

Validation of a Natural Language Processing solution (REALLI) compared to manual review by physicians of electronic health records (EHR) from patients with HER2⁺ metastatic breast cancer

AUTHORS

A. Groenez¹, B. Lebas², G. Rejasse², C. Roux², P.A. Squara¹, M Jouve²

¹Pfizer, Paris, France

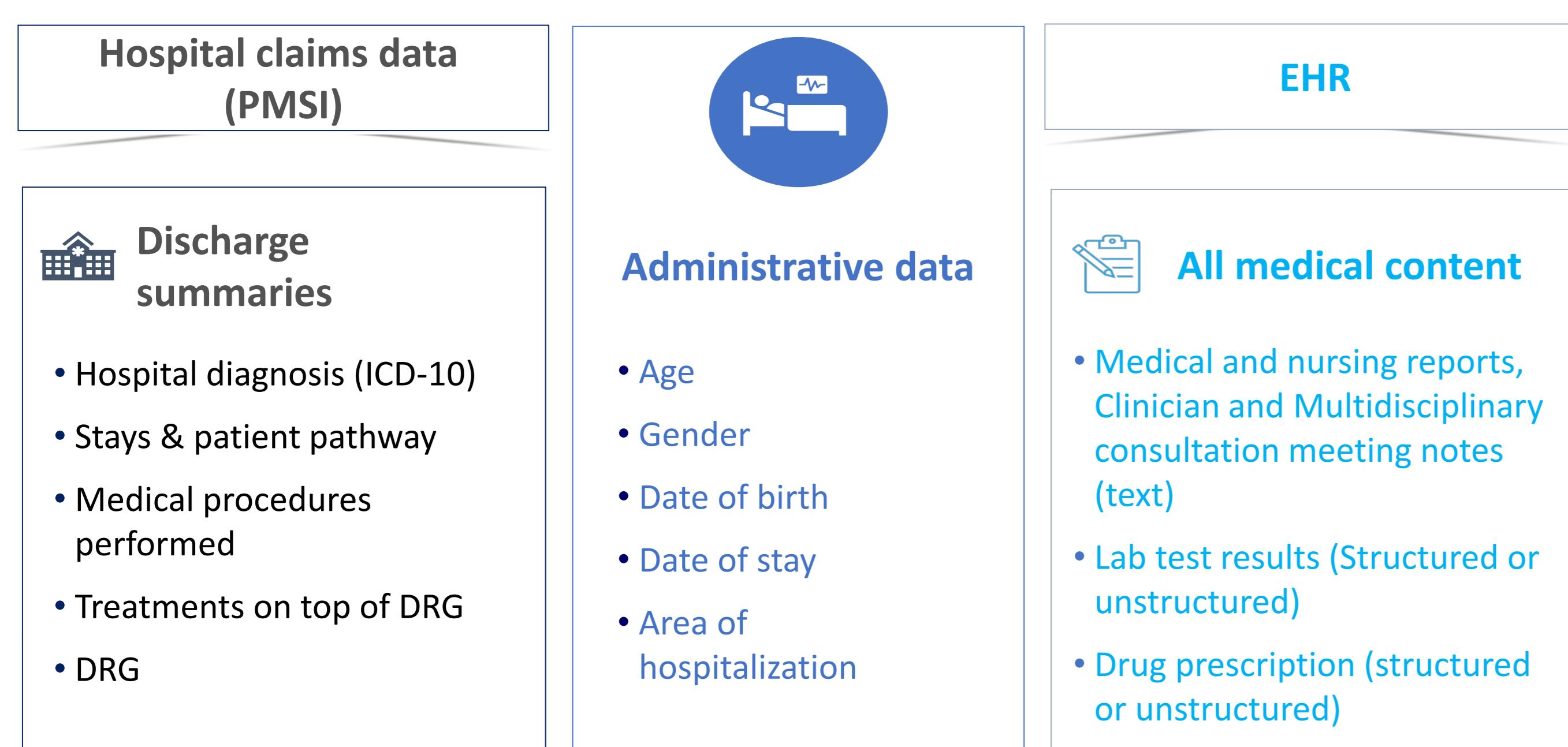
²Sancare, Paris, France

INTRODUCTION AND OBJECTIVES

Hospital electronic health records (EHR) contain valuable real-world data, but their unstructured free-text format limits utilization for real-world evidence generation. Natural Language Processing (NLP) methods, such as REALLI solution, offer the potential to efficiently identify specific patient populations without the need for time-consuming and costly manual review by physicians.

