

If you're a person with a disability or using assistive technology and are having difficulty accessing any of the content in this file, please contact Astrid Masfar at [astrid.masfar@nih.gov](mailto:astrid.masfar@nih.gov).

Methods: Mind the Gap  
Webinar Series

---

# **Risk Prediction Models for Breast Cancer in Black Women: Importance of Considering Molecular Subtypes**



Presented by:

Julie R. Palmer, Sc.D., M.P.H.

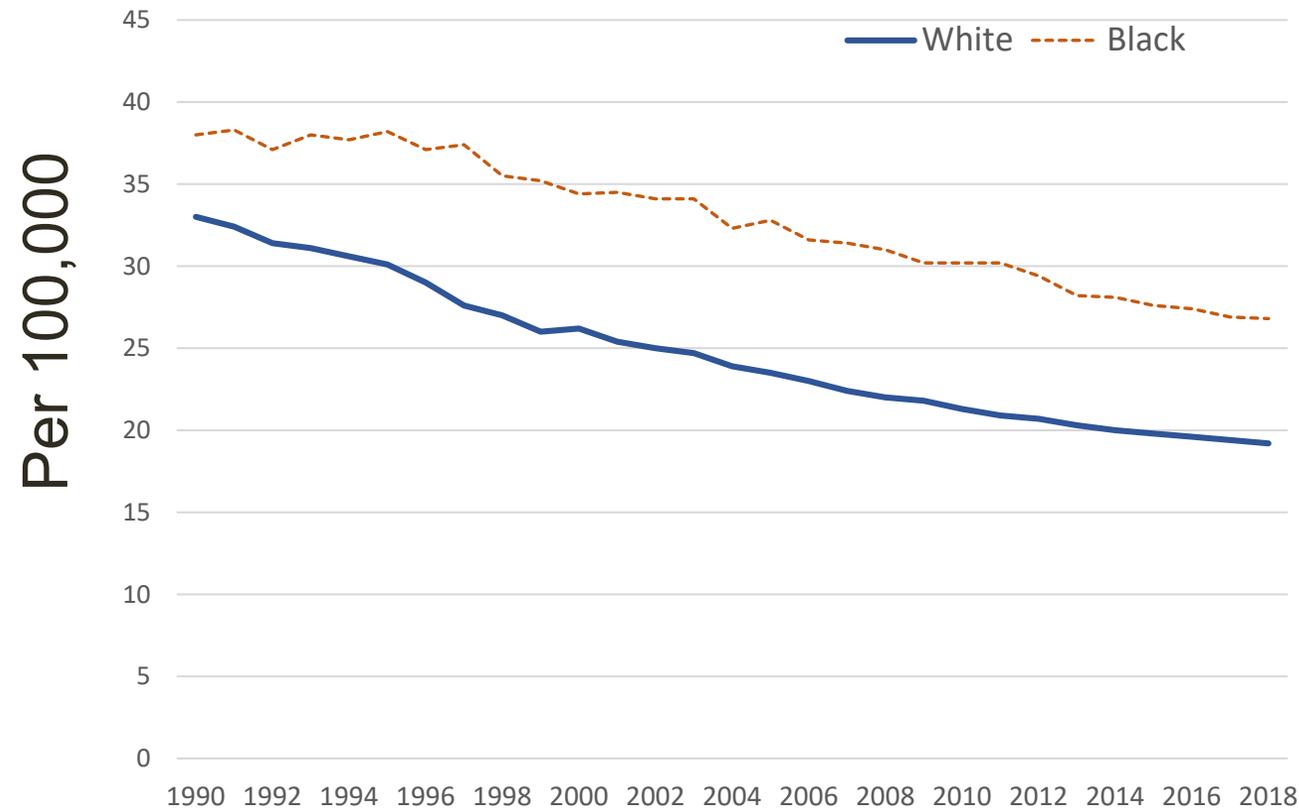
Boston University School of Medicine



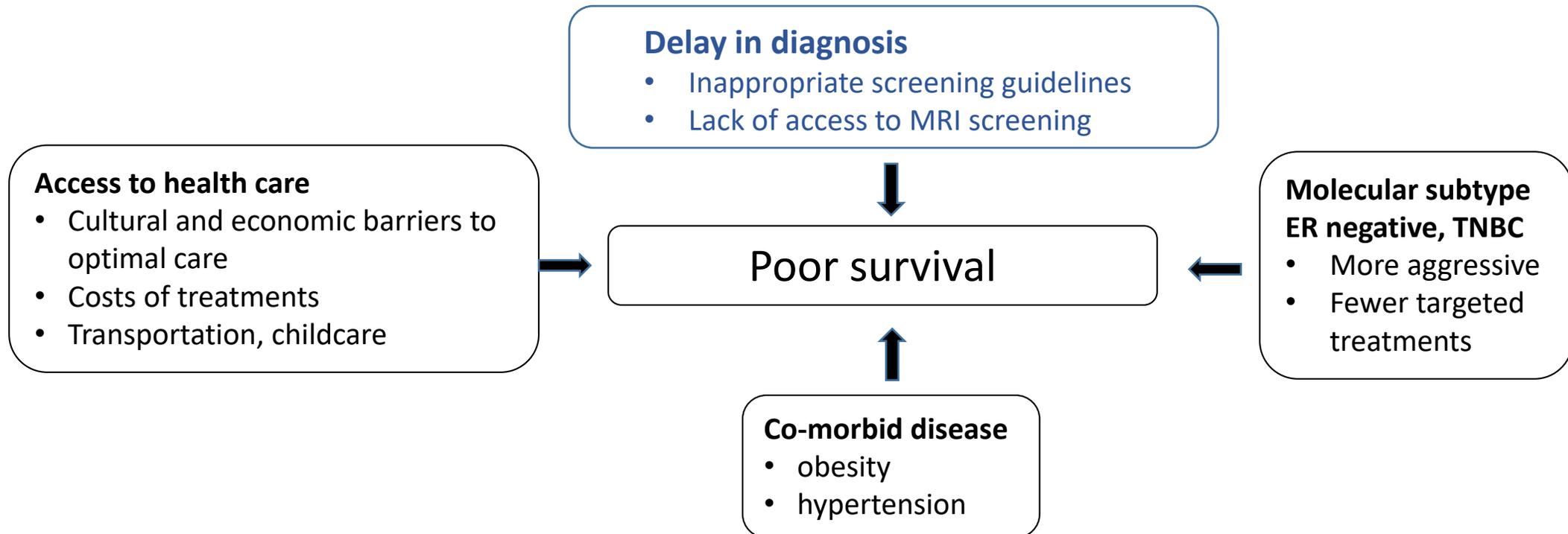
National Institutes of Health  
*Office of Disease Prevention*

# Disparities in breast cancer mortality

Age-adjusted breast cancer mortality rates, 1990-2018



# Factors contributing to excess breast cancer mortality



# Risk prediction models

---

Estimate the probability that an individual with defined risk factors and free of the disease at a given age will be diagnosed with the disease during a given risk period (absolute risk)

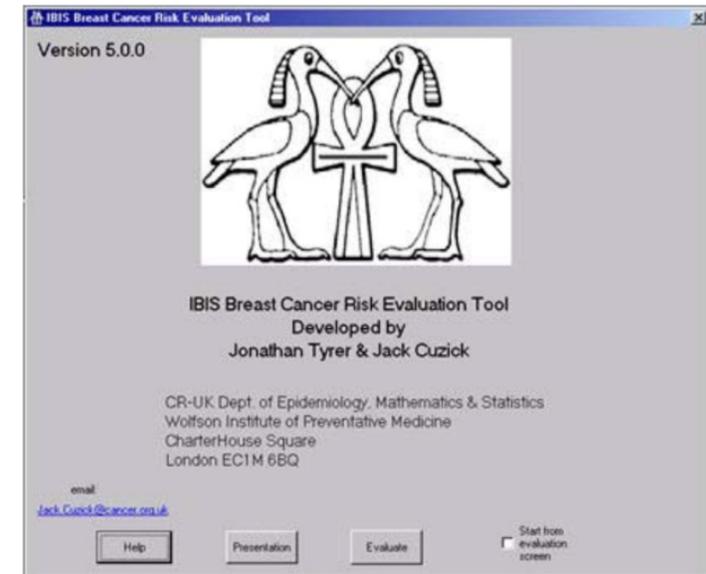
# How are risk prediction models used?

