CO201

Real-World Treatment Patterns and Clinical Outcomes in Previously Untreated Locally Advanced or Metastatic Triple-Negative Breast Cancer, by PD-L1 Status: A NorstellaLinQ US Linked Claims–Electronic Health Records (EHR) Study

Becky Hollenberg, MPH[1]; Allison Perry, PhD, MHS[1]; Rahul K. Das, PhD[1]; Karan Bhanot, PhD[1]; Ilan Behm, MPH[1]
[1]Norstella, One Pennsylvania Plaza, New York, NY 10119 USA

Background

Triple-negative breast cancer (TNBC) lacks estrogen receptor and progesterone receptor expression and HER2 overexpression or amplification, and has historically had few targeted options. For previously untreated locally advanced inoperable or metastatic (la/m) TNBC, first-line therapy is guided by PD-L1 status: chemotherapy plus pembrolizumab is an option for PD-L1–positive disease (CPS ≥10),[1] while chemotherapy alone remains standard for PD-L1–negative disease. Antibody-drug conjugates (ADCs), sacituzumab govitecan[2] and, for HER2-low disease, trastuzumab deruxtecan,[3] are approved after prior therapy.

PD-L1 results, metastatic sites, and performance status are inconsistently captured in structured fields. NorstellaLinQ links EHR, open claims, and closed claims with structured laboratory data and clinical notes, enabling biomarker-stratified real-world treatment patterns and outcomes to be described.

Objective

To characterize treatment patterns and real-world clinical outcomes among adults with previously untreated la/m TNBC in the United States, stratified by PD-L1 status.

Methods

Data Sources
NorstellaLinQ US real-world linked data: structured EHR, unstructured clinical notes, open claims, closed claims, and mortality data linked at the patient level. Data window: January 2018 to December 2024.

Study Design
Retrospective cohort study. Index date: initiation of first systemic therapy for la/m TNBC (start of line 1), 2020 to 2023. Pre-index baseline: variable lookback from first EHR activity (data from January 2018 allow up to 24 months for the earliest index dates). Follow-up: index to death, loss to follow-up, or data end (December 31, 2024); at least 1 day required, censored at last observed activity.

Population
- Source population: 4,437,456 patients with breast cancer (ICD-10-CM C50.x in structured EHR, open claims, or closed claims).
- Inclusion: age ≥18 years at index; triple-negative status (ER-, PR-, HER2- per ASCO/CAP, from at least 1 EHR laboratory result or clinical note per receptor, or a direct TNBC confirmation in notes); locally advanced inoperable or metastatic disease (stage 3b/3c/4 from LLM-curated notes, or at least 2 secondary malignancy codes [C78.x, C79.x] more than 14 days apart); initiation of systemic therapy for la/m disease with no prior systemic therapy in that setting ((neo)adjuvant therapy not assigned a line; first treatment within 90 days of metastatic diagnosis) Final cohort: **724 patients**.

PD-L1 Ascertainment
- PD-L1 status derived from (1) structured laboratory results and (2) LLM-assisted abstraction of clinical notes with human in the loop review, using results documented before index (closest to index if multiple).
- Positive: CPS ≥10 or IC ≥1%, with no baseline autoimmune disorder; or unknown PD-L1 with immunotherapy in the la/m setting. Negative: CPS <10 or IC <1%. Unknown/undocumented: no interpretable result. Source hierarchy when results conflicted: closest to index.

Line of Therapy
Lines of therapy (LOT; up to 3) constructed from claims and EHR administration records using NorstellaLinQ standardized rules adapted from published algorithms. LOT 1 starts at the first eligible drug on or after la/m diagnosis; other eligible drugs within 28 days (at least 1 day of overlap) join the regimen. A regimen change advances the line, as does an addition after day 28. Exceptions: same-class substitutions, targeted/IO drugs added within 28 days, CDK4/6 inhibitors added within 60 days, and dropping one drug while others continue. Gaps in therapy do not end a line. Regimens categorized as chemotherapy alone, chemotherapy plus immunotherapy, ADC-based, or other.

Outcomes & Statistics
rwTOT: start to end of each line (end = earliest of last administration or supply date, day before the next line starts, death, or end of follow-up). **rwTTNT:** start of a line to start of the next line or death, censored at last activity. **rwOS:** index to death from any cause, censored at last activity or data end.
Medians with 95% CIs estimated by Kaplan–Meier (Panalgo Instant Health Data analytics platform with integrated R). Baseline characteristics summarized descriptively overall and by PD-L1 status; no adjusted comparisons were performed.

LLM Extraction & Validation
Data were extracted from clinical notes by an open-weight large language model (LLM) with human-in-the-loop prompt development: prompts were refined through iterative review of sampled outputs against the source notes before application to the full corpus, and extracted values were harmonized to standardized categories. Extraction quality was assessed on 200 clinical notes by an artificial intelligence (AI) evaluation panel of three LLMs from different model families. Each model independently assessed each extracted value against the corresponding note text and the extraction specification. Because no single reviewer, whether human or AI, is error-free, the majority verdict of the panel served as the reference. In prior evaluations, this AI evaluation approach showed approximately 85–95% agreement with clinician reviewers. Precision, recall, and F1 score were calculated, with 95% confidence intervals (CIs) estimated by note-level bootstrap resampling.

