

Why are cancer phenotypes in RWE studies often wrong?

Asieh Golozar

10 Nov 2026



How studies define advanced or metastatic NSCLC

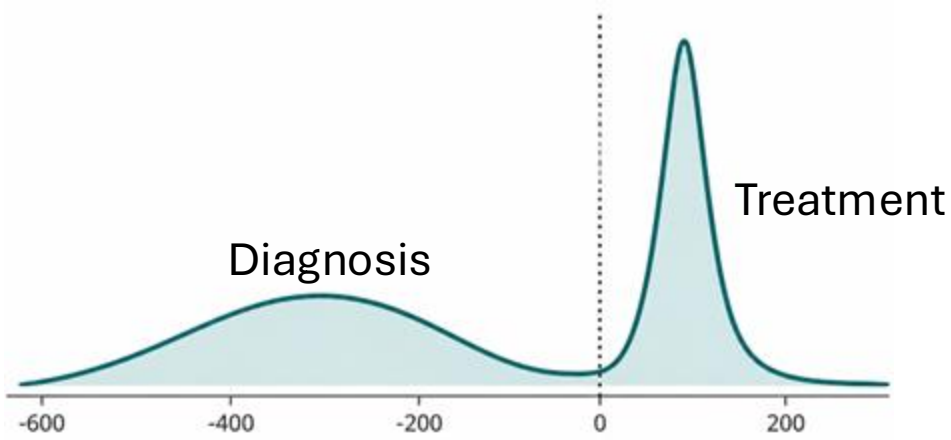
Study	Phenotype rules
PMID 29799325	Clinician-reported NSCLC. ≥2 metastatic claims ≤90 days apart, with ≥1 claim from −180 to +90 days relative to first treatment.
PMID 39498908	≥2 lung cancer claims 30–180 days apart and ≥1 metastasis code within ±60 days of the first lung claim. Excluded ≥2 other-cancer claims 30 days apart in the 6-mo pre-biopsy baseline.
EUPAS1000000112	Initial TNM stage IIIB or IV with NSCLC. Excluded other primaries any time before NSCLC diagnosis.
PMID 39011928	≥2 lung cancer claims ≥30 days apart and ≥1 secondary malignancy claim from 30 days before first lung diagnosis onward. Excluded other primaries from 6 mo before metastasis through the day before first-line EGFR-TKI.
PMID 37074589	≥2 lung cancer diagnoses. A metastasis diagnosis or first-line treatment within 30 days after first lung diagnosis qualified as advanced disease. Excluded other cancers before first lung diagnosis.
PMID 34455553	A lung cancer code followed by a metastasis code. Excluded prior other cancer or metastasis and SCLC-specific therapy. Required lung cancer treatment after first metastasis.

How studies define advanced or metastatic NSCLC

Study	Lung cancer codes	Metastasis evidence	Other-cancer exclusion	Treatment rule
PMID 29799325	Clinician-reported NSCLC	**≥2 claims ≤90 days apart**	--	Metastasis within **–180 to +90 days** of 1st treatment
PMID 39498908	**≥2 claims 30–180 days apart**	≥1 code within **±60 days** of first lung claim	**≥2 claims** in 6-month pre-biopsy baseline	--
EUPAS1000000112	Initial diagnosis, **no repeat-code rule**	**TNM stage IIIB or IV**	**Any time** before NSCLC diagnosis	Selected therapy **≤6 months** after diagnosis
PMID 39011928	**≥2 claims ≥30 days apart**	≥1 secondary malignancy claim from **30 days before** first lung diagnosis	From **6 months before** metastasis to first-line EGFR-TKI	1st line EGFR-TKI ends the exclusion window
PMID 37074589	**≥2 diagnoses**	Metastasis code **or first-line treatment** ≤30 days after first lung diagnosis	Before 1st lung diagnosis	Treatment can stand in for metastasis
PMID 34455553	**1 code, no repeated code**	Metastasis code after lung code	Prior cancer or metastasis; **SCLC-specific therapy**	Lung cancer treatment **after first metastasis**