- **Before mammographic screening**
  - Refer for earlier screening
  - Eligibility for prevention trials
  - Use of established chemopreventives
- **After mammographic screening**
  - Additional screening modalities
  - Genetic testing



Breast Cancer Risk Assessment Tool

RISK CALCULATOR ABOUT THE CALCULATOR



# What factors are included in breast cancer risk prediction models?

---

- Familial history of breast cancer, other cancers
- Nongenetic risk factors such as age at menarche, hormone use, reproductive factors, alcohol consumption
- **Polygenetic risk score**
- **Hormone levels**
- **Mammographic density**

# Performance of risk prediction models

---

## Calibration

- Ability of a model to predict the number of events that arise in an independent validation cohort; is related to average risk in a population
- Expected / observed

## Discrimination

- AUC - Area under the receiver operating characteristic curve
- How well a model separates risk between cases and noncases
- Probability that a randomly selected case has a higher projected risk than a randomly selected noncase

# Performance of prediction models: White women

---

## **BCRAT (or Gail model)**

- Tested in WHI, CPS-II, NHS, PLCO
- **AUC = 0.62**

## **Updated Rosner & Colditz model**

- Developed and tested in data from the Nurses Health Study
- **AUC = 0.64**

## **IBIS model (International Breast Cancer Intervention Study)**

- Tested in data from ProF-SC (Breast Cancer Prospective Family Study Cohort)
- **AUC = 0.62**

# Performance of prediction models: Black women

---

## **CARE model** (modified BCRAT model)

- Developed in data from Black women in Women's CARE Study
- tested in postmenopausal WHI Black women
- **AUC = 0.55**
- tested in BWHS participants, pre- and postmenopausal
- **AUC = 0.57**

## **BWHS model (Black Women's Health Study)**

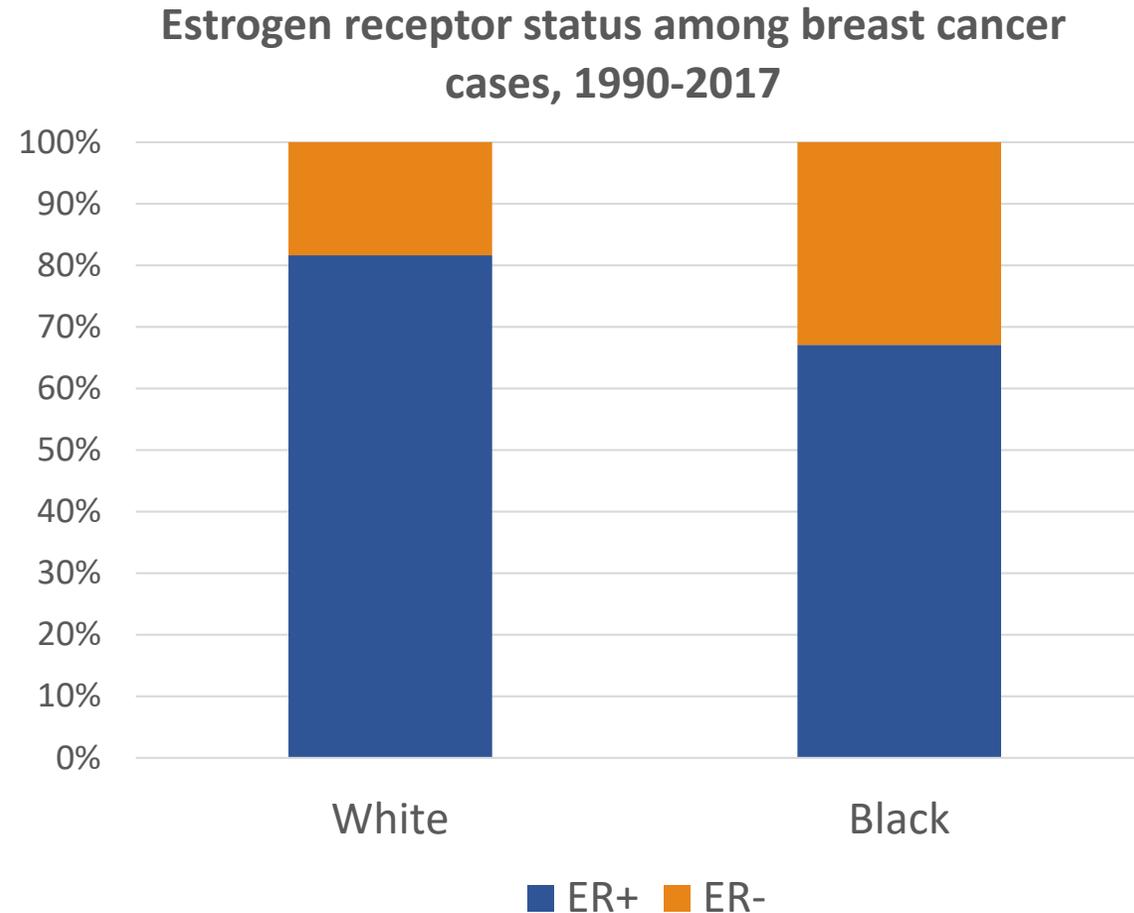
- based on first 10 years of follow-up in BWHS
- tested in next 5 years of follow-up
- **AUC = 0.59**

