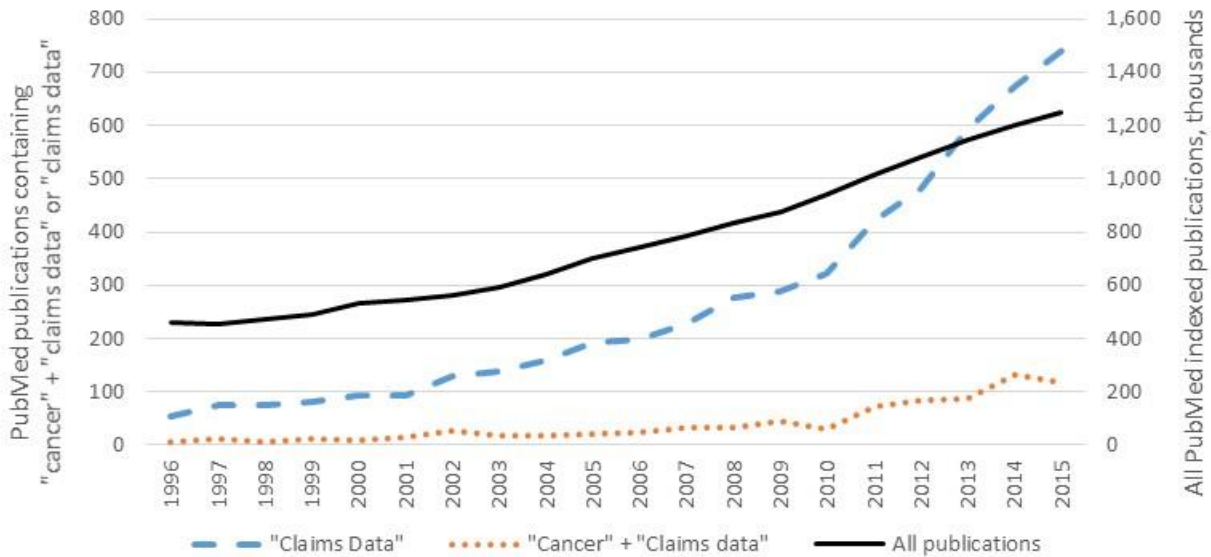
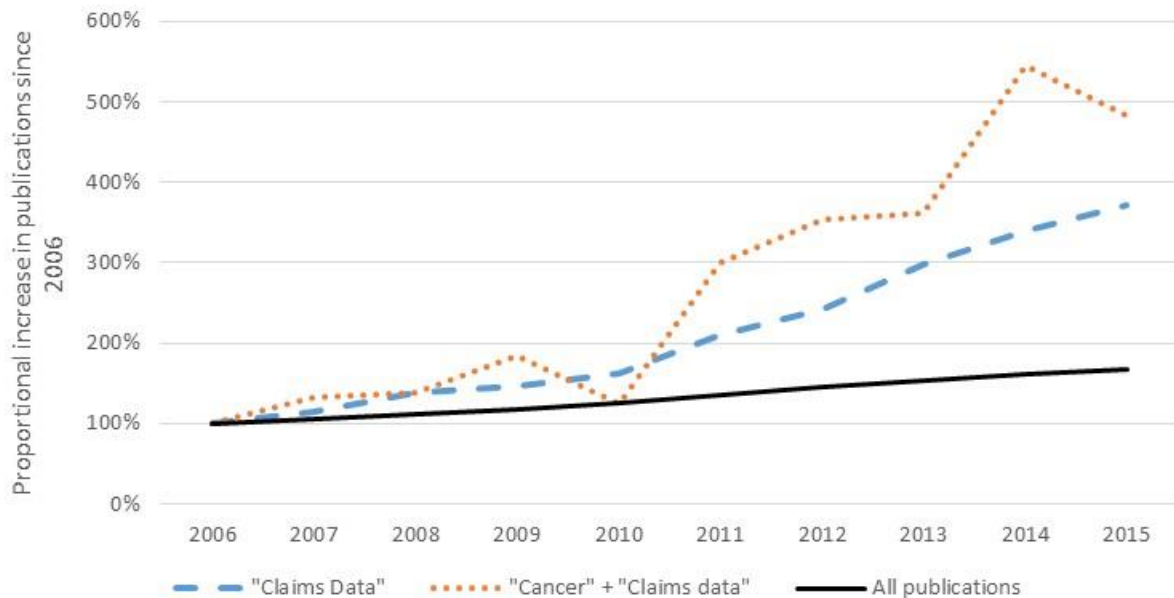


## Supplementary materials

**Supplementary figure S1:** Number (A) and proportional growth over time (B) of publications overall, those containing the phrase “claims data”, and those with both terms “cancer” and “claims data”.



(A) As the use of claims data in research overall has grown significantly during the last 20 years, and more recently in the last 10 years for cancer research, specifically.



