Understanding Breast Cancer Inequities Using Real World Data Mapped to the OMOP CDM from Pakistan and the United Kingdom

Rabia Ali Khan 1, Muhammad Usman Afzal 2, Faisal Jamil 2, Awais Tauseef 2, Marta Pineda Moncusi 1, Danielle Newby 1, Haroon Hafeez 2, Sara Khalid 1

1. Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK

2. Shaukat Khanum Memorial Cancer Hospital and Research Centre, Lahore, Pakistan

Background

Advancements in breast cancer (BC) diagnosis and treatment have notably improved survival in high-income countries. However, patients (pts) in low- and middle-income countries (LMICs) continue to face substantial disparities in treatment and clinical outcomes.

Objectives

1. Develop a BC phenotype suitable for use with databases mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) from diverse settings
2. Characterize pts with BC in Pakistan and the United Kingdom (UK)
3. Estimate crude annual incidence rate (IR) of BC in Pakistan and the UK

Conclusion

We demonstrate the feasibility of applying a BC phenotype to two international, heterogeneous real world data sources i.e **Shaukat Khanum Memorial Cancer Hospital and Research Centre (*SKMCH&RC*)** database, a cancer hospital database from Pakistan, and **The Health Improvement Network UK (*THIN® UK*),** a primary care database from the UK, using the OMOP-CDM to identify and analyse pts with BC.

Given the population coverage differences between the two databases, significant disparities were observed between the two BC cohorts. Pts in Pakistan presented at a substantially younger age at diagnosis. Between 2014-2022, SKMCH&RC data reported considerable year-to-year variability in crude annual IR (64.51–224.98), whereas the THIN® UK database demonstrated a more consistent trend in BC diagnoses (88.49–140.28).

Results

Creating BC Phenotype

- 5,997 concept IDs reviewed manually by RAK
- **1,154** concept IDs included in the finalized BC phenotype

Characterization

	SKMCH&RC	THIN® UK
Pts in Database	10,031,718	10,977,326
Pts with BC	29,714	63,621
Sex (% Female)	29,481 (99.22%)	62,957 (98.96%)
Median Age at Diagnosis (Years [IQR])	45 [38 - 54]	65 [54 - 75]
Age Groups (Years)		
0-17	6 (0.02%)	31 (0.05%)
18-64	27,518 (92.61%)	31,365 (49.30%)
65+	2,190 (7.37%)	32,225 (50.65%)
Top 5 Medications Prescribed at Any Time Following Diagnosis of BC	Metoclopramide* (80.56%); Omeprazole* (77.74%); Acetaminophen* (76.44%); Ondansetron* (74.81%); Dexamethasone* (74.70%)	Omeprazole* (33.53%); Letrozole** (31.42%); Acetaminophen* (30.31%); Amoxicillin* (29.51%); Tamoxifen** (26.69%)

* Supportive Medication; ** Therapeutic Medication

Trends in Crude Annual Incidence Rate of BC

- SKMCH&RC: Highly variable trend in crude IR, ranging from 64.51 per 100,000 person-years in 2017 to a peak of 224.98 per 100,000 person-years in 2015.
- THIN® UK: More stable trends, with IR ranging from 88.49 per 100,000 person-years in 2019 to a maximum of140.28 per 100,000 person-years in 2015.