# Why is performance worse among Black women?

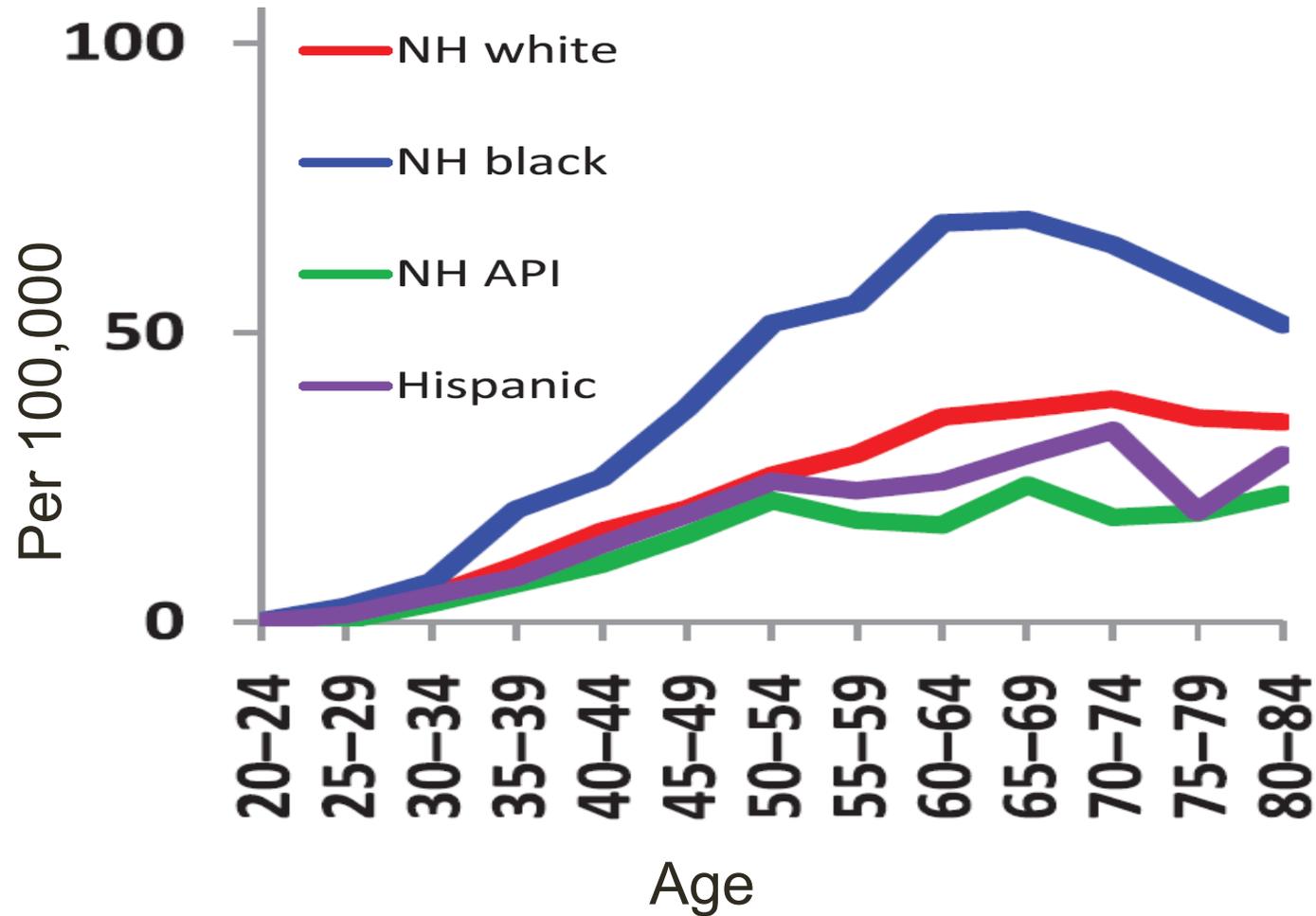
---

- Most of the risk factors identified are risk factors for hormonally responsive breast cancers (ER+)
- Proportion of breast cancers that are ER- approximately twice as high in NHB women as compared with NHW women

# Estrogen receptor (ER) status by race



# Age-specific incidence of TNBC in U.S. SEER data

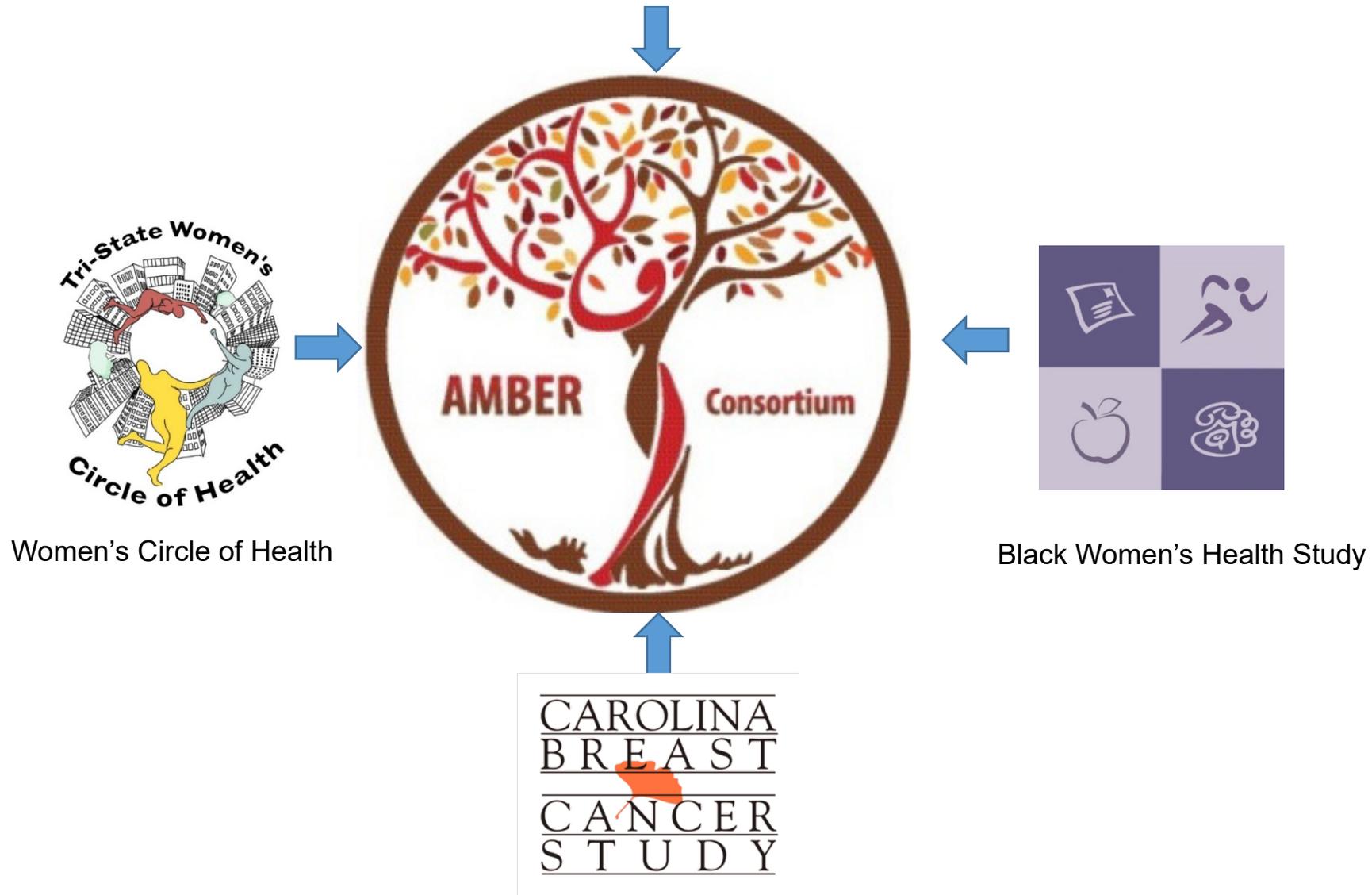


# Why is performance worse among Black women?

---

- Most of the risk factors identified are risk factors for hormonally responsive breast cancers (ER+)
- Proportion of breast cancers that are ER- approximately twice as high in NHB women as compared with NHW women
- **If differing risk factors for ER+ and ER- breast cancer, then greater impact on performance of risk prediction model in Black women**

