REGISTRAR PIP

Visit SEER*Educate: A comprehensive training platform for registry professionals

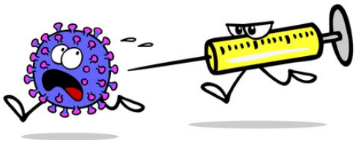
## September 2024 Registrar PIP
## A Visual Sharing Circle

### Introduction

In July, I reported on some of the more memorable posters seen hanging in the Exhibit Hall during the National Cancer Registrars Association's (NCRA) annual meeting. After attending the North American Association of Central Cancer Registries' (NAACCR) annual meeting held this year in Idaho, I decided to give equal time to the posters on display in Boise for this edition of the Registrar PIP.

Before highlighting the topic most likely to be of interest to registrars and our standard setters, I wanted to mention that I was struck by the number of COVID-19 related research posters pinned to the conference center's cork-covered boards. While the worst of the clinically-related impact of COVID-19 is behind us, investigators are still grappling with how best to interpret that disease's impact on cancer counts and outcomes.

### COVID-19 . . . It's back!



Clearly, COVID-19 had an impact on 2020 cancer incidence reporting when compared to prior years' case counts and age-adjusted rates. According to the poster by Abby Holt, Program Manager at Intermediate Care Facilities (ICF) in Little Rock, Arkansas, much of the decrease in reporting occurred during March through May of 2020 when the public health emergency was announced. For some of the major cancers, incidence rates and case counts reported still had not returned to pre-pandemic levels by the end of 2021. She reported that while female breast cases diagnosed at localized, regional, and distant stages returned to pre-pandemic levels, the following site and stage groups did not:

- Early staged prostate and colorectal cancers
- Early to late staged lung and bronchus cancers
- Unknown stage for breast, prostate, colorectal, lung, and bronchus

Cancer projections are used in not only time trend analyses, but they are used in planning future capital expenditures, patient care spending, and funding for implementing cancer risk reduction strategies. There is a lot of money at stake. Therefore, investigators believe it is worth taking a little time to consider whether it is necessary to modify projections when a pandemic disrupts things and how best to do so.

Given the critical uses of projected counts, I was intrigued by Oliver Bucher's, Epidemiologist with CancerCare Manitoba in Winnipeg, COVID-19 poster presentation, especially the description of the first two of three methods he used to project counts and age-standardized incidence rates (ASIRs) for the 1988-2045 timeline used in Manitoba, Canada:

- Replace the 2020 and 2021 counts/ASIRs with the previously expected counts/ASIRs (In other words, let the estimates stand as previously calculated.)
- Exclude the 2020 and 2021 counts/ASIRs (After all, the National Cancer Institute (NCI) proposed the exclusion of the 2020 incidence from trend estimates. At some later point, investigators could determine if that data should continue to be excluded or reincorporated into analyses.)

CSS is funded by the National Cancer Institute's SEER Program, Contract Number HHSN261201800004I

- Redo projections without any adjustment for the COVID-19 period (Reflect reality as we know it for diagnosis years 1988 through 2021and then recalculate future expected counts/ASIRs through 2045.)

Bucher concluded that for Manitoba any ***adjusting*** for the pandemic years in the projection models resulted in some unrealistically high projected counts/ASIRs, whereas their ***unadjusted*** incorporation did not. Researchers believe it is unlikely that they will soon observe a return to the trends at the diagnosis levels seen pre-pandemic. Ultimately, in Manitoba the decision was made to use the COVID-19 unadjusted counts/ASIRs in their projection models because they appear to fit well.

### Linkages . . . movement toward a more comprehensive registry data infrastructure

The decision to highlight the poster seen in Figure 1 for this edition seemed like a no-brainer. Believe it or not, I was not influenced by the fact that the poster was created by Hutch staff or that some of the regional data were used in this analysis; though, admittedly, I'm always curious about how the data we've spent years collecting is being used. The primary reason for selecting the poster below was to highlight the Surveillance, Epidemiology and End Results (SEER) Program's interest in the effort of population-based central registries to find cost effective ways to capture data on cancer recurrence and metastasis.

