Cancer type	Data source	# of CRC patients	Median follow-up, mo	Period of patient inclusion	Identification of gold-standard recurrence
Anaya (2012)[4]	CRC	Veterans Administration	308	-	1997-2008	chart review
Hassett (2014)[5]	CRC or others	Medicare and Kaiser Permanente*	1,708	60	2000-2005	chart review
Deshpande (2015)[6]	CRC	Barnes-Jewish Hospital	174	-	2005-2009	chart review
Warren (2016)[7]	CRC or breast	Medicare	6,910	60	1994–2003	none
Hassett (2017)[3]	CRC or lung	Medicare and Kaiser Permanente*	3,427	19	2000-2012	chart review
Le (2020)[8]	Colon	California Cancer Registry	2,077	31-46	2004-2012	chart review

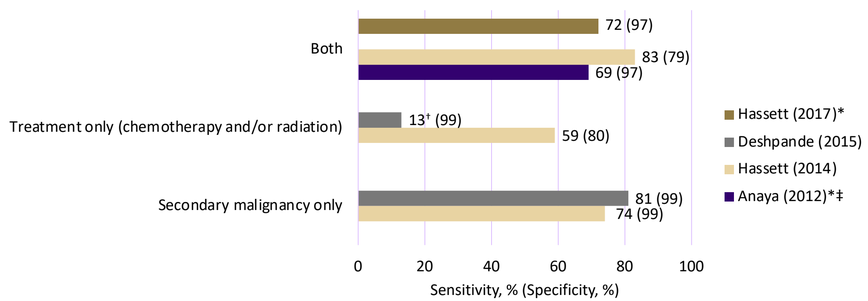
*Kaiser Permanente Colorado and Kaiser Permanente Northwest

Table 2. Algorithm Definition and Validation

Reference	Model type	Recurrence algorithm details	Specificity, % (95% CI)	Sensitivity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
Anaya (2012)[4]	Logistic Regression*	**Colorectal liver metastasis** and/or **partial hepatectomy**	97	69	87	92
Hassett (2014)[5]	Rules‡	1) Secondary malignancy **excluding nodes** 2) **Chemotherapy only ≥ 12 mo** after index diagnosis 3) Algorithm 1) or 2)	99 (98-99) 80 (77-83) 79 (76-82)	74 (68-80) 59 (52-66) 83 (77-88)	94 45 53	NR
Deshpande (2015)[6]	Rules	1) Secondary malignant neoplasm **≥ 3 mo** after index surgery 2) **Chemotherapy** and/or **radiation ≥ 8 mo** after surgery 3) **Chemotherapy ≥ 16 mo** and/or **radiation ≥ 12 mo** after surgery	99 (96-100) 100 (97-100) 99 (95-100)	81 (63-92) 6 (1-22) 13 (4-30)	96 (79-100) 100 (20-100) 67 (24-94)	96 (91-98) 83 (76-88) 83 (77-89)
Warren (2016)[7]	Rules	Mutually exclusive hierarchy for first indicator of recurrence: **additional therapy**, diagnosis of metastatic disease, or **end-of-life care**	NR	NR	NR	NR
Hassett (2017)[3]	Logistic Regression†	Components: secondary malignancy **including nodes, radiation ≥ 12 mo** after index diagnosis, **hospice, imaging (# per year)**	97 (96-99)	72 (58-81)	83 (78-92)	96 (94-97)
Le (2020)	Rules§	**Peritoneal involvement** occurring **≥ 6 mo** after initial diagnosis of colon cancer	NR	NR	NR	NR

*Algorithm specific to colorectal liver metastasis
‡ Validation of model reported for Kaiser Permanente sample
† Validation of model reported for Medicare sample and performance results maximized accuracy (94%, 95% CI 92-96) with an ROC-AUC of 0.95 (95% CI, 0.92-0.97)
§ Algorithm specific to metachronous peritoneal metastasis

Figure 2. Sensitivity and Specificity by Algorithm Component



Sensitivities were not reported for two studies (Le and Warren)
*Logistic regression model (Anaya 2012, Hassett 2017), other papers used rule-based algorithms
‡ Algorithm specific to colorectal liver metastasis
† Sensitivity reported is algorithm 3 (Deshpande 2015), algorithm 2 had lower sensitivity (6%, 95% CI 1.1 – 22.2)

Summary of Results

> Algorithm components included diagnoses of **secondary malignancies, treatments/procedures, or both**
> Each algorithm used a **different combination** of codes and formats (ICD-9-CM, DRG, CPT, HCPC, NDC) to define components
> Two studies did **not validate** the recurrence algorithm used to identify patients (Warren, Le)
> The inclusion of a higher number and different formats of codes resulted in higher specificities, but lower sensitivities
> The **specificity** was generally **high** across the different algorithms, but **sensitivities** were highly variable **between and within** model type and algorithm components

References 1. Izci H, Tambuyzer T, Tuand K, et al. A Systematic Review of Estimating Breast Cancer Recurrence at the Population Level With Administrative Data. *J Natl Cancer Inst*. 2020;112(10):979-988. doi:10.1093/jnci/djaa05. 2. Figueredo A, Rumble RB, Maroun J, Earle CC. Follow-up of patients with curatively resected colorectal cancer: a practice guideline. BMC Cancer 2003;3:26. 3. Hassett MJ, Uno H, Cronin AM, Carroll NM, Hornbrook MC, Ritzwoller D. Detecting Lung and Colorectal Cancer Recurrence Using Structured Clinical/Administrative Data to Enable Outcomes Research and Population Health Management. *Med Care*. 2017;55(12):e88-e98. 4. Anaya DA, Becker NS, Richardson P, Abraham NS. Use of administrative data to identify colorectal liver metastasis. *J Surg Res*. 2012;176(1):141-146. doi:10.1016/j.jss.2011.07.022. 5. Hassett MJ, Ritzwoller DP, Taback N, et al. Validating billing/encounter codes as indicators of lung, colorectal, breast, and prostate cancer recurrence using 2 large contemporary cohorts. *Med Care*. 2014;52(10):e65-e73. doi:10.1097/MLR.0b013e318277eb6f. 6. Deshpande AD, Schootman M, Mayer A. Development of a claims-based algorithm to identify colorectal cancer recurrence. *Ann Epidemiol*. 2015;25(4):297-300. doi:10.1016/j.annepidem.2015.01.005. 7. Warren JL, Mariotto A, Melbert D, et al. Sensitivity of Medicare Claims to Identify Cancer Recurrence in Elderly Colorectal and Breast Cancer Patients. *Med Care*. 2016;54(8):e47-e54. doi:10.1097/MLR.0000000000000058. 8. Le VH, Thornblade L, Ituarte PHG, Lai LL, Melstrom KA. Metachronous peritoneal metastases following curative resection for colon cancer: Understanding risk factors and patterns of recurrence. *J Surg Oncol*. 2021;123(2):622-629. doi:10.1002/jso.26322