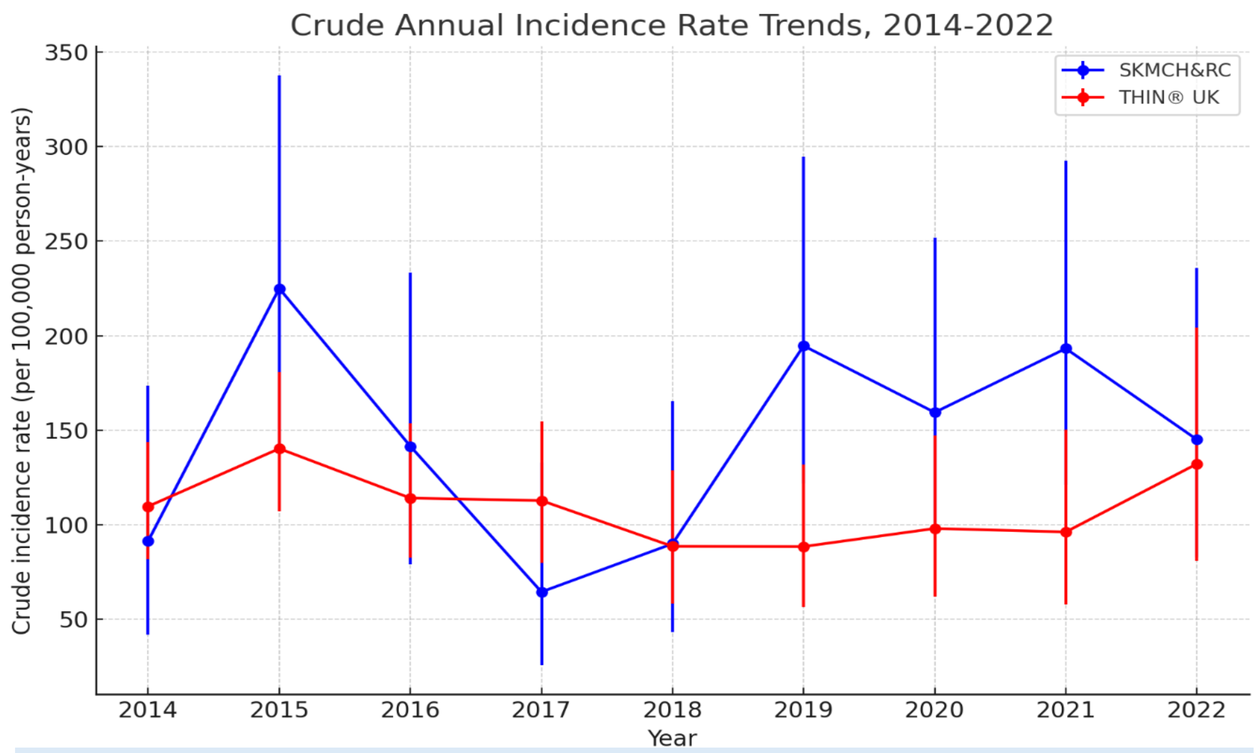

Figure 1. Crude Annual Incidence Rate Trends for Breast Cancer in SKMCH&RC and THIN® UK, 2014–2022

Methods

Developing the BC Phenotype

- *CodelistGenerator* package (keywords: “malignant neoplasm of breast", "breast carcinoma" and "breast cancer”)[1]
  - Manually reviewed code list and verified with two independent clinicians to finalize the phenotype

Characterization Analysis

Data Sources

1. SKMCH&RC (Cancer hospital database, Pakistan) [Start Date: 1994- 12-29; End Date: 2025-04-18]
2. THIN® UK (Primary care database, UK) [Start Date: 2011-01-01; End Date: 2023-01-01]

- Characterise pts using *PhenotypeR* package [2]
- Pts ≥ 18 years with a diagnosis of **incident primary BC**, and with at least **one year** of available clinical history prior to diagnosis, were included
- Pts were followed up until either database exit, death, or the study end date

Estimating Crude Annual Incidence Rate

- Crude annual IR (95% confidence interval) per 100,000 person-years were calculated for patients with BC for the time period **2014-01-01 to 2022-12-31**

References

1. C. X. Burn E, Mercade-Besora N "CodelistGenerator*: Identify Relevant Clinical Codes and Evaluate Their Use. R package version* 3.5.0, ." https://darwin-eu.github.io/CodelistGenerator/ (accessed 2025).
2. C. M. Burn E, Chen X, Alcalde-Herraiz M, Prats-Uribe A "PhenotypeR*: Assess Study Cohorts Using a Common Data Model. R package version 0.1.4.*" https://ohdsi.github.io/PhenotypeR/ (accessed 2025).