It is unrealistic for standard setters to insist on multiple manual reviews of every patient's medical record to check on whether and when each patient developed disease progression or metastasis, and the sites of involvement when it does occur. The collection of recurrence and metastatic events would add to the already time-consuming number of manually abstracted data items. However, researchers, clinicians, and, most importantly, cancer patients want to understand more about the disease outcomes than we currently collect to help answer the questions and address the concerns they have. They need more. The rub is that realistically, our human and financial resources limit what we can do manually. The option to expand the dataset using linkages to other sources offers opportunities to audit, validate and enrich the data currently available in the registry.

Abstracts provide the supporting text for the initially collected data items that have proven critical in past and current cancer surveillance efforts, which are used to assess approaches to reduce the cancer burden. Surveillance data are also used to plan and evaluate cancer prevention and control interventions. While many can appreciate the usefulness of the existing data, they want a future that includes a more expanded potential use of registry data when it is linked with other sources.

In addition to information on disease progression and metastasis, registries typically lack detailed information on all available biomarkers, exposures, longitudinal treatment information, and reliable insurance data relevant to understanding the burden of cancer. Currently, linkages occurring between central registries with other data sources such as medical claims, hospital discharge data, and pathology reports, hope to fill in the missing pieces of information related to disease progression and metastasis. These same linkages can also be used to fill in the gaps observed in the treatment information recorded in registries. With a more enhanced dataset, research questions related to care patterns, costs, comorbidities, disparities, and late effects of treatment could be addressed.

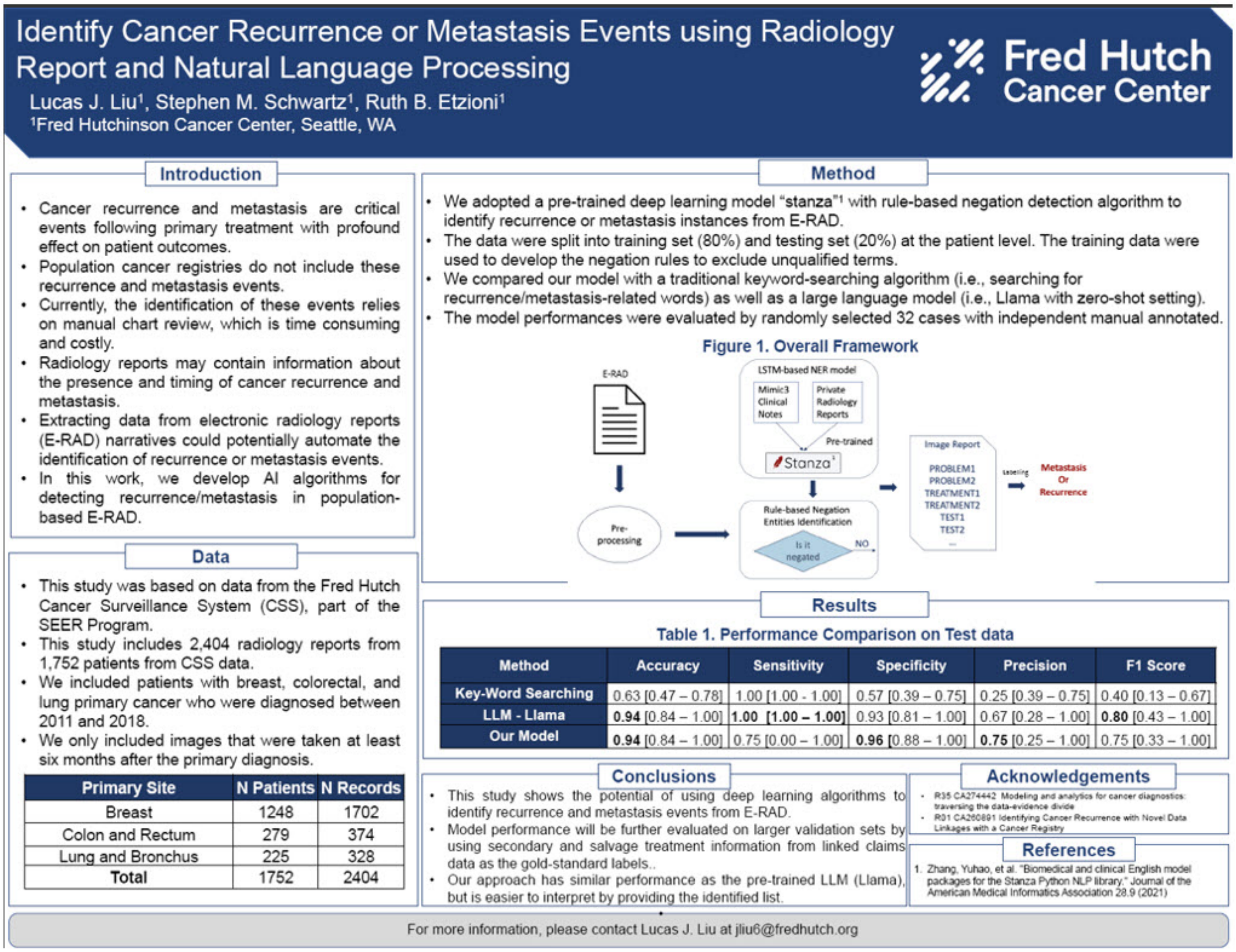
### Pathology may be everyone's first choice . . . but scans are far from being sloppy seconds!