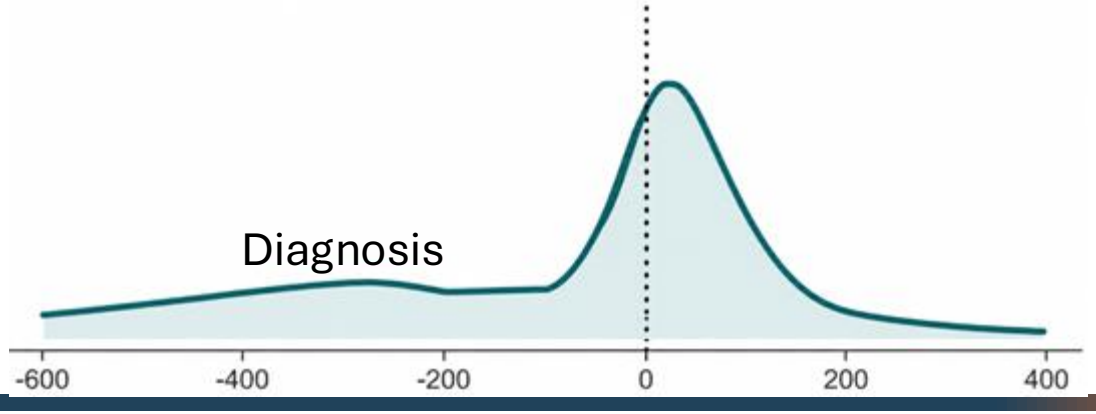
The clinical narratives and the real-world

