Reproductive history differs by molecular subtypes of breast cancer among women aged ≤50 years in Scotland in 2009-16: A population-based retrospective cohort study

Anushri Chitkara
University of Edinburgh Division of Reproductive and Developmental Sciences: The University of Edinburgh Division of Health Sciences

Ines Mesa-Eguiagaray
University of Edinburgh

Sarah H Wild
University of Edinburgh

Peter S Hall
University of Edinburgh

David Cameron
University of Edinburgh

Jonine Figueroa (jonine.figueroa@ed.ac.uk)
University of Edinburgh https://orcid.org/0000-0002-5100-623X

Research Article

Keywords: Breast cancer, molecular subtypes, reproductive factors, ER status, HER2, parity, age at first birth, time since last birth

Posted Date: September 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-940729/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background

The aetiology of breast cancers diagnosed ≤ 50 years of age remains unclear. We aimed to compare reproductive risk factors between molecular subtypes of breast cancer thereby suggesting possible aetiologic clues, using routinely collected cancer registry and maternity data in Scotland.

Methods

We conducted a population-based retrospective cohort study of 4,108 women aged ≤ 50 years with primary breast cancer diagnosed between 2009–2016 linked to maternity data. Molecular subtypes of breast cancer were defined using immunohistochemistry (IHC) tumour markers, oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER2), and tumour grade. Age-adjusted logistic regression models were used to estimate odds ratios (OR) and 95% confidence intervals (CI) for the association of number of births, age at first birth and time since last birth with IHC-defined breast cancer subtypes: luminal A-like, luminal B-like (HER2-), luminal B-like (HER2+), HER2-overexpressed and triple negative breast cancer (TNBC).

Results

Mean (SD) for number of births, age at first birth and time since last birth was 1.4 (1.2) births, 27.2 (6.1) years and 11.0 (6.8) years, respectively. Luminal A-like was the most common subtype (40%), while HER2-overexpressed and TNBC represented 5% and 15% of cases, respectively. Larger numbers of births were recorded among women with TNBC (p-trend = 0.05) and HER2-overexpressed (p-trend = 0.03) compared with luminal A-like tumours. Women with their most recent birth > 10 years prior to being diagnosed were less likely to have TNBC or luminal B-like (HER2+) tumours compared to luminal A-like (p-trend = 6X10^-6 and 0.01, respectively. We found limited evidence for differences by subtype with age at first birth.

Conclusion

Number of births and time since last birth differed by subtypes of breast cancer among women aged ≤ 50 years. Analyses using linked routine electronic medical records by molecularly defined tumour pathology data can be used to investigate the aetiology and prognosis of cancer.

Introduction

Breast cancer is the most common malignancy worldwide [1]. In Scotland, it constitutes 28.1% of all cancers, with 1 in every 8 women carrying a risk of developing it in her lifetime [2].
Breast cancer has been classified into ‘intrinsic’ or molecular subtypes based on mRNA expression profiling that have different treatment and survival outcomes [3]. The characteristics of these molecular subtypes are largely distinguished by expression of various combinations of tumour markers such as oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER2) and Ki67 tumour proliferation marker. Although gene profiling is considered the gold standard for classification of molecular subtypes, given the cost and lack of genetic profiling in clinical practice, a similar classification defined by immunohistochemistry (IHC) staining is a well-accepted surrogate [4, 5]. The St. Gallen Expert Panel recommends using ER, PR and HER2, along with tumour grade as a proxy for Ki67 index in defining the subtypes when the latter is unknown [5]. Based on IHC characterization, the molecular subtypes are: Luminal A-like, luminal B-like (HER2-), luminal B-like (HER2+), HER2-enriched and triple negative breast cancer (TNBC). As the luminal-like cancers (ER/PR+) express hormone receptors, they can be effectively treated with molecularly targeted hormone therapy and generally have better prognosis. Due to the absence of therapeutic targets i.e., ER/PR or HER2 in TNBC, the most aggressive subtype, chemotherapy is the only current treatment option [6, 7].

Reproductive factors have been well documented as key breast cancer risk factors with direct associations observed with early age at menarche, nulliparity, late age at menopause and first birth, and limited breastfeeding [8, 9]. Data also suggest that there is a temporal relationship with time since last birth, where a short-term increase in breast cancer risk is observed 3–5 years after last birth [10, 11], before a long-term protective effect of parity is observed compared to nulliparity.

Within Scotland’s renowned, high-quality routine electronic health records, the Scottish Cancer Registry (SMR06) is an excellent resource to investigate risk factors for cancer incidence. In Scotland, ER status data collection began in 1997, and PR and HER2 data collection started in 2009, almost a decade earlier than other registries in the UK. We have recently reported the high quality of these data and shown distinct temporal trends by molecular subtypes and observed increasing incidence of ER+ subtypes among women of screening age (50–70 years), among whom about half of all cases are diagnosed [12].

In this study, we aimed to assess whether there are differences in reproductive risk factors among invasive breast cancer cases diagnosed in Scotland using a ‘case-case’ approach. A case-case analysis compares the risk factor associations of breast cancer by comparing cases of a certain molecular subtype to cases of another subtype, without also describing risk factor patterns in women without breast cancer [13].

**Methods**

**Data sources and study population**

The Information Services Division (ISD) of Public Health Scotland holds population-level National Health Service (NHS) data for Scotland which can be deterministically linked using the Community Health Index (CHI) number, a unique patient identifier, with additional probabilistic linkage providing <4% false positive
and < 2% false negative linkage [14]. Incident primary breast cancer cases were identified using data from the Scottish Cancer Registry [15] which attains an average of 95.4% breast cancer case ascertainment and is over 99% complete [16, 17]. All tumours diagnosed in women 20+ years of age, with a primary invasive breast cancer (defined on the basis of the International Classification of Diseases, 10th revision code of C50) between 1997 and 2016 were ascertained [17, 18]. Approval for the analysis was obtained from the Public Benefit and Privacy Panel (PBPP) of NHS Scotland, and analyses were conducted in the Scottish National Safe Haven (PBPP reference number 1718-0057).