Full text electronic pathology reports were initially used to streamline and improve the completeness of casefinding procedures; support comprehensive rapid case ascertainment for studies; and provide the central registry an opportunity to generate preliminary incidence data. Given the current availability of these reports in many central registries, investigators across the country began to evaluate whether these same reports could be used to accurately identify cancer recurrence in patients following their initial treatment.

Joan Warren at the National Cancer Institute lead a team of investigators that conducted a **retrospective study** to evaluate how effective ***pathology reports*** were in identifying recurrences. The study included patients with known **previously recurrent breast** (n=214) or **colorectal** (n=203) cancers who had been followed in depth for 5 years after. After linking subsequently identified pathology reports to these patients, her results indicated half of cancer patients had a pathology report near the time of recurrence. For patients with a pathology report, the report's sensitivity in identifying recurrence was 98.1% for breast cancer cases and 95.7% for colorectal cancer cases. However, she concluded that electronic submission of ***pathology reports alone couldn't measure population-based recurrence for all solid cancers***, but they could identify specific cohorts of recurrent cancer patients in near real-time.

**Figure 1**

**Fred Hutchison Cancer Center**
**Annual Meeting Poster Presentation**
**North American Association of Central Cancer Registries**
**June 2024**



## Identify Cancer Recurrence or Metastasis Events using Radiology Report and Natural Language Processing

Lucas J. Liu[1], Stephen M. Schwartz[1], Ruth B. Etzioni[1]
[1]Fred Hutchinson Cancer Center, Seattle, WA

Fred Hutch Cancer Center

### Introduction

- Cancer recurrence and metastasis are critical events following primary treatment with profound effect on patient outcomes.
- Population cancer registries do not include these recurrence and metastasis events.
- Currently, the identification of these events relies on manual chart review, which is time consuming and costly.
- Radiology reports may contain information about the presence and timing of cancer recurrence and metastasis.
- Extracting data from electronic radiology reports (E-RAD) narratives could potentially automate the identification of recurrence or metastasis events.
- In this work, we develop AI algorithms for detecting recurrence/metastasis in population-based E-RAD.

