Concordance of ER and HER2 Status by Claims-based Algorithm and Large Language Models (LLM) Abstracted Clinical Notes in Patients with Metastatic Breast Cancer (mBC)

Authors: Jiemin Liao[1], Aaron Hardin[1], Caroline Weipert[1], Sara Wienke[1], Jose Mena[2], Amar Das[1]

Affiliations: [1]Guardant Health, Palo Alto, CA; [2]Mendel AI, San Jose, CA
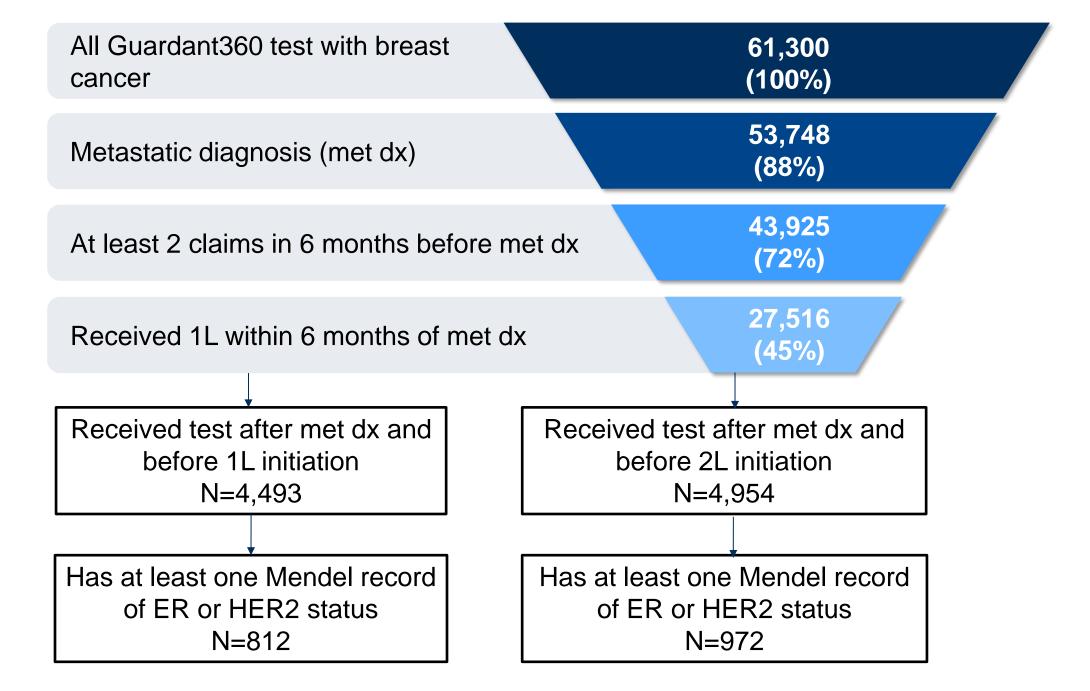
GUARDANTINFORM™

Introduction

- Breast cancer is a highly heterogenous disease, comprised of different subtypes with distinct clinical behaviors that require different treatment pathways.
- Breast cancer subtypes are defined based on three receptors: estrogen receptor (ER); progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2)
- As more real-world data is utilized for studying and defining breast cancer subtypes, there is a need to understand the accuracy of defining subtypes using claims-based algorithms.
- **Objective: We seek to compare the concordance of ER and HER2 status between claims-derived information and abstracted clinical notes.**