# The Multiethnic Cohort Study



# Breast cancer cases and controls in AMBER consortium

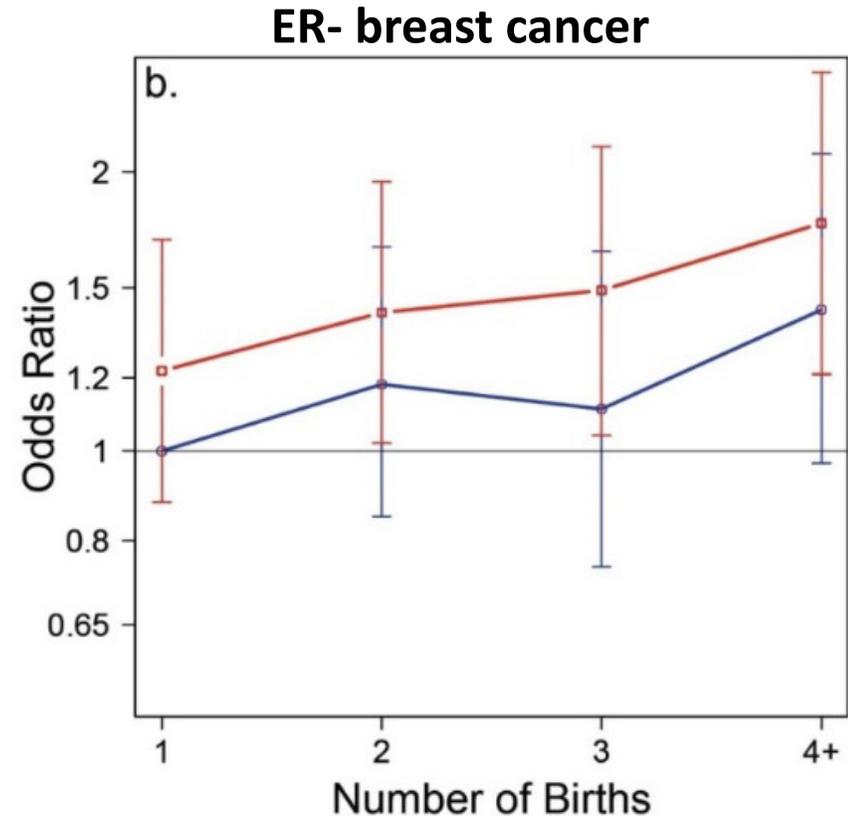
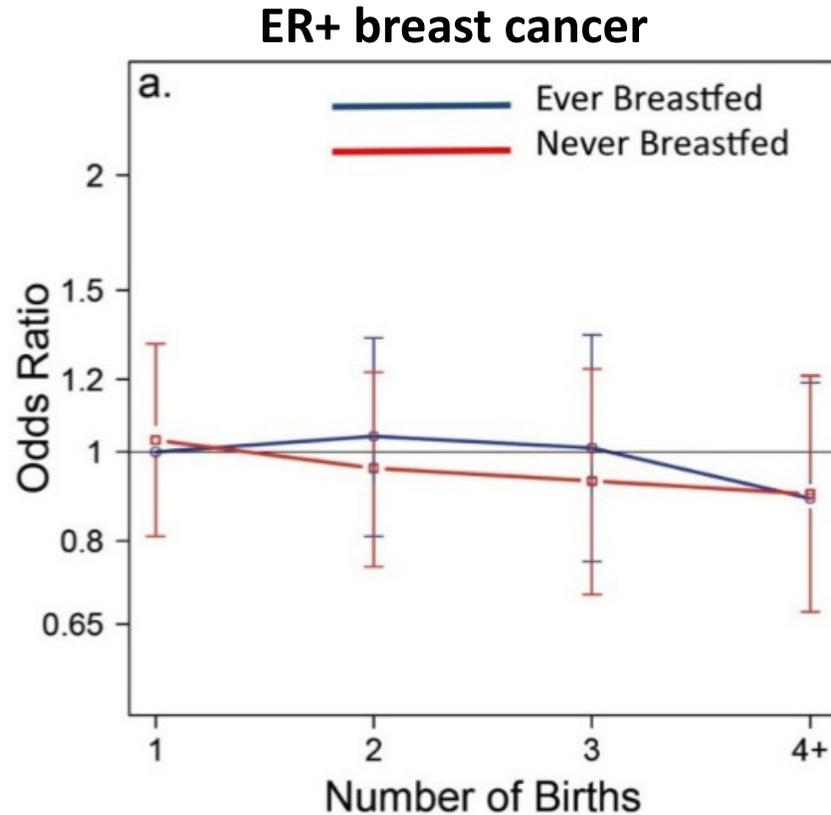
---

Cases and Controls	Black Women's Health Study	Women's Circle of Health Study	Multi-ethnic Cohort Study	Carolina Breast Cancer Study	Total
Breast cancer cases	2,533	1,691	1,221	894	<b>6,339</b>
ER+	1,444	937	753	405	<b>3,539</b>
ER-	614	352	260	401	<b>1,627</b>
TNBC	262	207	107	233	<b>809</b>
Controls	11,771	1,271	4,895	788	18,725

# AMBER: Parity, breastfeeding, and ER subtypes

Categories Tested	N	ER+ OR (95% CI)	N	ER- OR (95% CI)
Nulliparous	444	1.00 Reference	170	1.00 Reference
Parous	<b>2,006</b>	<b>0.91 (0.80-1.02)</b>	<b>1,084</b>	<b>1.33 (1.11-1.58)</b>
Number of births				
1	509	1.00 Reference	238	1.00 Reference
2	605	0.97 (0.85-1.11)	343	1.17 (0.97-1.40)
3	397	1.04 (0.89-1.22)	224	1.25 (1.01-1.54)
≥ 4	495	0.92 (0.79-1.08)	279	1.29 (1.04-1.60)
Breastfeeding				
Never	823	1.00 Reference	556	1.00 Reference
Ever	<b>646</b>	<b>1.02 (0.90-1.15)</b>	<b>329</b>	<b>0.84 (0.72-0.99)</b>