### Data

- This study was based on data from the Fred Hutch Cancer Surveillance System (CSS), part of the SEER Program.
- This study includes 2,404 radiology reports from 1,752 patients from CSS data.
- We included patients with breast, colorectal, and lung primary cancer who were diagnosed between 2011 and 2018.
- We only included images that were taken at least six months after the primary diagnosis.

| Primary Site | N Patients | N Records |
|---|---|---|
| Breast | 1248 | 1702 |
| Colon and Rectum | 279 | 374 |
| Lung and Bronchus | 225 | 328 |
| **Total** | 1752 | 2404 |

### Method

- We adopted a pre-trained deep learning model "stanza"[1] with rule-based negation detection algorithm to identify recurrence or metastasis instances from E-RAD.
- The data were split into training set (80%) and testing set (20%) at the patient level. The training data were used to develop the negation rules to exclude unqualified terms.
- We compared our model with a traditional keyword-searching algorithm (i.e., searching for recurrence/metastasis-related words) as well as a large language model (i.e., Llama with zero-shot setting).
- The model performances were evaluated by randomly selected 32 cases with independent manual annotated.

Figure 1. Overall Framework

### Results

Table 1. Performance Comparison on Test data

| Method | Accuracy | Sensitivity | Specificity | Precision | F1 Score |
|---|---|---|---|---|---|
| Key-Word Searching | 0.63 [0.47 – 0.78] | 1.00 [1.00 - 1.00] | 0.57 [0.39 – 0.75] | 0.25 [0.39 – 0.75] | 0.40 [0.13 – 0.67] |
| LLM - Llama | **0.94** [0.84 – 1.00] | **1.00 [1.00 – 1.00]** | 0.93 [0.81 – 1.00] | 0.67 [0.28 – 1.00] | **0.80** [0.43 – 1.00] |
| Our Model | **0.94** [0.84 – 1.00] | 0.75 [0.00 – 1.00] | **0.96** [0.88 – 1.00] | **0.75** [0.25 – 1.00] | 0.75 [0.33 – 1.00] |

### Conclusions

- This study shows the potential of using deep learning algorithms to identify recurrence and metastasis events from E-RAD.
- Model performance will be further evaluated on larger validation sets by using secondary and salvage treatment information from linked claims data as the gold-standard labels..
- Our approach has similar performance as the pre-trained LLM (Llama), but is easier to interpret by providing the identified list.

### Acknowledgements

- R35 CA274442 Modeling and analytics for cancer diagnostics: traversing the data-evidence divide
- R01 CA260891 Identifying Cancer Recurrence with Novel Data Linkages with a Cancer Registry

### References

1. Zhang, Yuhao, et al. "Biomedical and clinical English model packages for the Stanza Python NLP library." Journal of the American Medical Informatics Association 28.9 (2021)

For more information, please contact Lucas J. Liu at jliu6@fredhutch.org

To improve the capture of recurrence data, some central registries are more recently turning their attention to radiology reports. Figure 1 indicates the ***retrospective study*** observed results by Hutch investigators who evaluated how effective different methods were in identifying recurrence in electronically received ***radiology reports*** linked to the CSS dataset. The study included 1248 breast patients with 1,702 radiology reports, 279 ***colorectal*** patients with 374 radiology reports, and 225 ***lung/bronchus*** patients with 328 radiology reports. They compared the effectiveness of the following models used to try to classify the radiology reports:

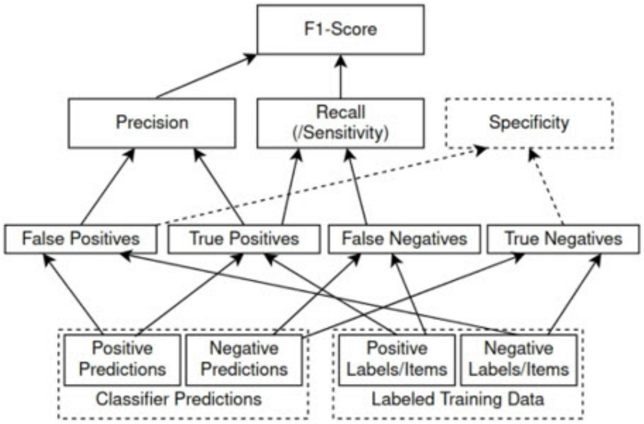
- **Key-word searching** – searches for keywords in a report that have been specified by the user to indicate a recurrence