Methods

- **Claims Data Source:** Patients were identified from the GuardantINFORM database, which links cell-free circulating tumor DNA (cfDNA) results to de-identified claims data, with study time period from January 2014 to June 2024.
- **Clinical Notes Data Source:** Clinical notes submitted at the time of a Guardant test were abstracted with Mendel.ai's LLM and symbolic reasoning tools with human review
- **Inclusion and exclusion criteria:**
  - Adult patients in the US with breast cancer diagnosis indicated on their Guardant360 test requisition form
  - Received at least one metastatic diagnosis code from claims information
  - Received at least 2 claims in the 6 months before first metastatic diagnosis
  - Initiated first line therapy (1L) within 6 months of metastatic diagnosis
  - Have at least one clinical record of ER or HER2 status based on FISH, IHC, or tissue testing
  - 1L cohort: Received Guardant360 test after metastatic diagnosis and before 1L initiation
  - 2L cohort: Received Guardant360 test after metastatic diagnosis and before 2L initiation
- **Concordance analysis:**
  - Concordance between claims-derived ER and HER2 status, and abstracted clinical ER and HER2 status were compared using:
    - Cohen's Kappa statistic (κ), where κ>0.8 indicates almost perfect agreement, 0.6-0.8 indicates substantial agreement and 0.4-0.6 indicates moderate agreement.
    - Total number of patients with concordant values out of all patients.
- **Primary analysis:** Included all clinically abstracted information (with or without date of test)
- **Secondary analysis:** Included only clinically abstracted information with dates from 2014 onwards
- **ER+ status:** we investigated 2 definitions, (A) only including treatments (B) including treatment and ICD 9/10 code V86.0 or Z17.0 for estrogen receptor positive status
- **HER2+ status:** we investigated 2 definitions, (A) with fam-trastuzumab-deruxtecan-nxki (T-DXd) (B) without T-DXd

Figure 1. Patient demographic and clinical characteristics

CONCLUSIONS

- We demonstrated concordance of ER and HER2 status between claims-based algorithm and clinical notes was high, highlighting the value of treatment information for identify ER or HER2 positivity.
- Inclusion of ICD diagnosis codes for defining ER positivity did not improve the concordance consistently.
- Inclusion of T-DXd for defining HER2 positivity did not impact concordance
- One should be cautious when using claims-based algorithm only to identify ER status, especially in settings where neoadjuvant or adjuvant therapies are not well captured or distinguished from metastatic treatment lines

Defining ER and HER2 status using claims information

ER+

Treatments
tamoxifen
toremifen
fulvestrant
letrozole
anastrozole
exemestane
megestrol
estradiol
palbociclib
ribociclib
abemaciclib
everolimus
alpelisib
elacestrant

HER2+

Treatments
trastuzumab
pertuzumab
ado-trastuzumab emtansin
fam-trastuzumab-deruxtecan-nxki (T-DXd, received prior to August 2022)

Patients **with** the identified treatments at or any time before metastatic 1L or 2L were considered ER+ or HER2+ at 1L or 2L

Patients **without** any identified treatment at or any time before metastatic 1L or 2L were considered ER- or HER2- at 1L or 2L

Defining ER and HER2 status using extracted clinical notes

ER+ status: Using IHC or Tissue Tests, results with positive, strongly positive or weakly positive
ER- status: negative, weakly negative, strongly negative
Missing: high, intermediate, low, equivocal

HER2+ status: Using IHC (3+), FISH or Tissue Tests, results with positive, strongly positive or weakly positive
HER2- status: equivocal, negative, weakly negative, strongly negative
Missing: high, intermediate, low