Real-world data may not be what you expect

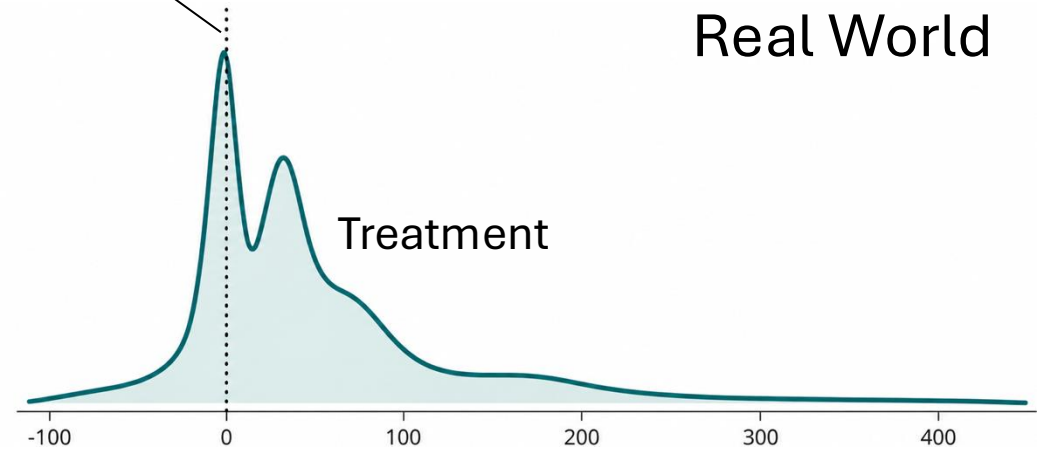


Expected narrative

Detection of metastasis



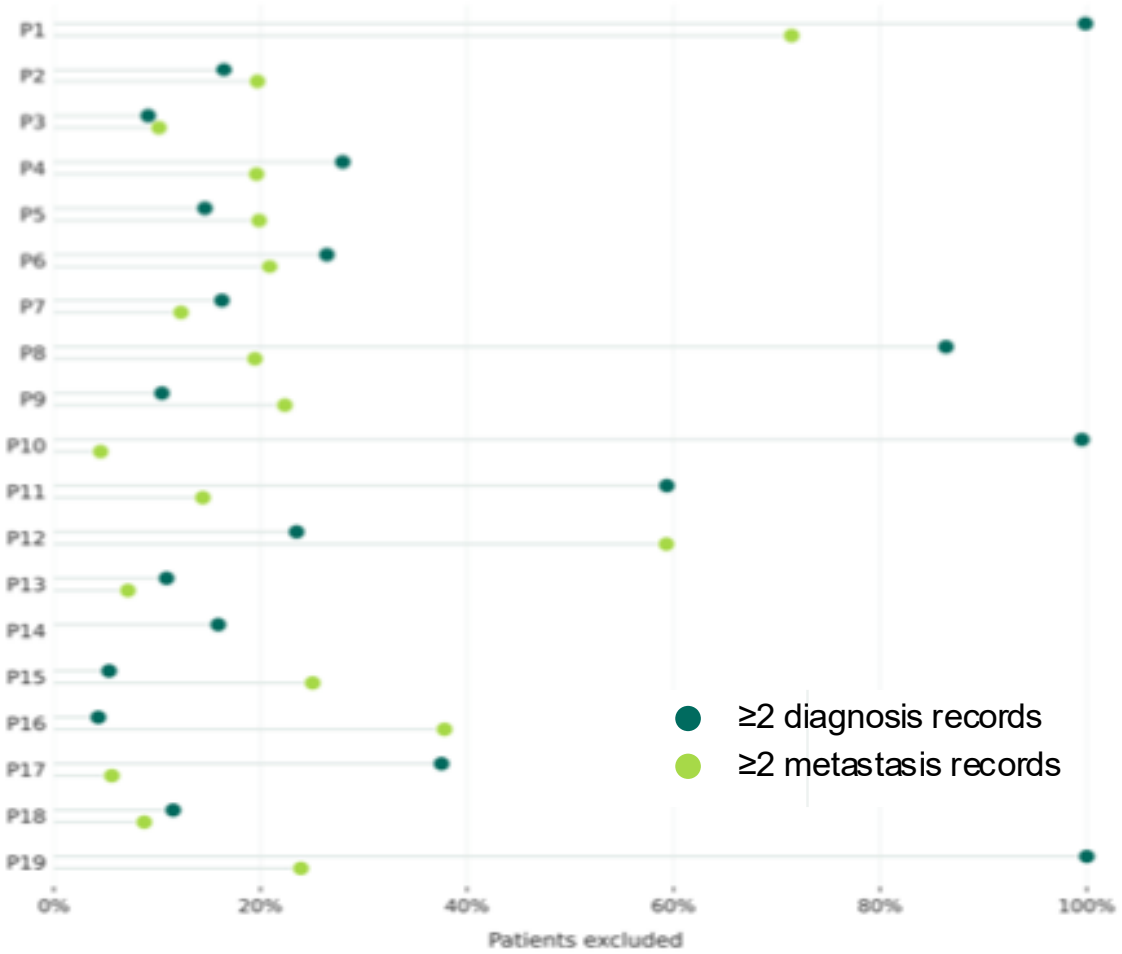
Real World



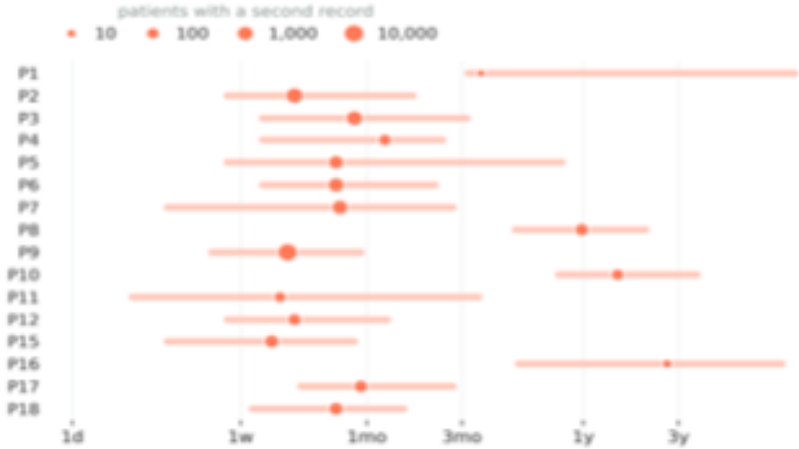
Requiring a second record

6–100% of mBC patients have **only one diagnosis or metastasis record**

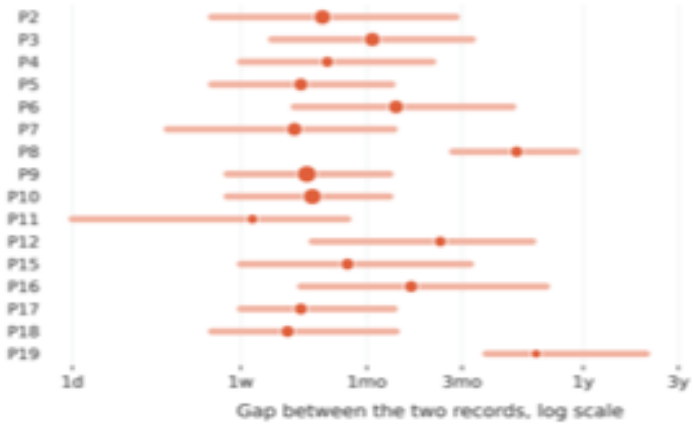
A. % patient loss by requiring 2 diagnosis or metastasis codes