# AMBER: Parity and breast cancer, by breastfeeding status

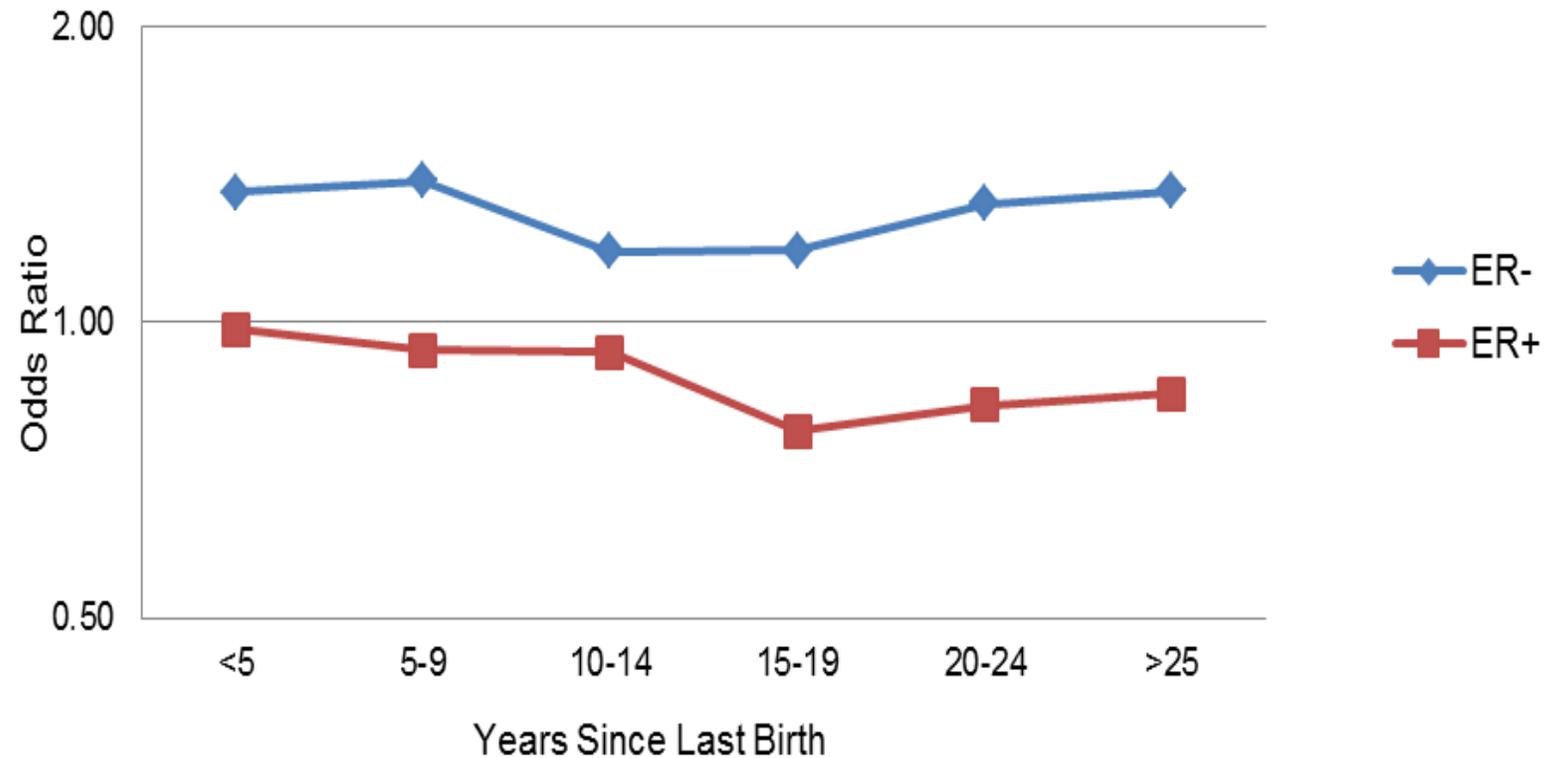


Reference category is uniparous women who breastfed

# AMBER: Parity and breast cancer, by age at first birth

ER Subtype	Age first birth <25 years		Age first birth ≥25 years	
	Case/Control	OR (95% CI)	Case/Control	OR (95% CI)
<b>ER+</b>				
Nulliparous	444/2,700	1.00	444/2,700	1.00
Parous	1,400/8,209	0.83 (0.73-0.94)	577/3,031	1.03 (0.89-1.19)
<b>ER-</b>				
Nulliparous	170/2,700	1.00	170/2,700	1.00
Parous	810/8,209	1.33 (1.10-1.60)	266/3,031	1.33 (1.07-1.65)

# AMBER: Parity and breast cancer, by time since last birth



Multivariable odds ratios for parous relative to nulliparous

# Breast Cancer Family Registry: Parity, breastfeeding, and ER subtypes

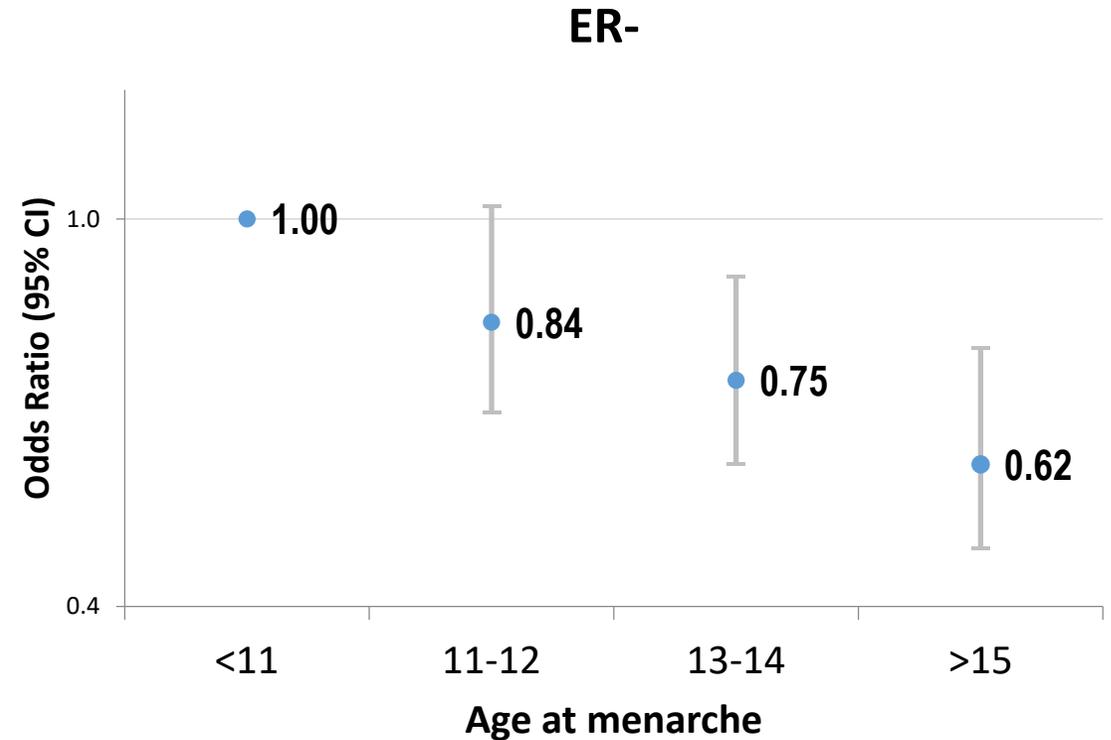
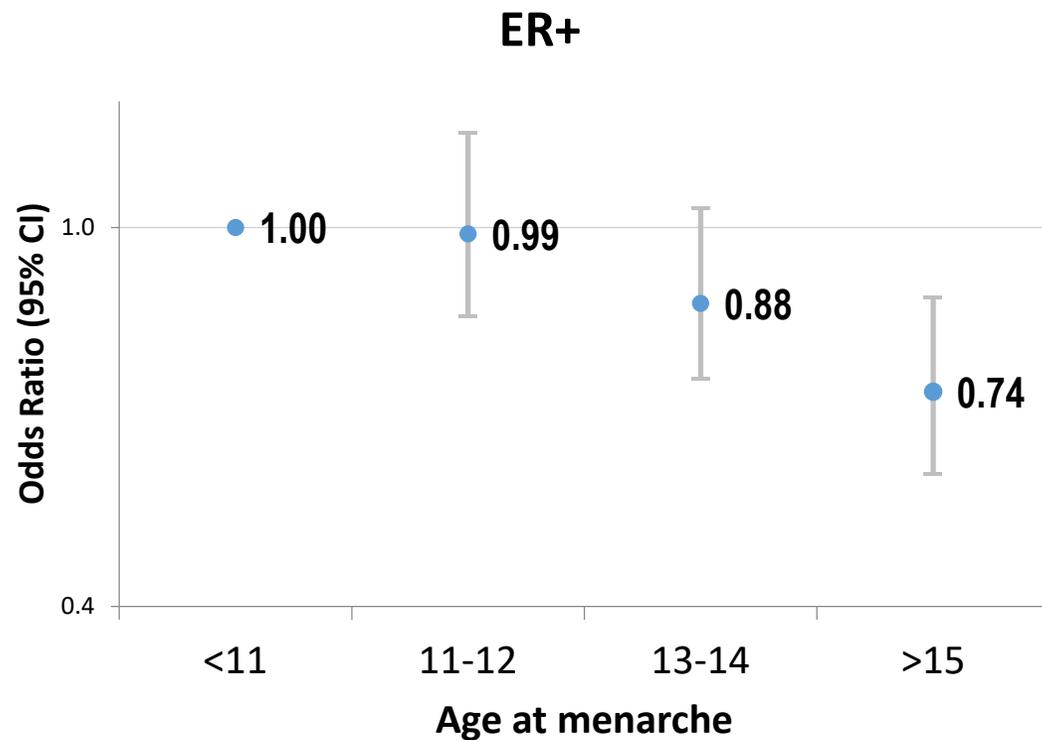
Categories	ER+PR+	ER-PR-
<b>Parity</b>		
Nulliparous	1.0 (ref)	1.0 (ref)
1-2	0.80 (0.65-0.99)	1.33 (1.00-1.76)
≥3	0.93 (0.73-1.17)	1.59 (1.15-2.18)
<b>Breastfeeding duration</b>		
Never	1.0 (ref)	1.0 (ref)
<12	1.04 (0.87-1.23)	0.72 (0.57-0.91)
≥12	0.80 (0.66-0.98)	0.52 (0.40-0.68)
<b>Parity and breastfeeding (BF)</b>		
Nulliparous	1.0 (ref)	1.0 (ref)
1-2 live births, never BF	0.80 (0.63-1.00)	1.30 (0.96-1.75)
≥3 live births, never BF	0.90 (0.68-1.19)	1.57 (1.10-2.24)
1-2 live births, ever BF	0.78 (0.64-0.93)	0.88 (0.68-1.14)
≥3 live births, ever BF	0.82 (0.67-0.99)	0.93 (0.71-1.22)