METHODOLOGY

Data Source : Data were collected from EHRs in 3 general hospitals and 1 university hospital in France.



Study Period : Between January 1st, 2017, and December 31st, 2021.

Study Population : Adult women hospitalized for breast Cancer identified through ICD10 diagnosis code

The aim of the study was to assess the performance of an NLP-powered solution (REALLI) compared to gold standard (manual review) for the detection of the metastatic status, metastatic diagnostic date and HER2+ status in EHRs of patients with Breast Cancer.

Analysis : Out of 3,214 BC patients identified within the study period, a random sample of 200 patients was selected for comparison between REALLI and manual review by physicians. REALLI used a combination of algorithm based on ICD 10 codes or ATC, UCD codes and NLP methods from semantic rules-based research and Large Language Model (LLM) to identify metastatic and HER2+ status from hospital EHRs.



Performance metrics included sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The concordance of metastatic diagnosis dates between REALLI and manual review was also assessed.

RESULTS

> METASTATIC BREAST CANCER PATIENT DETECTION

Among the random sample of 200 BC patients, 67 metastatic patients have been identified by the NLP solution REALLI and 60 by manual review (Figure 1).

	Manual review			Total
	Metastatic status positive	Metastatic status negative		
REALLI	60	7	69	
Metastatic status negative	0	133	133	
Total	60	140	200	

Figure 1: Contingency matrix for metastatic status of breast cancer patient

> DIAGNOSIS DATES OF METASTASES

Among the 60 metastatic patients identified by the manual review, a concordance of 90% (54/60) on the metastatic diagnosis dates was observed between REALLI and the manual review.

> METASTATIC BREAST CANCER PATIENTS WITH HER2⁺ STATUS

HER2 status detection have been assessed on 43 metastatic BC patient (1 center, representing 17 metastatic patients, was excluded because patients were treated in the nearby comprehensive cancer center).

	Manual review			Total
	HER2 status positive	HER2 status negative		
REALLI	9	0		9
HER2 status negative	0	34		34
Total	9	34		43

Figure 2: Contingency matrix for HER2 status of metastatic breast cancer patients

Among the 43 metastatic breast cancer patients identified, 9 were found to be HER2+ by the NLP solution and 9 by manual review (Figure 2).

Tableau 1 : Performances of REALLI to identify metastatic breast cancer and HER2+ status of metastatic BC patients

		Results	CI95	
Metastatic Status	Sensitivity	1.0	0.94	1.0
	Specificity	0.95	0.90	0.98
	PPV	0.90	0.80	0.96
	NPV	1.0	0.97	1.0
HER2+ Status	Sensitivity	1.0	0.66	1.0
	Specificity	1.0	0.89	1.0
	PPV	1.0	0.66	1.0
	NPV	1.0	0.89	1.0

PPV: Positive predictive value; NPV: Negative Predictive Value

CONCLUSION

The REALLI NLP solution is a time-saving, cost-effective, and reliable alternative to manual review for extracting key clinical information from unstructured hospital EHRs. Its high performance in identifying metastatic and HER2+ status, as well as diagnosis dates, supports its use in real-world oncology research. However, multicenter validation is required to account for differences in coding and reporting practices. This approach has the potential to enhance the utilization of unstructured EHR data for real-world evidence generation in oncology.



DOWNLOAD THE
DIGITAL VERSION

Scan this QR code to download a copy of the poster. Please note that reproducing copies of this poster and additional content via the Quick Response (QR) code is prohibited without permission from both ISPOR and the authors.

www.sancare.fr