Figure 2. Defining ER and HER2 status using claims and extracted clinical notes

Results

ER CONCORDANCE

PRIMARY ANALYSIS

	CLAIMS	
CLINICAL NOTES	ER Negative	ER Positive

FIRST LINE

	ER Negative	ER Positive
ER Negative	78	17
ER Positive	36	235

(A) Concordance=86%; κ=0.65
(B) Concordance=87%; κ=0.65

SECOND LINE

	ER Negative	ER Positive
ER Negative	60	23
ER Positive	26	334

(A) Concordance=89%; κ=0.64
(B) Concordance=88%; κ=0.53

SECONDARY ANALYSIS

	CLAIMS	
CLINICAL NOTES	ER Negative	ER Positive

FIRST LINE

	ER Negative	ER Positive
ER Negative	25	5
ER Positive	10	78

(A) Concordance=87%; κ=0.68
(B) Concordance=89%; κ=0.71

SECOND LINE

	ER Negative	ER Positive
ER Negative	20	9
ER Positive	7	104

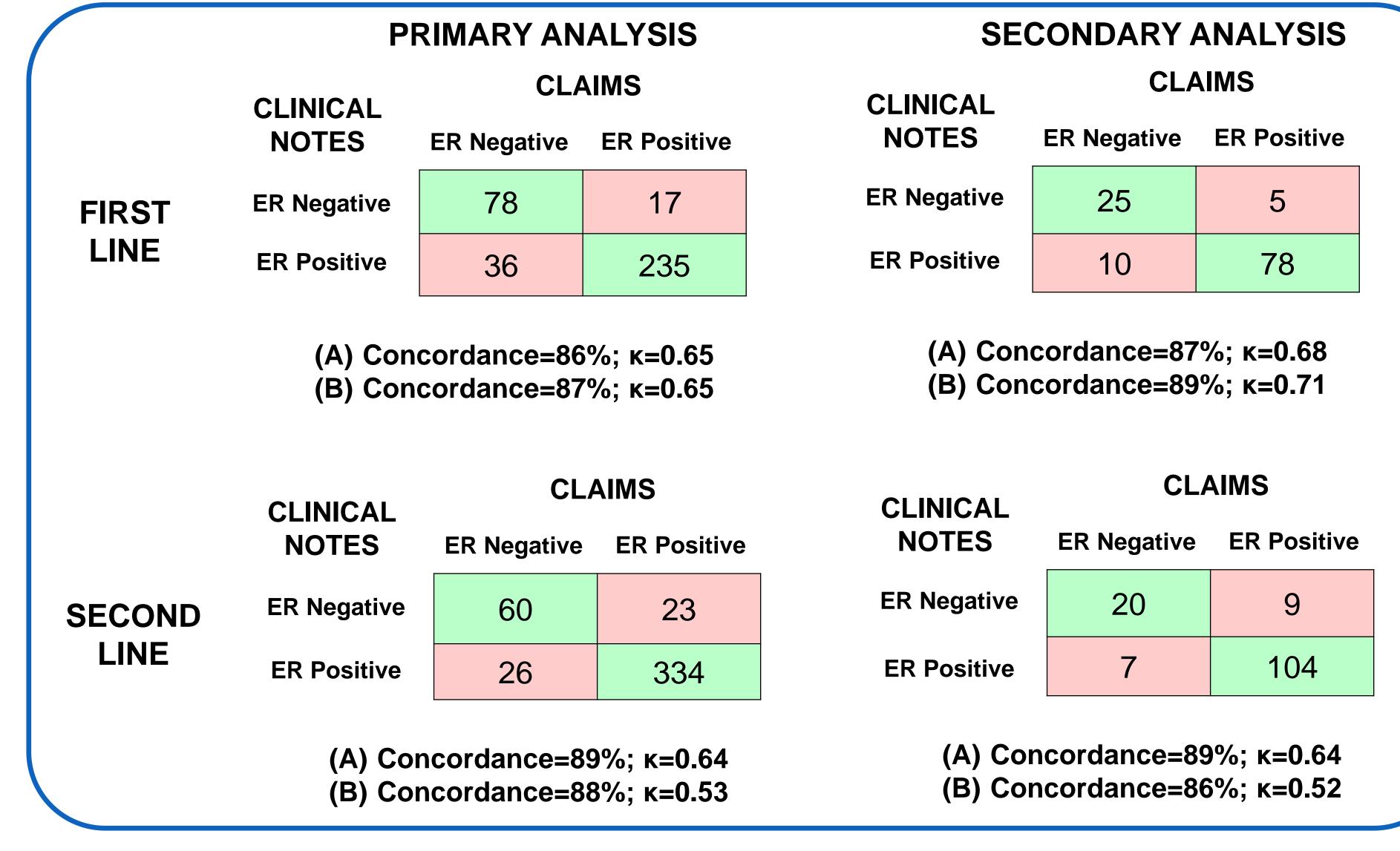(A) Concordance=89%; κ=0.64
(B) Concordance=86%; κ=0.52

Figure 3. Concordance and agreement between claims and abstracted clinical notes for ER status, stratified by first and second line of therapy. (A) includes treatment information only for claims definition; (B) includes treatment and ICD diagnosis codes information for claims definition

