

Enriching the TNBC Patient Journey with Clinical Note Abstraction

Challenge

EHR data provides a longitudinal overview of oncology care, but clinically important details such as treatment intent, biomarker testing, surgical outcomes, and pathologic response are often documented primarily in unstructured clinical notes. OMNY Health and Savant evaluated how targeted abstraction could build on OMNY's structured oncology data to create a more detailed view of the TNBC patient journey.

OMNY Health: The Foundational Dataset

OMNY Health has provider-led, longitudinal real-world data captured directly from the point of care, providing rich clinical insights across encounters, patient demographics, diagnoses, procedures, prescriptions, problem lists, clinical assessment scores, and social determinants of health. OMNY's network represents more than 175 million patient lives, including more than 10 million oncology patient journeys from 2017 to the present. For this analysis, OMNY provided both de-identified structured EHR data and unstructured clinical notes from a sample cohort of 1,000 patients with TNBC. Unstructured clinical notes were de-identified using 'hiding in plain sight' methodology and the overall dataset was certified as HIPAA compliant using expert determination methodology.

Savant Bio: Targeted Abstraction Layer

Savant Bio helps healthcare and life sciences organizations unlock the value of unstructured healthcare data by converting it into structured, analysis-ready datasets. For this analysis, Savant provided the clinical data curation layer to extract and standardize key clinical details from OMNY's unstructured clinical notes, complementing the information available in the structured EHR data. Savant abstracted treatment intent, menopausal status, BRCA testing, surgery details, pathologic response, and radiotherapy to identify patients who received neoadjuvant treatment and reconstruct their longitudinal journeys with greater clinical context.

Key Results

Savant abstracted approximately 385,000 clinical notes across a 1,000-patient TNBC cohort. OMNY's structured oncology EHR data provided broad longitudinal coverage across biomarkers, systemic treatment, surgery, staging, and more. Savant's targeted abstraction of unstructured clinical notes built on this foundation and substantially increased the availability of key clinical details, including treatment intent from 51.2% to 87.7%,

BRCA results from 17.3% to 65.5%, and surgical laterality from 42.4% to 92.4%. Abstraction also added clinical context not represented as discrete fields in the structured data, including margin status, pathologic response, and radiotherapy site and intent, enabling more detailed longitudinal analysis of the TNBC patient journey.

TNBC Dataset Overview

Overview of the 1,000-patient TNBC cohort and the clinical information captured across OMNY's structured EHR data and unstructured clinical notes.

CLINICAL CONCEPT	FOUNDATIONAL STRUCTURED DATA	FOLLOWING NOTE ABSTRACTION
TREATMENT & STATUS		
Treatment Intent Neoadjuvant, Adjuvant, Curative, Palliative, Maintenance	512 patients (51.2%)	877 patients (87.7%)
Menopausal Status	394 patients (39.4%) with menopause-related symptom codes across diagnoses & problem lists (no pre-menopausal status code in the code set)	941 patients (94.1%) with documented menopausal status
BIOMARKERS		
Biomarkers: ER, PR, HER2	100% of the cohort has ER, PR, HER2 testing and results confirming the 1,000 patient TNBC cohort	Fully captured in structured data; no additional abstraction required
Biomarkers: BRCA	173 patients (17.3%) with a BRCA result across biomarker rows and diagnosis codes	655 patients (65.5%) with documented BRCA result
SURGERY		
Breast or nodal surgery	626 patients (62.6%)	927 patients (92.7%)
Surgery laterality	424 patients (42.4%)	924 patients (92.4%)
Margin Status	Not represented as a discrete field	628 patients (62.8%)
RADIOTHERAPY		
Radiotherapy	530 patients (53.0%)	770 patients (77.0%)
Radiotherapy site	Not represented as a discrete field	717 patients (71.7%)
Radiotherapy intent	Not represented as a discrete field	672 patients (67.2%)
PATHOLOGIC RESPONSE		
Explicit pathologic response	Not represented as a discrete field	pCR or residual-disease status for 440 patients (44.0%)

Sample Patient Journey

OMNY STRUCTURED EHR FOUNDATION

WITH SAVANT ABSTRACTION

 Clinical characteristics Triple-negative, infiltrating ductal carcinoma with clinical stage cT2 cN0 cM0	Core disease characteristics were comprehensively captured in structured data, with no additional abstraction required
 BRCA testing Structured biomarkers established the patient's molecular profile	Germline testing was documented in November 2023, with negative BRCA1 and BRCA2 results and test origin captured
 Neoadjuvant treatment Administration records documented receipt of paclitaxel, carboplatin, and cyclophosphamide	Clinical notes established the neoadjuvant intent and full KEYNOTE-522 treatment sequence: carboplatin, paclitaxel, and pembrolizumab beginning April 24, 2023, followed by doxorubicin, cyclophosphamide, and pembrolizumab beginning July 26, 2023
 Surgery Procedure records identified sentinel lymph node mapping on November 20, 2023	Notes confirmed a right lumpectomy and right sentinel lymph node biopsy on November 20, 2023, with negative surgical margins
 Pathologic response Structured staging supported postoperative disease characterization	Pathologic complete response was confirmed: ypT0 ypN0, no residual tumor, and no positive lymph nodes among one examined
 Radiotherapy Radiation oncology records confirmed that radiotherapy was delivered	Notes identified adjuvant breast radiotherapy from January 23 through February 20, 2024, beginning 64 days after surgery

Impact and Conclusion

These findings demonstrate the complementary value of the structured EHR data and abstraction of unstructured clinical notes within the OMNY dataset. The structured oncology data provided the scale and longitudinal foundation to define the TNBC cohort and capture key clinical events across treatment, surgery, staging, and radiotherapy. Savant Bio's abstraction added the clinical context needed to clarify treatment intent, expand key variables, and connect these events across the patient journey.

CONTACT OMNY HEALTH AT info@omnyhealth.com

CONTACT SAVANT BIO AT hello@savantbio.com