- **LLM** - Llama (large language model meta artificial intelligence) – uses machine learning, a form of artificial intelligence involving the use of natural language processing (NLP), to train its model to take a sequence of words as input in a report to determine whether the report indicates a recurrence
- **Stanza** – uses deep learning, a form of artificial intelligence involving machine learning, to instruct a computer to learn through observation and repetition how to classify whether the report indicates a recurrence

It is important to understand the terms used to validate the different classification models in order to determine whether an automated method could successfully be used to help us collect recurrence information from scans. Which method has the greatest potential? Most classifications metrics evaluate the following:

- **Accuracy**: ability to measure what it's supposed to, and how well it identifies and excludes a condition
- **Sensitivity:** ability to designate an individual with disease as positive
- **Specificity:** ability to designate an individual who does not have a disease as negative
- **Precision:** ability of a method to reproduce its measurements

**Figure 2**



Teemu Kanstrén, a technology and software engineer at the Qt Company in Oulu, Finland, created the graphic shown in Figure 2 that outlines the elements involved in creating F1 Score, another metric used to evaluate the accuracy of a machine learning model. It provides a more comprehensive evaluation than the accuracy measure by taking into account both false positive and false negative results. According to Danish Hasarat, a DevsOps engineer at IBM, "In scenarios where false positives and false negatives have an implication in medical diagnoses, the F1-score may be preferred over accuracy."

In the Hutch investigators study results, LLM and Stanza models have the same accuracy rates and both outperform key-word searching models. The LLM model's F1 score is slightly better than the Stanza model. Their future plans include evaluations on larger validation sets using "secondary and salvage treatment information from linked claims data."

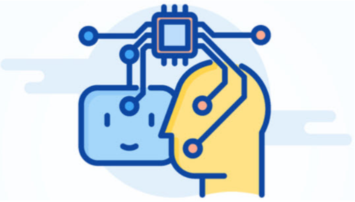
The initial study results suggest we could potentially use electronic methods to improve data collection methods which will:

- Enhance registry operation efficiency
- Expand the registry dataset
- Improve the accuracy and completeness of the data
- Permit data analysis to occur more quickly
- Allow stakeholders access to data necessary to address research issues and the concerns of cancer survivors post first course of treatment

## Conclusion

Population-based cancer registry data related to progression and metastatic patterns is seldom recorded and available because it is cost prohibitive to collect manually. With increasing survivorship, there is a growing demand to understand outcomes other than death that reflect the post-diagnosis course of the disease for cancer patients. Understanding cancer progression and metastasis is important for clinicians and patients because it can help determine future treatment, prognosis, and clinical management for this disease. According to Stacey Tinianov, Vice President of Patient Advocacy and Engagement at Rabble Health, "Without metastatic recurrence data, the registry is neither complete nor reflective of the patient population."

CSS is funded by the National Cancer Institute's SEER Program, Contract Number HHSN261201800004I



Realistically, registries will need to incorporate and combine automated extraction of data from electronic medical records and the use of artificial intelligence, machine learning, and deep learning to increase data collection efficiency and enhance the research and clinical value of registry data. Extracting additional data items from pathology and radiology reports, structured clinical reporting systems, and administrative hospital data is possible and should be part of the future data collection gold standard because it has the potential to affect and personalize cancer patient care management from screening, diagnosis, and treatment to survivorship.

CSS is funded by the National Cancer Institute's SEER Program, Contract Number HHSN261201800004I