# Nurses' Health Studies: Parity, breastfeeding, and ER subtypes

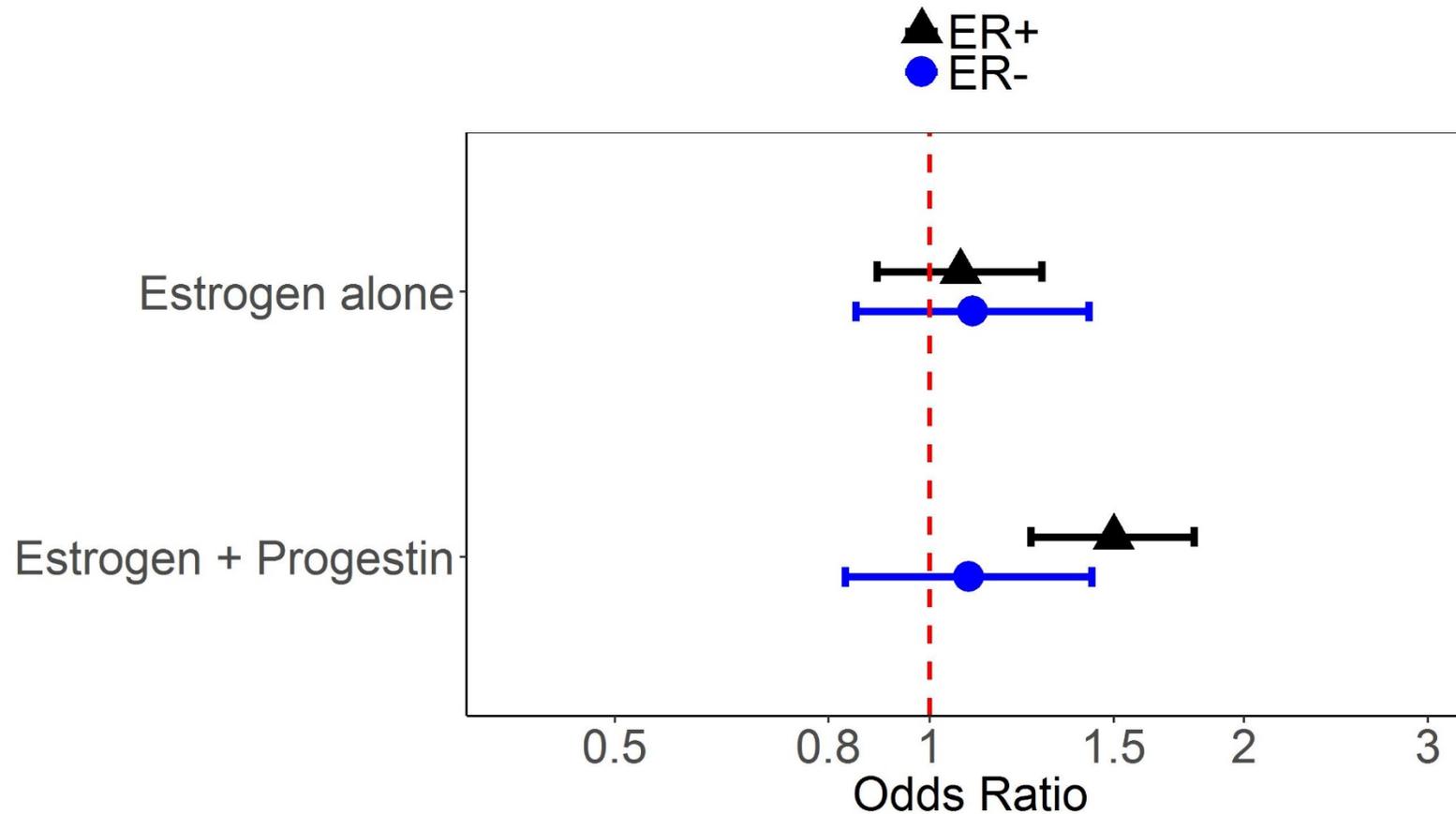
---

Categories	ER+	ER-
Parity		
Nulliparous	1.0 (ref)	1.0 (ref)
Parous	0.82 (0.77-0.88)	0.98 (0.84-1.13)
Breastfeeding		
Never breastfed	1.0 (ref)	1.0 (ref)
Ever breastfed	0.99 (0.94-1.04)	0.83 (0.75-0.92)
Parity/breastfeeding		
Nulliparous	1.0 (ref)	1.0 (ref)
Parous, never breastfed	0.83 (0.77-0.90)	1.11 (0.94-1.31)
Parous, ever breastfed	0.82 (0.76-0.88)	0.92 (0.79-1.08)

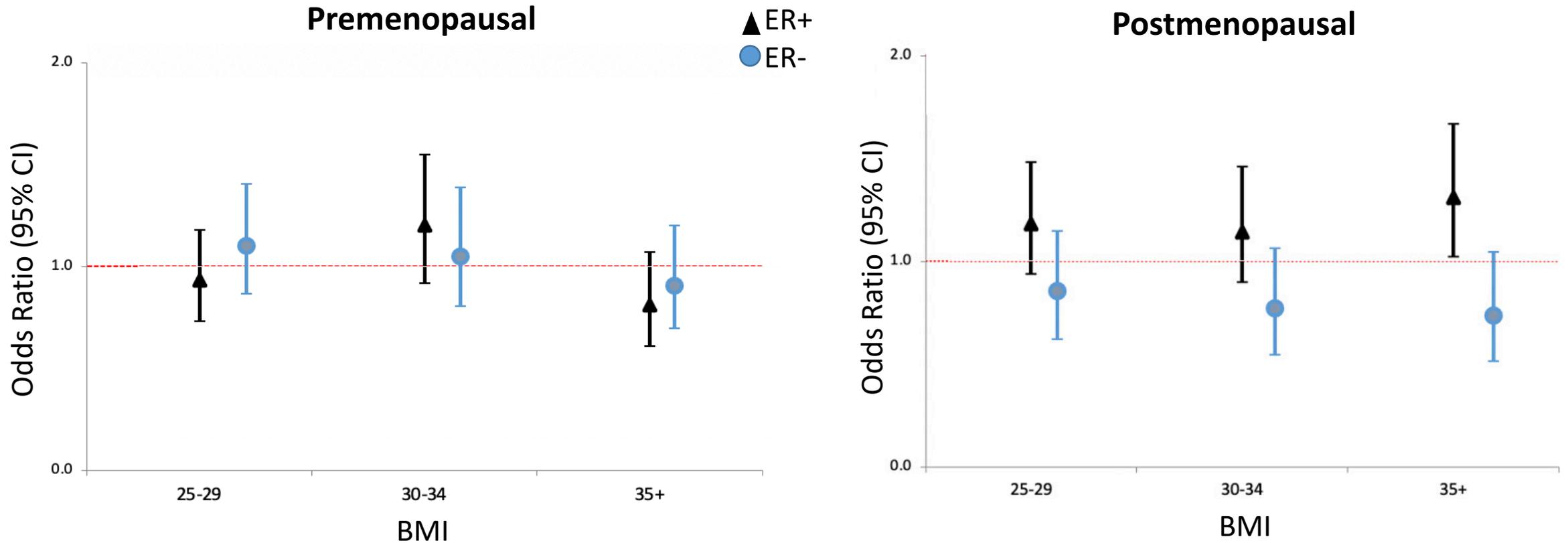
# AMBER: Age at menarche in relation to breast cancer risk, by ER subtype