Results

724
Adults with previously untreated la/m TNBC (index 2020 to 2023)

59% / 34%
Initiated second-line / third-line therapy

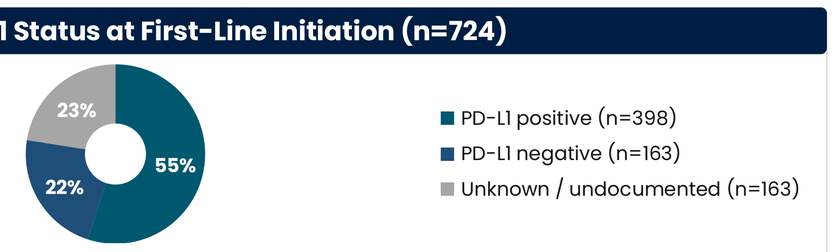
<3 months
Median rwTOT in every line (2.6, 2.8, 2.9 months)

Not reached
Median rwOS during available follow-up (108 deaths)

Patient Selection

4,437,456	Patients with breast cancer, Jan 2018 to Dec 2024
15,932	Triple-negative (ER-, PR-, HER2-)
4,203	Locally advanced inoperable or metastatic disease
1,838	Initiated systemic therapy 2020 to 2023, aged ≥18 years
724	Final cohort after follow-up and exclusion criteria

PD-L1 Status at First-Line Initiation (n=724)



Positivity definition: CPS ≥10 or IC ≥1% (EHR laboratory results or clinical notes); negative if CPS <10 or IC <1%
Unknown group: positive PD-L1 with autoimmune disease, or unknown PD-L1 without immunotherapy in the la/m setting (n=163, 22.5%).

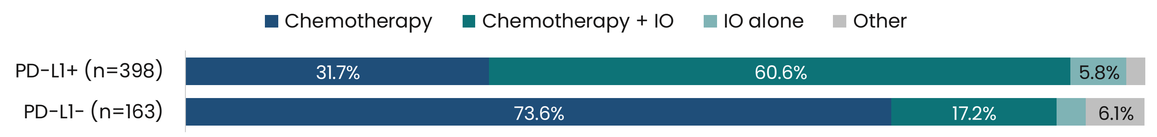
Baseline Characteristics

Characteristic	Overall (n=724)	PD-L1+ (n=398)	PD-L1- (n=163)
Female	721 (99.6%)	397 (99.7%)	162 (99.4%)
Age, mean (SD), years	59.30 (12.90)	58.07 (12.52)	59.76 (13.38)
ECOG 0–1	506 (69.9%)	282 (70.9%)	119 (73.0%)
Mets: lymph nodes	249 (34.4%)	137 (34.4%)	55 (33.7%)
Mets: bone	149 (20.6%)	68 (17.1%)	49 (30.1%)
Mets: lung	142 (19.6%)	63 (15.8%)	48 (29.4%)
Mets: liver	119 (16.4%)	58 (14.6%)	37 (22.7%)

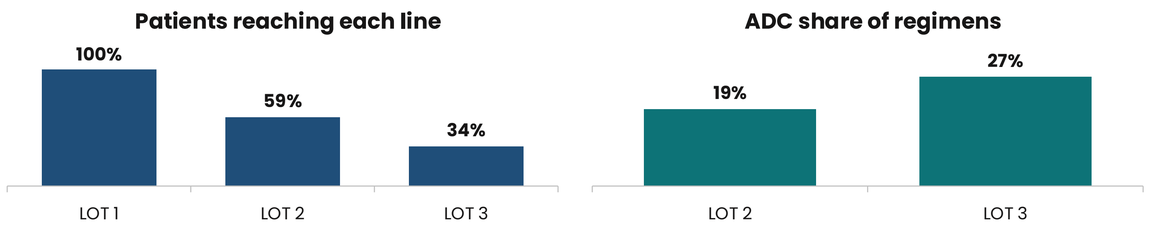
First-Line Regimen by PD-L1 Status

61%
of PD-L1+ received chemotherapy plus immunotherapy

74%
of PD-L1- received chemotherapy alone



Treatment Sequencing and ADC Use by Line



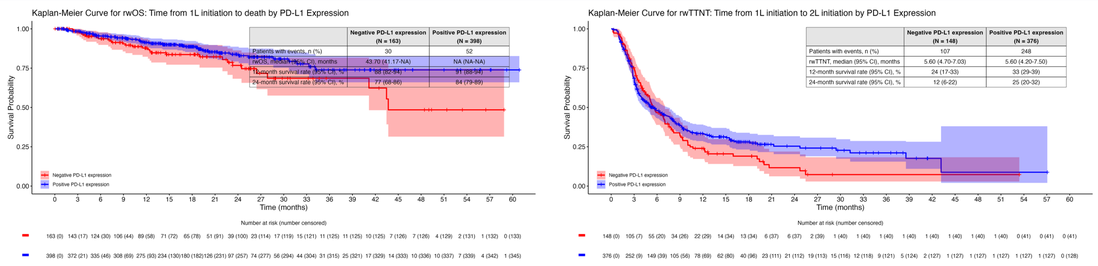
ADC share of regimens: PD-L1- 30.8% (LOT 2) and 31.0% (LOT 3); PD-L1+ 12.9% (LOT 2) and 25.2% (LOT 3).