(B) In the last decade yearly publications of cancer research using claims data has grown 5-fold and those using claims data in general has grown nearly 4-fold, while publications in general have less than doubled.

Data used to create graphs was obtained from: <http://dan.corlan.net/medline-trend.html>

## MEDICAL RECORD ABSTRACTION FORM

### SECTION 1: BASIC PATIENT INFORMATION AND CANCER TYPE

**Q1: Chart ID:**

**Q2: Patient gender (Please choose ONLY ONE of the following option):**

- ☐ Male
- ☐ Female
- ☐ Not specified

**Q3: According to the medical chart, what cancer was the patient initially diagnosed with? (Please choose ONLY ONE of the following option):**

- ☐ Breast cancer (Go to SECTION 2)
- ☐ Lung cancer (Go to SECTION 3)
- ☐ Colorectal cancer (Go to SECTION 4)
- ☐ None of the above (End of the abstraction)
- ☐ Unknown/No cancer was diagnosed (End of the abstraction)

## SECTION 2: BREAST CANCER RELATED QUESTIONS

**Note: Please only fill out the questions in the section if “Breast Cancer” in Section 1, Q6 is selected. Otherwise, please skip this section.**

**Q1: According to the medical chart, what is the most recent stage (based on TNM staging) for the diagnosed breast cancer? (Please choose ONLY ONE of the following option)**

- ☐ Stage 0
- ☐ Stage I
- ☐ Stage II
- ☐ Stage III
- ☐ Stage IV
- ☐ Unknown / Not Documented

**Q2: According to the medical chart, what is the patient's most recent breast cancer stage (based on TNM staging) at a more granular level? (Please choose ONLY ONE of the following option)**

- ☐ Stage 0
- ☐ Stage I
- ☐ Stage IA
- ☐ Stage IB
- ☐ Stage IIA
- ☐ Stage IIB
- ☐ Stage IIIA
- ☐ Stage IIIB
- ☐ Stage IIIC
- ☐ Stage IV
- ☐ Unknown / Not Documented

**Q3: According to the medical chart, what is the status/result of the patient's ESTROGEN RECEPTOR test? (Please choose ONLY ONE of the following option)**

- ☐ Positive
- ☐ Negative
- ☐ Conflicting results
- ☐ Unknown/Test was not done/Results were not documented

**Q4: According to the medical chart, what is the status/result of the patient's PROGESTERONE RECEPTOR test? (Please choose ONLY ONE of the following option)**

- ☐ Positive
- ☐ Negative
- ☐ Conflicting results
- ☐ Unknown/Test was not done/Results were not documented

**Q5: According to the medical chart, what is the status/result of the patient's HER2 test? (Please choose ONLY ONE of the following option)**

- ☐ Positive
- ☐ Negative
- ☐ Equivocal
- ☐ Conflicting results
- ☐ Unknown/Test was not done/Results were not documented

**Q6: According to the medical chart, what is the patient's MENOPAUSAL status? (Please choose ONLY ONE of the following option)**

- ☐ Pre-menopausal
- ☐ Post-menopausal
- ☐ Peri-menopausal
- ☐ Conflicting results
- ☐ Unknown/ Status was not documented

### **SECTION 3: LUNG CANCER RELATED QUESTIONS**

***Note: Please only fill out the questions in the section if “Lung Cancer” in Section 1, Q6 is selected. Otherwise, please skip this section.***

**Q1: According to the medical chart, what is the most recent stage (based on TNM staging) for the diagnosed lung cancer? (Please choose ONLY ONE of the following option)**

- ☐ Stage 0
- ☐ Stage I
- ☐ Stage II
- ☐ Stage III
- ☐ Stage IV
- ☐ Limited
- ☐ Unknown / Not Documented

**Q2: According to the medical chart, what is the patient's most recent lung cancer stage (based on TNM staging) at a more granular level? (Please choose ONLY ONE of the following option)**

- ☐ Stage 0
- ☐ Stage IA
- ☐ Stage IB
- ☐ Stage IIA
- ☐ Stage IIB
- ☐ Stage IIIA
- ☐ Stage IIIB
- ☐ Stage IV
- ☐ Limited
- ☐ Unknown / Not Documented

**Q3: According to the medical chart, what is the patient's ALK mutation status? (Please choose ONLY ONE of the following option)**

- ☐ Positive
- ☐ Negative
- ☐ Conflicting results
- ☐ Unknown/Test was not done/Results were not documented

**Q4: According to the medical chart, what is the patient's EGFR mutation status? (Please choose ONLY ONE of the following option)**

- ☐ Wild type
- ☐ Mutation
- ☐ Conflicting results
- ☐ Unknown/Genotype was not documented

**Q5: According to the medical chart, what is the patient's histology classification for lung cancer? (Please choose ONLY ONE of the following option)**