HER2 CONCORDANCE

PRIMARY ANALYSIS

	CLAIMS	
CLINICAL NOTES	HER2 Negative	HER2 Positive

FIRST LINE

	HER2 Negative	HER2 Positive
HER2 Negative	641	6
HER2 Positive	33	83

(A) Concordance=95%; κ=0.78
(B) Concordance=95%; κ=0.78

SECOND LINE

	HER2 Negative	HER2 Positive
HER2 Negative	734	22
HER2 Positive	31	134

(A) Concordance=94%; κ=0.80
(B) Concordance=95%; κ=0.81

SECONDARY ANALYSIS

	CLAIMS	
CLINICAL NOTES	HER2 Negative	HER2 Positive

FIRST LINE

	HER2 Negative	HER2 Positive
HER2 Negative	143	1
HER2 Positive	2	17

(A) Concordance=98%; κ=0.91
(B) Concordance=98%; κ=0.91

SECOND LINE

	HER2 Negative	HER2 Positive
HER2 Negative	155	11
HER2 Positive	1	24

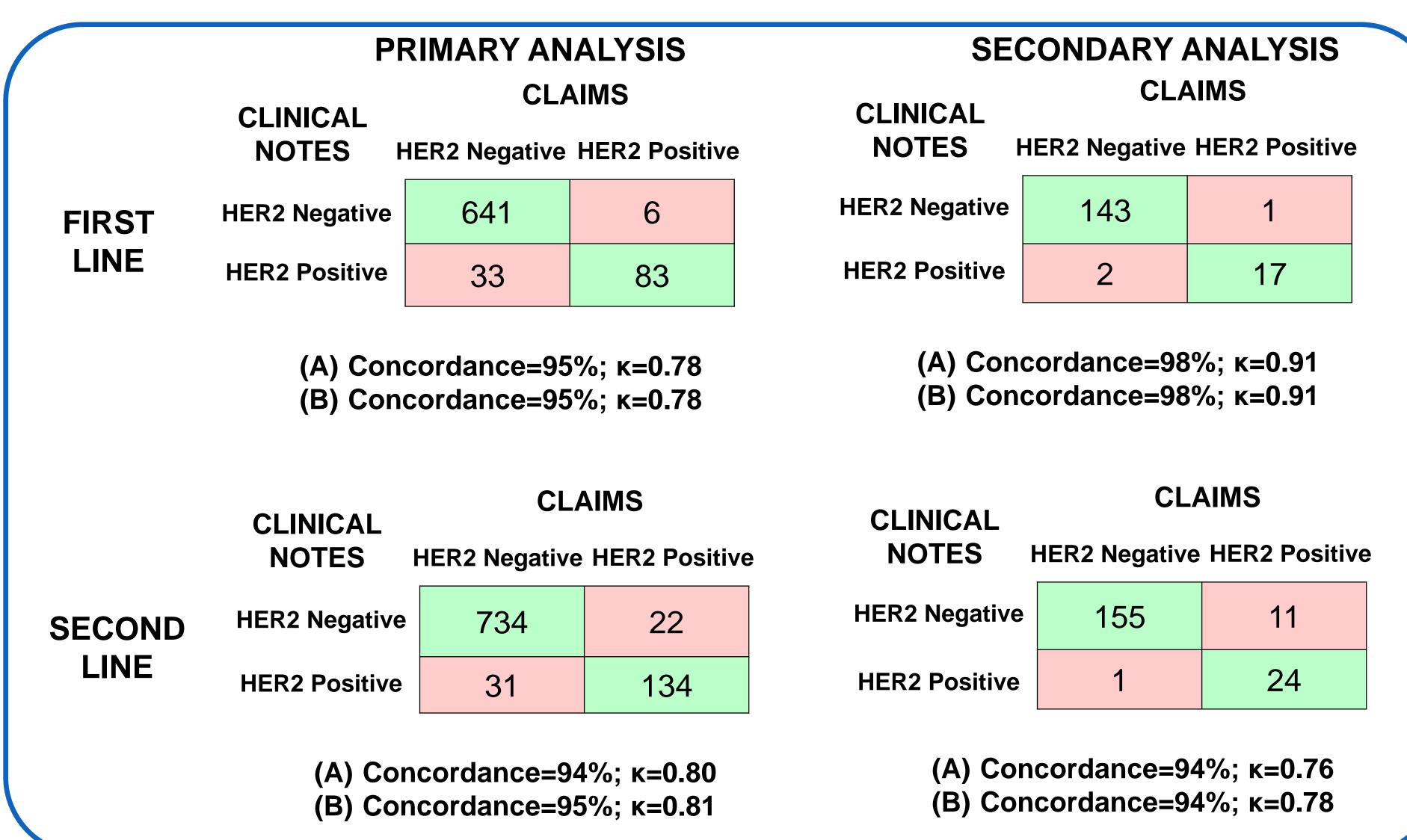(A) Concordance=94%; κ=0.76
(B) Concordance=94%; κ=0.78

Figure 4. Concordance and agreement between claims and abstracted clinical notes for HER2 status, stratified by first and second line of therapy. (A) includes T-DXd (from August 2022 onwards); (B) excludes T-DXd from treatment list