B. Timing to a second diagnosis code

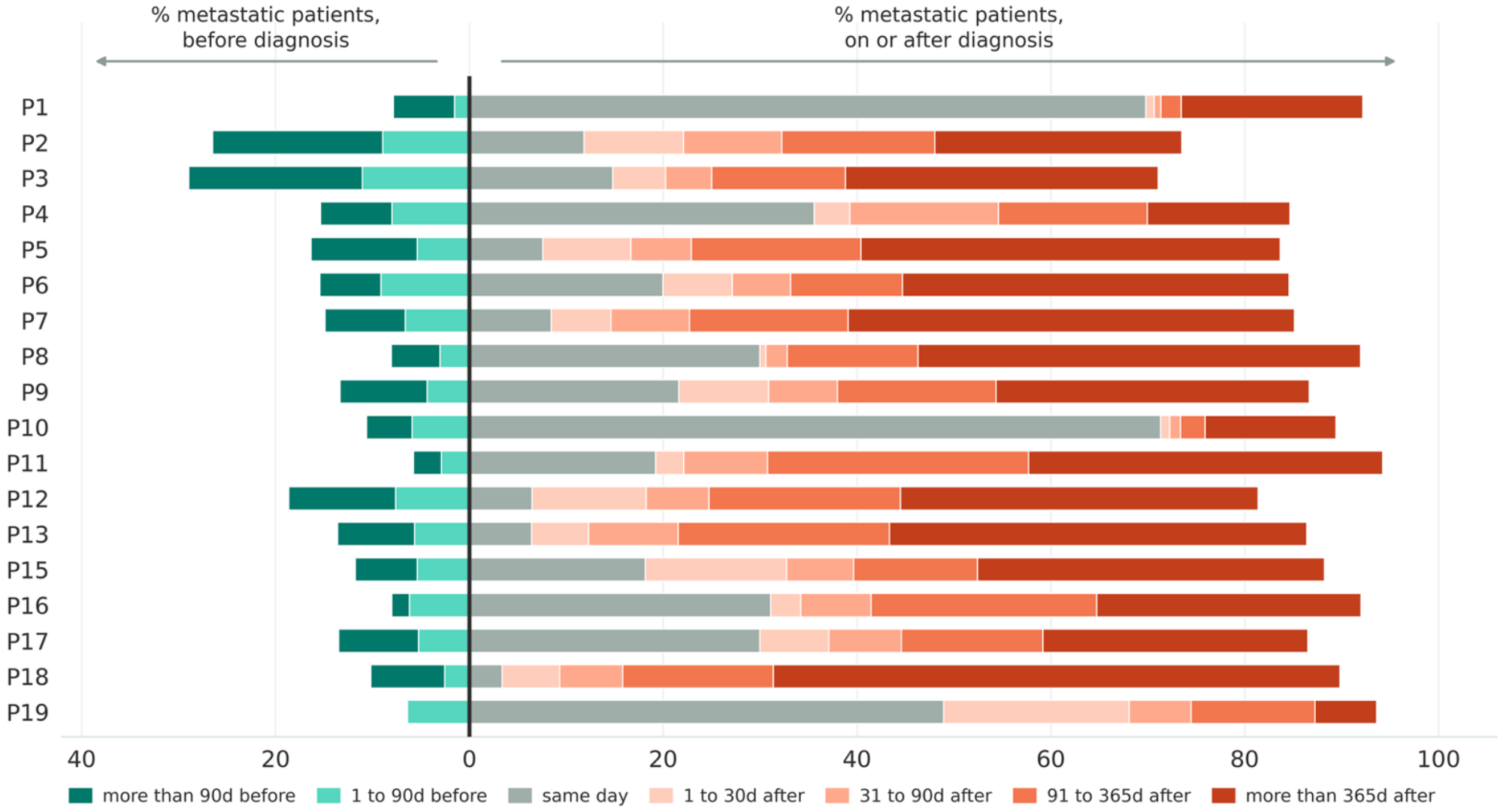


C. Timing to a second metastasis code