- ☐ Small cell lung cancer
- ☐ Non-small cell lung cancer
- ☐ Conflicting results
- ☐ Unknown/Histology was not documented

## SECTION 4: COLORECTAL CANCER RELATED QUESTIONS

**Note: Please only fill out the questions in the section if “Colorectal Cancer” in Section 1, Q6 is selected. Otherwise, please skip this section.**

**Q1: According to the medical chart, what is the most recent stage (based on TNM staging) for the diagnosed colorectal cancer? (Please choose ONLY ONE of the following option)**

- ☐ Stage 0
- ☐ Stage I
- ☐ Stage II
- ☐ Stage III
- ☐ Stage IV
- ☐ Unknown / Not Documented

**Q2: According to the medical chart, what is the patient's most recent colorectal cancer stage (based on TNM staging) at a more granular level? (Please choose ONLY ONE of the following option)**

- ☐ Stage 0
- ☐ Stage I
- ☐ Stage IIA
- ☐ Stage IIB
- ☐ Stage IIC
- ☐ Stage IIIA
- ☐ Stage IIIB
- ☐ Stage IIIC
- ☐ Stage IV
- ☐ Stage IVA
- ☐ Stage IVB
- ☐ Unknown / Not Documented

**Q3: According to the medical chart, what is the patient's KRAS genotype? (Please choose ONLY ONE of the following option)**

- ☐ Wild type
- ☐ Mutation
- ☐ Conflicting results
- ☐ Unknown/Genotype was not documented



**Supplementary Data: Cross tabulation of all cancer stage and biomarker status data according to the medical record and the HIRE-Oncology (HIRE-O)**

**A. Breast Cancer**

Cancer Stage		Medical Record						Total
		0	I	II	III	IV	Missing	
HIRE-O	0	0	0	1	0	0	0	1
	I	0	45	3	0	0	3	51
	II	0	6	90	3	1	7	107
	III	0	4	10	25	2	2	43
	IV	0	4	6	3	81	4	98
Total		0	59	110	31	84	16	300

Estrogen Receptor		Medical Record			Total
		Positive	Negative	Missing	
HIRE-O	Positive	193	3	10	206
	Negative	5	77	4	86
	Missing	2	5	1	8
Total		200	85	15	300

Progesterone Receptor		Medical Record			Total
		Positive	Negative	Missing	
HIRE-O	Positive	137	14	13	164
	Negative	6	104	9	119
	Missing	7	9	1	17
Total		150	127	23	300

HER2		Medical Record				Total
		Positive	Negative	Equivocal	Missing	
HIRE-O	Positive	90	5	1	8	104
	Negative	4	170	3	9	186
	Equivocal	3	6	0	1	10
	Missing	0	0	0	0	0
Total		97	181	4	18	300

Highlighted areas in all tables indicates non-missing data for which the validation analysis was conducted

## B. Lung Cancer

Cancer Stage		Medical record							Total
		0	I	II	III	IV	Limited	Missing	
HIRE-O	0	0	0	0	0	1	0	0	1
	I	0	0	0	0	0	0	1	1
	II	0	0	3	0	1	0	0	4
	III	0	0	0	17	2	2	0	21
	IV	0	0	1	10	152	2	2	167
	Limited	0	0	0	0	0	3	0	3
Total		0	0	4	27	156	7	3	197

ALK Mutation		Medical Record			Total
		Positive	Negative	Missing	
HIRE-O	Positive	7	6	4	17
	Negative	0	61	12	73
	Missing	0	25	82	107
Total		7	92	98	197

EGFR Mutation		Medical Chart (Reference)			Total
		Mutation	Wild Type	Missing	
HIRE-O	Mutation	13	2	6	21
	Wild Type	0	54	9	63
	Missing	3	35	75	113
Total		16	91	90	197

Pathology		Medical Chart (Reference)			Total
		SC	NSC	Missing	
HIRE-O	Small cell	22	2	0	24
	Non-small cell	4	146	4	154
	Missing	3	15	1	19
Total		29	163	5	197

Highlighted areas in all tables indicates non-missing data for which the validation analysis was conducted

**C. Colorectal  
Cancer**

Cancer Stage		Medical Record						Total
		0	I	II	III	IV	Missing	
HIRE-O	0	0	0	0	0	2	0	2
	I	0	0	0	1	0	0	1
	II	0	0	4	1	0	0	5
	III	0	0	0	19	3	0	22
	IV	0	0	0	5	160	2	167
Total		0	0	4	26	165	2	197

KRAS mutation		Medical Record			Total
		Wild Type	Mutation	Missing	
HIRE-O	Wild Type	71	3	15	89
	Mutation	3	37	5	45
	Missing	6	8	49	63
Total		80	48	69	197

Highlighted areas in all tables indicates non-missing data for which the validation analysis was conducted