Real-World Outcomes, Median (95% CI), Months

Outcome	Overall	PD-L1+	PD-L1-
rwTOT, LOT 1	2.63 (2.47-2.83)	2.77 (2.57-3.03)	2.80 (2.33-3.23)
rwTOT, LOT 2	2.83 (2.60-3.27)	3.03 (2.57-3.70)	2.73 (2.27-3.27)
rwTOT, LOT 3	2.93 (2.57-3.73)	3.40 (2.77-4.67)	2.27 (1.63-3.97)
rwTTNT, LOT 1 to 2	6.53 (5.47-7.80)	5.60 (4.20-7.50)	5.60 (4.70-7.03)
rwTTNT, LOT 2 to 3	7.20 (6.53-9.27)	7.43 (6.27-10.57)	6.30 (5.03-7.07)
rwOS, median	NR	NR	43.70 (41.17-NR)
rwOS, 12-month rate, %	90 (87-92)	91 (88-94)	88 (82-94)
rwOS, 24-month rate, %	81 (78-85)	84 (79-89)	77 (68-86)

NR, not reached. Medians (95% CI) in months by Kaplan–Meier; 108 deaths overall; median follow-up 15.67 months (IQR 7.62-24.01).

Kaplan–Meier Curves



Key Limitations

1. PD-L1 status was unknown or undocumented for 163 of 724 patients (22.5%); patients with documented results may differ from those without.
2. PD-L1 assays and scoring algorithms differ; documented positivity may not map to a single label-defined threshold.
3. rwTOT and rwTTNT are derived from claims and EHR administration records; incomplete capture of administrations can shorten estimated durations.
4. Death capture may be incomplete; median rwOS not reached reflects limited follow-up (index through 2023, data through 2024).
5. Analyses are descriptive; differences between PD-L1 groups are unadjusted and may reflect confounding. The cohort reflects contributing health systems.

LLM Extraction Quality vs. AI Evaluation Panel (200 clinical notes)

Relative to the panel reference, extraction achieved a precision, recall, and F1 score of 99.2% (95% CI, 97.4–100.0), 92.0% (87.4–96.3), and 95.5% (92.6–97.8) for grouped stage; 96.8% (94.6–98.5), 98.7% (98.0–99.4), and 97.8% (96.6–98.7) for biomarker value and unit; 95.0% (91.1–98.0), 98.5% (96.3–100.0), and 96.7% (94.3–98.9) for metastasis sites; and 94.2% (91.1–97.0), 95.0% (92.3–97.4), and 94.6% (92.0–96.8) for tumor subtype.

Table 1. Agreement (%, 95% CI) between LLM extraction and a three-model AI evaluation panel (majority vote).

Field	n (notes)	P	R	F1
Grouped stage	200	99.20 (97.4–100.0)	92.00 (87.4–96.3)	95.50 (92.6–97.8)
Biomarker value & unit	200	96.80 (94.6–98.5)	98.70 (98.0–99.4)	97.80 (96.6–98.7)
Metastasis sites	200	95.00 (91.1–98.0)	98.50 (96.3–100.0)	96.70 (94.3–98.9)
Tumor subtype	200	94.20 (91.1–97.0)	95.00 (92.3–97.4)	94.60 (92.0–96.8)

Conclusion

1. In real-world la/m TNBC, first-line treatment varied by PD-L1 status: 61% of PD-L1–positive patients received chemotherapy plus immunotherapy, whereas 74% of PD-L1–negative patients received chemotherapy alone.
2. Time on treatment was short across all lines (median rwTOT under 3 months), 59% and 34% of patients proceeded to second- and third-line therapy, and ADCs were used predominantly after first line (19% of second-line and 27% of third-line regimens), indicating substantial unmet need.
3. Linked claims and EHR data, supplemented with PD-L1 ascertainment from clinical notes, enabled characterization of biomarker-defined treatment patterns that are difficult to evaluate using structured data alone.

References
1. Cortes J, Cescon DW, Rugo HS, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355). Lancet. 2020;396(10265):1817-1828.
2. Bardia A, Hurvitz SA, Tolaney SM, et al. Sacituzumab govitecan in metastatic triple-negative breast cancer. N Engl J Med. 2021;384(16):1529-1541.
3. Modi S, Jacot W, Yamashita T, et al. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. N Engl J Med. 2022;387(1):9-20.

This research was sponsored by Norstella, which all authors were employed by at the time of conduct. The authors declare no conflict of interest in connection with the presented work.