Requiring diagnosis before metastasis

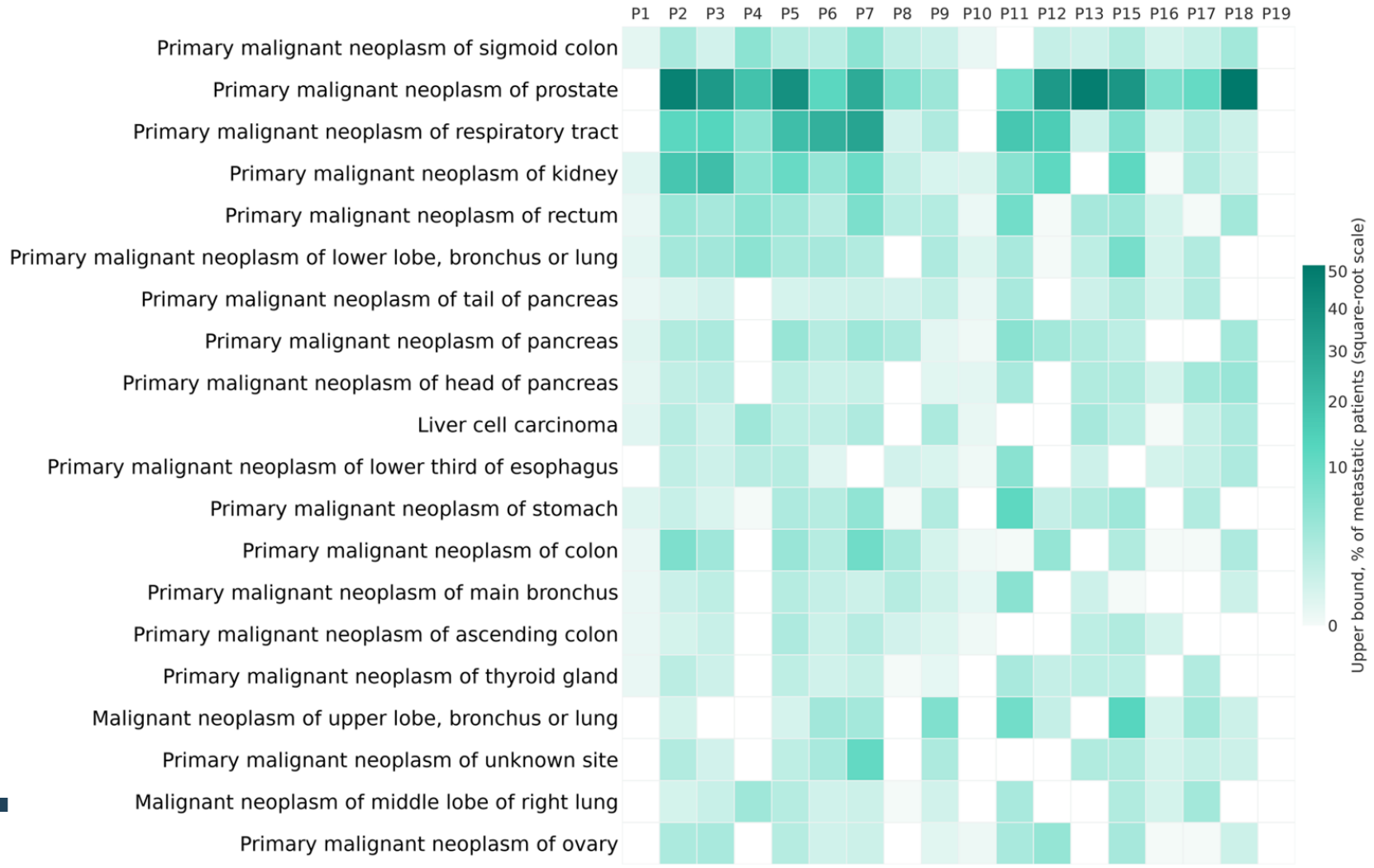
3–29% of mBC patients have **the metastasis record before the diagnosis**