# AMBER: Hormone supplement use and breast cancer risk in postmenopausal women, by ER subtype



# AMBER: Body mass index (BMI) in relation to breast cancer risk, by ER subtype and menopausal status



# AMBER: Family history of cancer and breast cancer risk

---

First degree family history	ER+	ER-
None	1.00 (reference)	1.00 (reference)
Breast cancer only	1.62 (1.39-1.89)	1.50 (1.21-1.86)
Prostate cancer only	1.24 (1.04-1.48)	0.98 (0.75-1.28)
Breast and prostate	3.40 (2.42-4.79)	2.09 (1.21-3.63)

# What goes into a cancer risk prediction model?

---

- Stable estimates of relative risk for factors related to the cancer

# Risk factors for ER+ and ER- breast cancer: Results from the AMBER consortium

---

<b>Risk Factors</b>	<b>ER+ (95% CI)</b>	<b>ER- (95% CI)</b>
Ever parous	<b>0.92</b> (0.81-1.03)	<b>1.33</b> (1.11-1.59)
Ever breastfed (among all parous)	1.04 (0.91-1.18)	<b>0.81</b> (0.69-0.95)
Estrogen + progestin supplements		
Ever use	<b>1.50</b> (1.25-1.79)	1.09 (0.83-1.43)
Recent use, <5 years ago	<b>1.55</b> (1.21-1.99)	1.04 (0.73-1.47)
Obesity (postmenopausal)	<b>1.31</b> (1.02-1.67)	0.75 (0.54-1.04)
1 <sup>st</sup> degree family hx breast ca	1.76 (1.57-1.97)	1.67 (1.42-1.95)
1 <sup>st</sup> degree family hx prostate ca	<b>1.24 (1.04-1.48)</b>	0.98 (0.75-1.28)

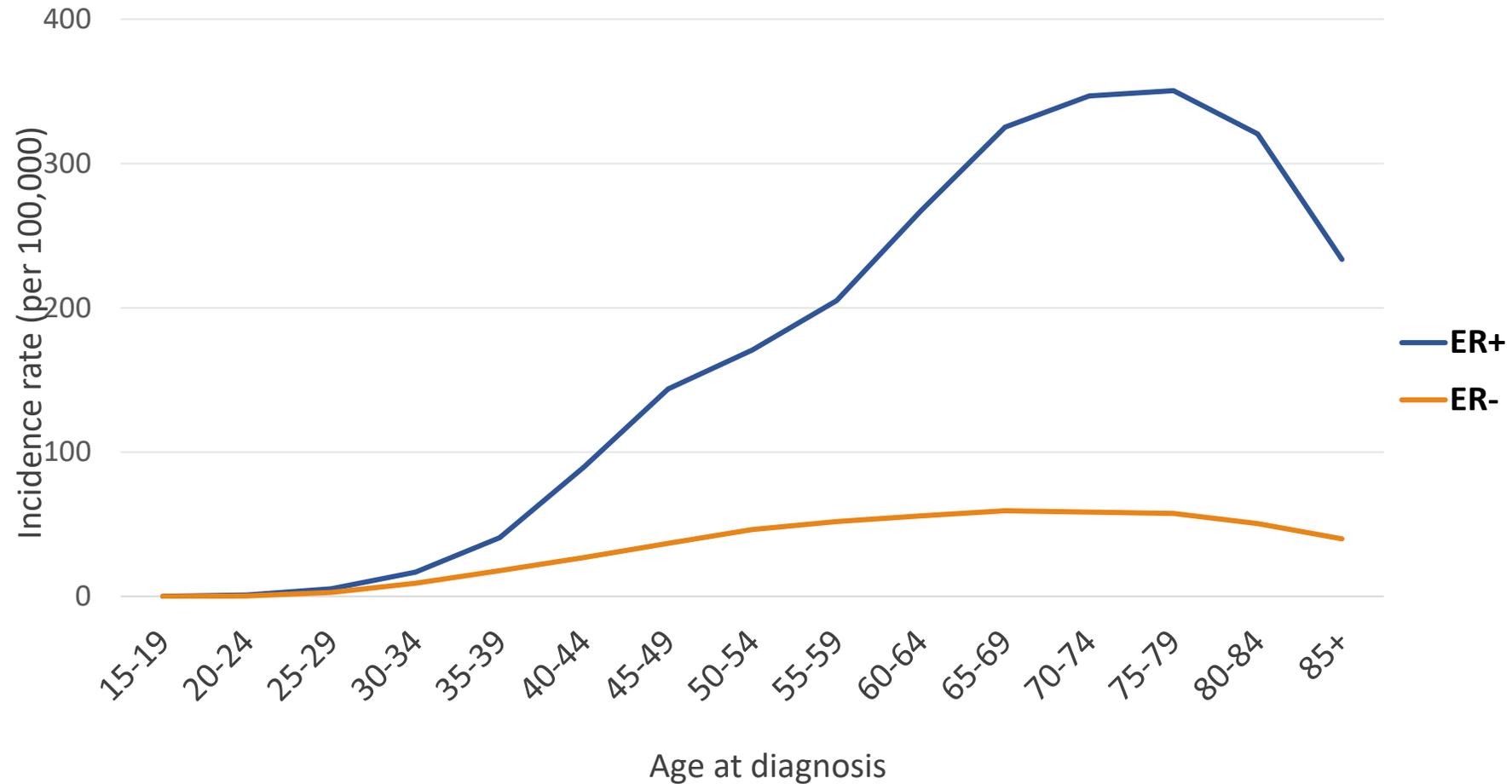
---

# What goes into a cancer risk prediction model?

---

- Stable estimates of relative risk for factors related to the cancer
- **Age-specific cancer incidence rates**

# Age-specific incidence of female breast cancer by ER status, 2000-2017



# Hypothesis

---

Poor performance of breast cancer risk prediction models in Black women may in part be due to considering breast cancer as a single disease

A risk prediction model that takes into account ER-specific risk factors and the differing age-incidence patterns of ER+ and ER- breast cancer may have improved discriminatory accuracy

# Methods of model development

---

- Develop 2 separate relative risk models  $rr_i(x)$  for  $i = ER_+, ER_-$  breast cancer based on case-control data
  - ✓ Variable selection by backward elimination with Akaike's Information Criterion
  - ✓ Multiple imputation to handle missing covariate values (50 imputed datasets)
- Estimate attributable risk  $AR_i$  from relative risk and risk factor distribution in cases

$$AR_i = 1 - \frac{\sum \text{cases } rr_i^{-1}}{\# \text{ cases}}$$

# Breast cancer hazard

---

- Obtain composite incidence rates  $h_i^*$  for  $i = ER_+, ER_-$  breast cancer from SEER data for non-Hispanic Black women

- Compute each age-specific **baseline hazard** rate from attributable risks
$$h_i(t) = h_i^*(1 - AR_i) \quad i = ER_+, ER_-$$

- Compute each hazard as

$$h_i(t, x) = h_i(t) rr_i(x) \quad i = ER_+, ER_-$$

with  $rr_i(x)$  **relative risk** part including risk factors  $x$

- Finally, compute hazard for breast cancer as sum of hazards for ER+ and ER- cancer

$$h_1(t, x) = h_{ER+}(t, x) + h_{ER-}(t, x)$$

# Calculation of absolute risk of breast cancer

---

Estimate 5-year absolute risk of **any** breast cancer for a given risk profile, while accounting for competing risk of death from causes other than breast cancer:

$$r(x, a, \tau) = P(T \leq a + \tau | T > a; x) = \int_a^{a+\tau} h_1(t, x) \exp \left[ - \int_a^t \{h_1(t, x) + h_2(u)\} du \right] dt$$