Excluding other primaries

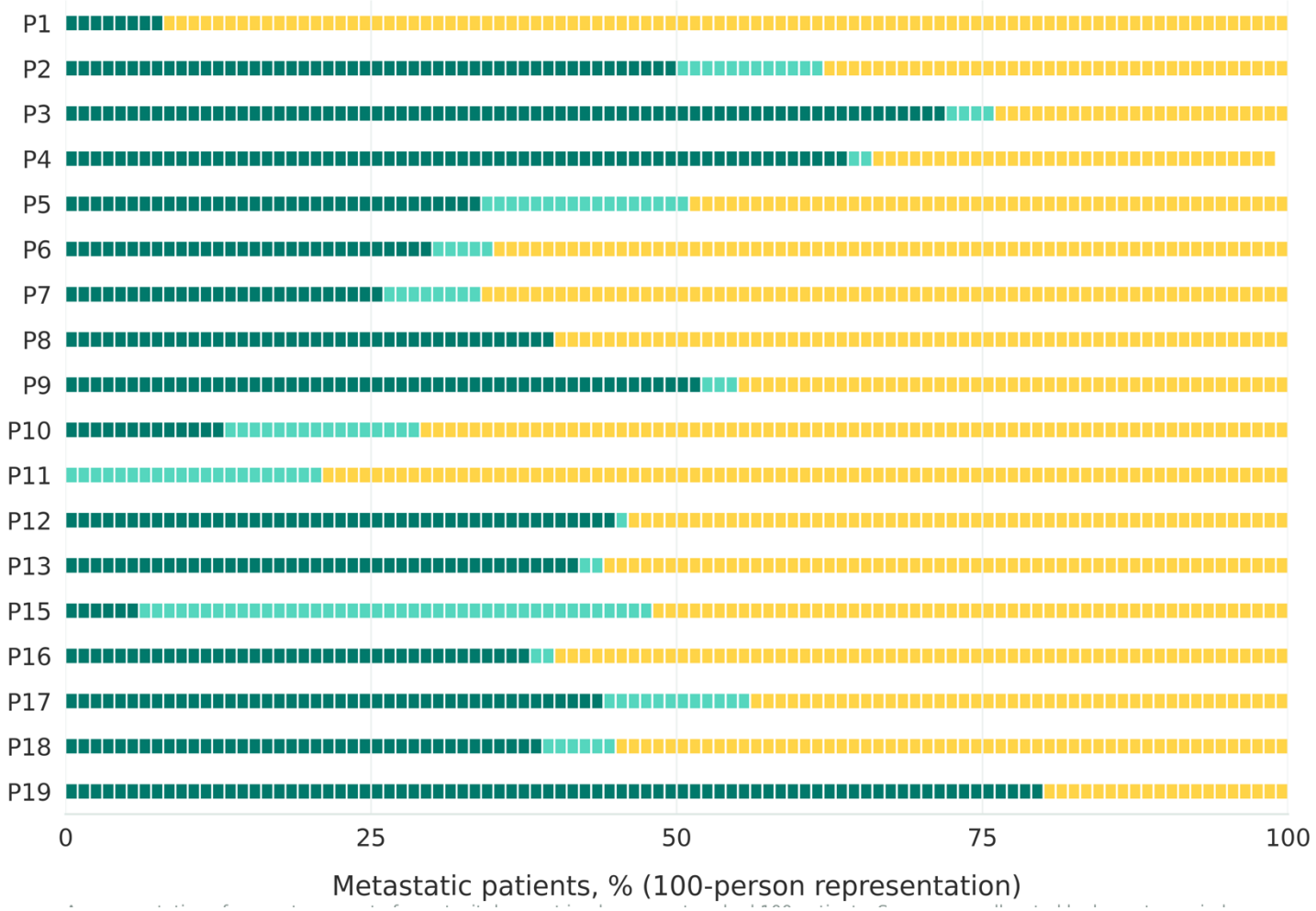
mBC patients have a **median of 14 (range 1–94) other cancer diagnoses**

Top 20 other cancer codes within 90 days of metastasis



Treatment after metastasis

Treatment information after metastasis varies by data partner



Each square is 1 of 100 metastatic patients, colored by treatment signal on or after the metastasis date

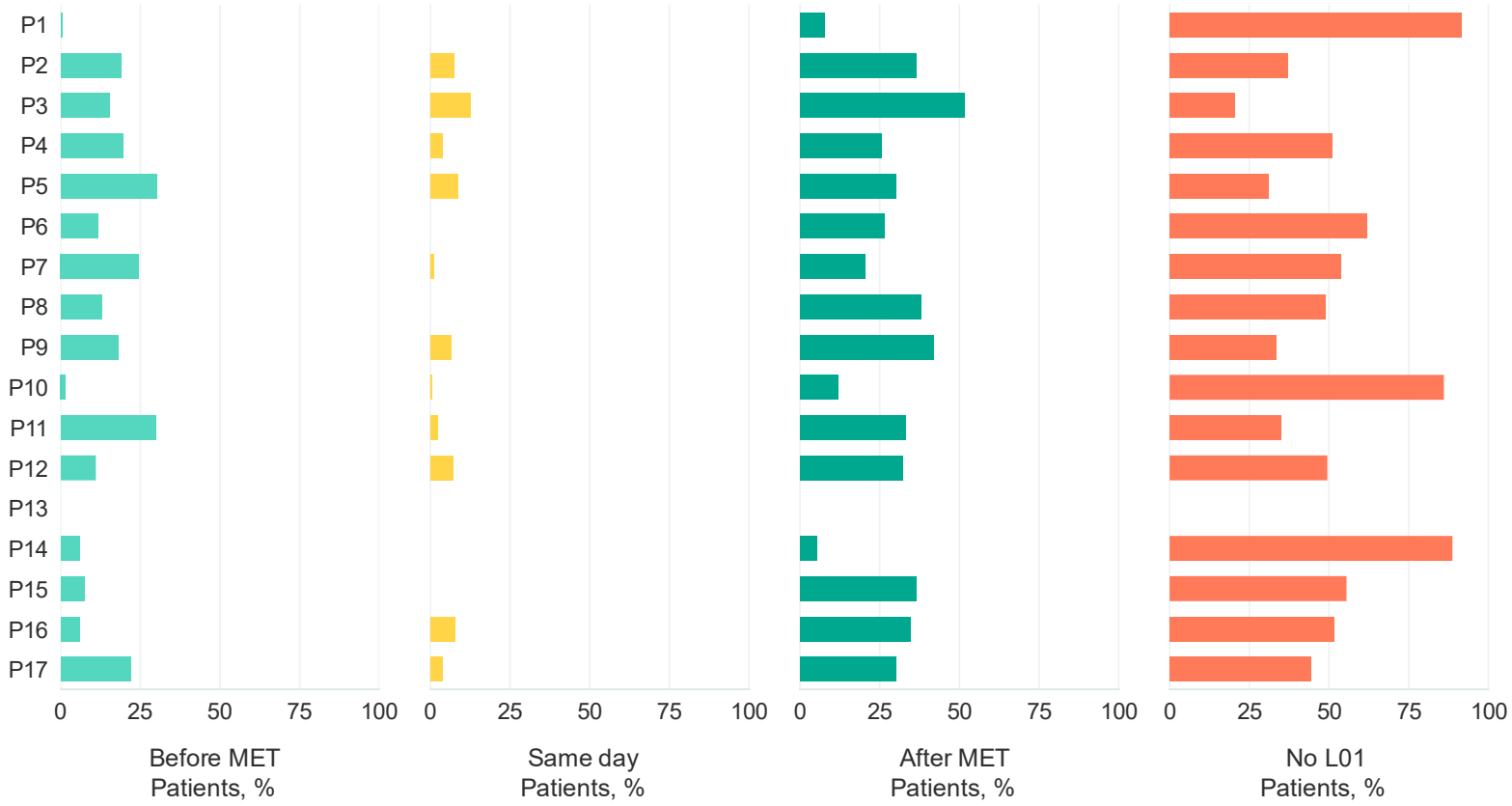
L01 drugs

Procedure only

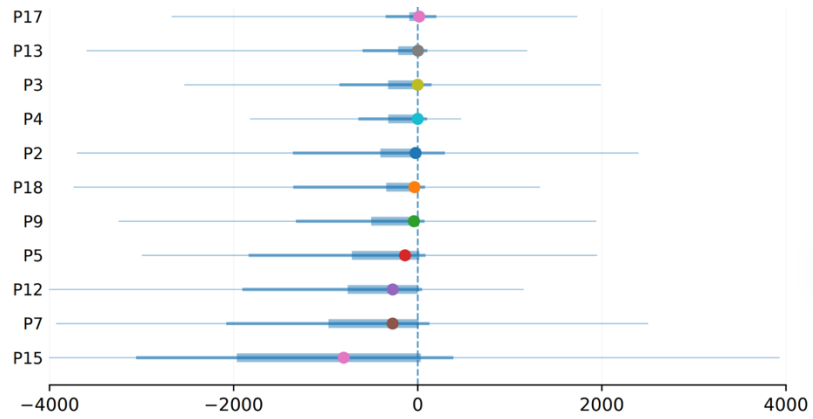
No treatment

Treatment timing relative to first metastasis

Patients by systemic treatment timing, %

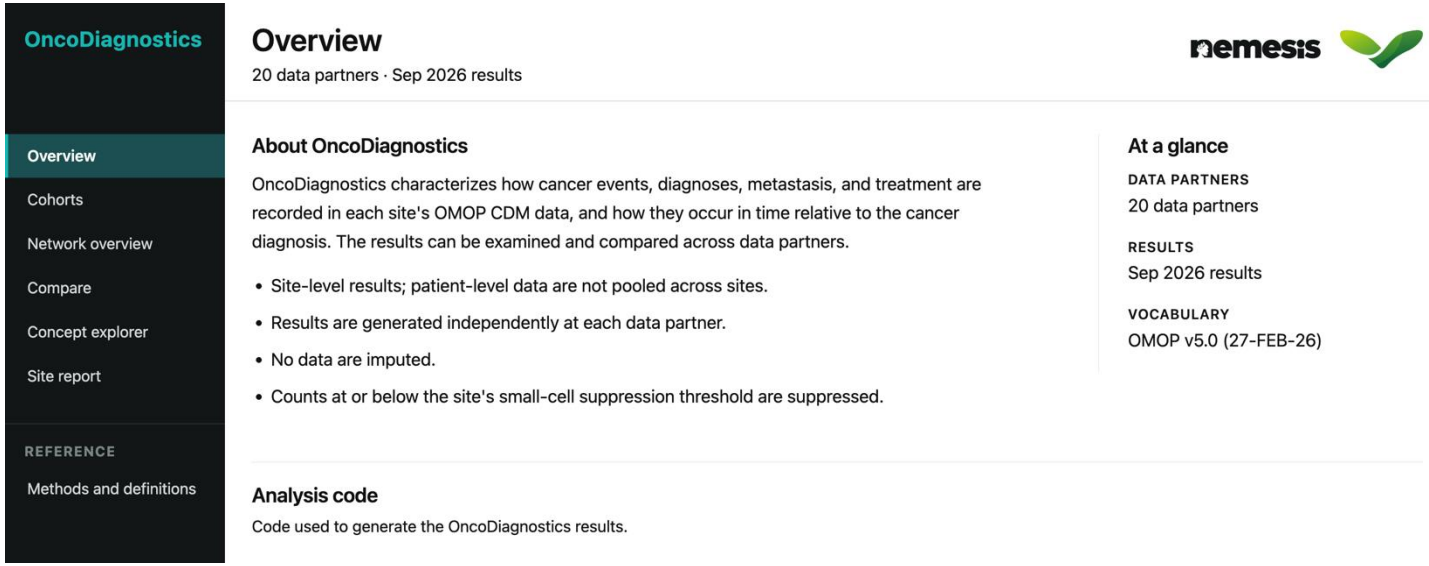


Closest antineoplastic drug to first metastasis



OncoDiagnostics profiles cancer data at each site

[OncoDiagnostics]



What it does

Characterizes diagnoses, metastasis and treatment in each site's OMOP CDM data, and when they occur relative to diagnosis

How it runs

Generated independently at each data partner; site-level results only, no pooling

Safeguards

No imputation; small cell counts suppressed

Current release

Focused on diagnosis, metastasis and treatment

Why cancer phenotypes in RWE studies are often wrong?

FALCON Oncology Research Network