$x$ : individual risk or protective factors

$a$  : age at start of projection

$\tau$  : 5 years, the length of projection

$h_1(t, x)$  : breast cancer hazard at age  $t$

$h_{death}(t)$  : mortality hazard at age  $t$  derived from CDC WONDER

# Breast cancer cases available for development and testing of a risk prediction model for breast cancer in Black women

---

Available Cases	ER+	ER-
Model development		
Women's CARE Study	736	579
Carolina Breast Cancer Study	405	401
Women's Circle of Health	1,134	402
<b>Total for model development</b>	<b>2,275</b>	<b>1,382</b>
Model testing		
<b>BWHS prospective data, 2000-2015</b>	<b>1,302</b>	<b>623</b>

# Potential benefits of the models that will be developed

---

## **Model that predicts absolute risk of any breast cancer**

- High risk young women can be referred for screening before they reach the guideline-recommended ages
- Improved accuracy of risk prediction for women of all ages

## **ER+ risk prediction model**

- Determine risk/benefit ratio of taking Tamoxifen or other chemopreventives
- Determine eligibility for prevention trials of drugs targeting hormone receptors

## **ER- risk prediction model**

- Determine eligibility for prevention trials targeting molecular pathways implicated in ER- or triple negative breast cancer

# Why not add a polygenic risk score (PRS)?

---

- PRS combines the effects of many SNPs, weighting by the relative effect of each
- Appears to be independent of epidemiologic risk factors; thus adding a PRS may increase discriminatory accuracy of a breast cancer risk prediction model
- Research in populations of European ancestry women have demonstrated utility of adding PRS to risk prediction models

# PRS for breast cancer in women of European ancestry

---

Pooled data from 69 studies

- 94,075 breast cancer cases
- 75,017 controls

PRS based on 313 SNPs

- AUC = 0.630 (0.628-0.651)
- OR per 1 standard deviation of PRS = **1.61 (1.57-1.65)**

# PRS for breast cancer in women of African ancestry

---

## Women's Health Initiative cohort

- 421 breast cancer cases among AA participants
- PRS based on 75 SNPs from GWAS of European ancestry women
- OR per one standard deviation of PRS = **1.24 (1.12-1.37)**

## ROOT Consortium (Nigeria, USA, Barbados)

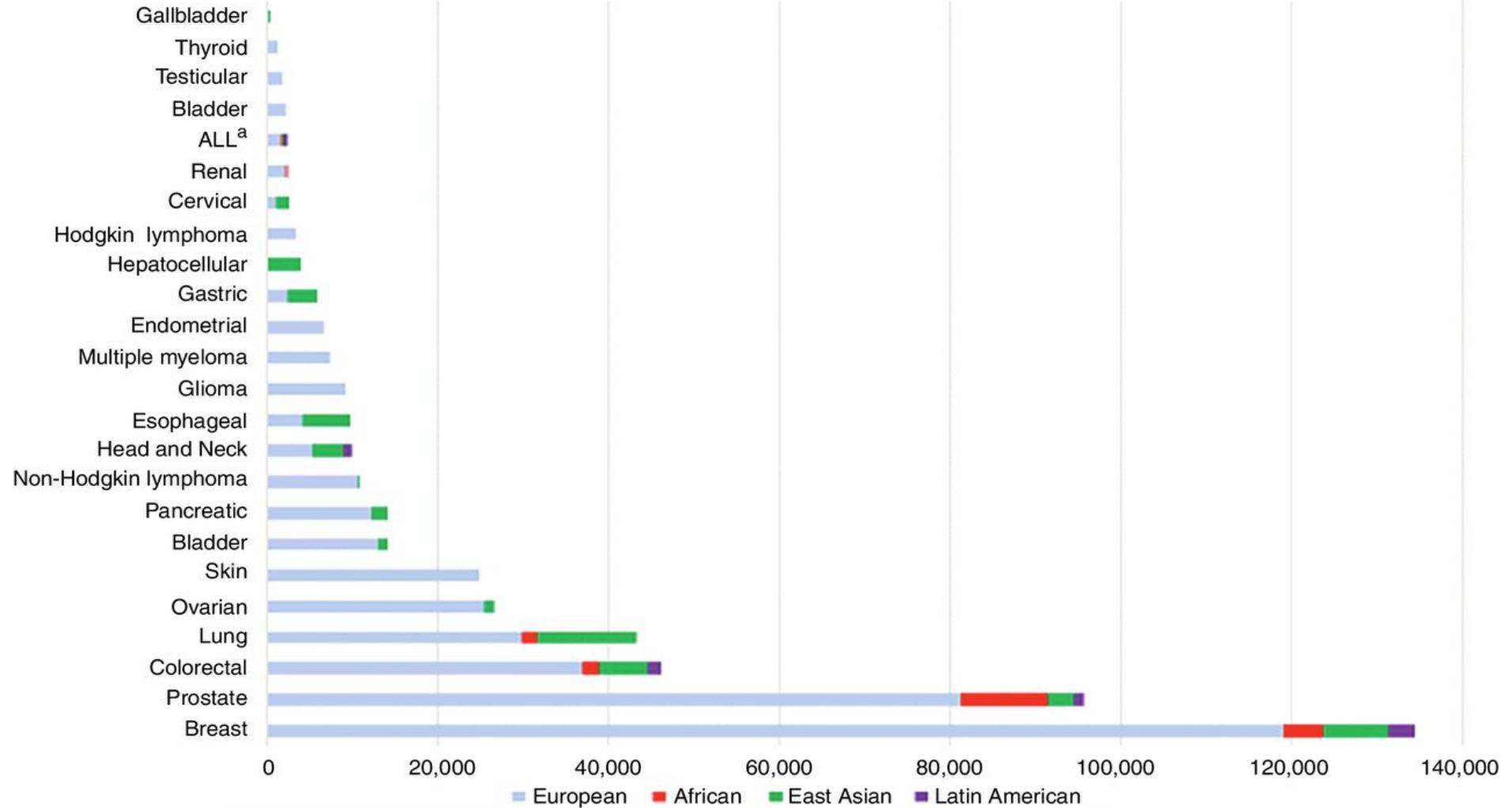
- 1657 breast cancer cases, 2029 controls
- PRS based on 34 SNPs from GWAS of women of European and Asian ancestry
- AUC = 0.53 (0.51-0.55)
- OR per 1 standard deviation of PRS = **1.13 (1.06-1.20)**

# Disparities in PRS accuracy

---

- Greater genetic diversity in populations in Africa or recently emerged from Africa, with different patterns of linkage disequilibrium
- Overwhelming abundance of European descent studies and dearth of well-powered discovery studies in globally diverse populations

# Number of cancer cases included in discovery stage of GWAS by ancestral population and cancer site



# Summary (1)

---

- Risk prediction models have a variety of uses, including to identify high risk individuals and aid in clinical decision-making
- Prediction models have historically not worked as well among Black women as among white women
- ER- breast cancer risk factors differ somewhat from risk factors for ER+ positive breast cancer, as do age-incidence curves for ER- and ER+ cancer

## Summary (2)

---

- Risk prediction models for breast cancer in Black women may have better performance if developed based on considering separate RRs and incidence rates for ER+ and ER- disease
- Sufficient data are now available to use this approach
- At a minimum, developing a model based on the three largest case-control studies and testing in 15 years of prospective data from the Black Women's Health Study has potential for improved model

# Acknowledgements

---

## **Boston University**

Lynn Rosenberg

Ludovic Trinquart

Gary Zirpoli

Kimberly Bertrand

Kirsten Woolpert

## **National Cancer Institute**

Ruth Pfeiffer

## **Roswell Park Cancer Institute**

Christine Ambrosone

## **University of North Carolina**

Melissa Troester

## **City of Hope Cancer Center**

Leslie Bernstein

National Institutes of Health grants R01CA228357, R01CA058420, U01CA164974, P01CA151135

Susan G. Komen Foundation SAC180086

Karin Grunebaum Cancer Research Foundation