Maternity data

CHI number and probabilistic matching were used to link cancer registry data (SMR06) to Scottish Morbidity Records maternity inpatient and day case records (SMR02) which was available from 1981. To improve completeness of maternity data, the study excluded women who were ≥ 16 years (i.e., already in their reproductive years) in 1981, resulting in a cohort of women born in 1966 or thereafter. Data on number of births, age at first birth and time since last birth, including both live births and stillbirths, were calculated. The number of births was derived from the number of maternity records each woman held in SMR02. The maternal age from the first maternity record for a parous woman was considered as her age at first birth. Time since last birth was calculated as the time from the most recent birth preceding a cancer diagnosis.

Molecular subtypes definition

The Scottish Cancer Registry (SMR06) records the receptor status for breast cancers using immunohistochemistry (IHC) staining for ER, PR and HER2, and for borderline IHC HER2 results the status based on fluorescence in situ hybridization [19]. While ER status for breast cancer became available in SMR06 in 1997, recording of information on PR and HER2 status commenced only in 2009 [19]. As we aimed to evaluate the subtypes based on ER, PR and HER2 status, we focused on cases diagnosed from 2009 onwards. Due to non-availability of data on Ki67 labelling index, tumour grade was employed as a proxy for distinguishing the luminal subtypes [5]. The outcome variable, breast cancer subtype, was derived from four variables in SMR06: ER status, PR status, HER2 status and histological grade of the tumour. The five subtypes were defined as: ‘luminal A-like’ [ER/PR + HER2- grade 1 or 2], ‘luminal B-like (HER2-)’ [ER/PR + HER2- grade 3], ‘luminal B-like (HER2+)’ [ER/PR + HER2+], ‘HER2-overexpressed’ [ER-PR-HER2+], and ‘triple negative breast cancer’ or ‘TNBC’ [ER-PR-HER2-]. SMR02 and SMR06 datasets were linked by ISD using a pseudonymised CHI.

The cohort was limited to women with complete data on IHC-defined molecular marker status and tumour grade. Further restricting to women born in 1966 or later and with a breast cancer diagnosis between 2009–2016, resulted in a cohort of women diagnosed at 50 years of age or younger.

Statistical analyses

Age distribution at diagnosis of breast cancer, number of births, age at first birth and time since last birth were computed for each breast cancer subtype. Pearson's chi-square tests were used to test for
differences between subtypes in the distribution of reproductive risk factors of interest. Logistic regression models adjusted for age at diagnosis of breast cancer were used to estimate odds ratios (OR) and 95% confidence intervals (CI) with the most common subtype, luminal A-like, as the reference group. The models were run separately for each reproductive risk factor of interest. Chi-squared tests for trend were performed. In order to avoid introduction of redundancy in the statistical models, absence of collinearity was confirmed between age at diagnosis and each reproductive risk factor by computing Spearman's correlation coefficients [20]. Tests were considered statistically significant at the 5% level. R version 3.6.0 [21] was used for all analyses.

Results

The final study population included 4,108 women with breast cancer diagnosed at or below 50 years of age with data available to assign breast cancer subtype, after excluding 9.7% of the initial cohort with missing hormone status or tumour grade data (Supplementary table S1). Luminal A-like was the most common type (40%) and HER2-overexpressed was the least common (5%, Fig. 1).

Distribution of age at diagnosis of breast cancer, number of births, age at first birth and time since last birth by the five breast cancer subtypes are presented in Table 1. Overall, 34% of breast cancers occurred in patients of age 40 years or younger years of age and 66% in those between 41–50 years of age. The proportion of all luminal A-like tumours diagnosed in the age group 41–50 years was 76% as compared to around 60% of the other subtypes being diagnosed in this age group. Women with luminal A-like subtype had the highest proportion of absence of birth records (assumed nulliparity of 31%) and breast cancer diagnoses that were six or more years following their most recent birth (83%), while women with HER2-overexpressed and TNBC had the highest proportions of one or more birth records (79% and 76%, respectively) and diagnoses made within six years after last birth (30% and 31%, respectively). Chi-square test revealed no statistically significant differences for age at first birth by subtype (Table 1). A significant correlation between age at diagnosis and time since last birth was observed (Spearman $R^2 = 0.63$) and no correlation for number of births or age at first birth with age at diagnosis (Spearman $R^2 < 0.05$).
Table 1
Descriptive characteristics of the cohort comprising of women born after 1965 and diagnosed with primary invasive breast cancer between 2009 and 2016 in Scotland stratified by surrogate molecular subtypes

|                      | Luminal A-like (HER2-) | Luminal B-like (HER2-) | Luminal B-like (HER2+) | HER2-overexpressed | TNBC | Overall |
|----------------------|------------------------|------------------------|------------------------|-------------------|------|---------|
| Total number of cases (%) | 1,650 (40.2)           | 998 (24.3)             | 629 (15.3)             | 214 (5.2)         | 617  | 4,108   |

|                      | N (%) | N (%) | N (%) | N (%) | N (%) | N (%) |
|----------------------|-------|-------|-------|-------|-------|-------|
| Age at diagnosis (years) |       |       |       |       |       |       |
| ≤ 40                 | 405 (24.5) | 372 (37.3) | 257 (40.9) | 92 (43.0) | 268 | 1,394   |
| 41–50                | 1,245 (75.5) | 626 (62.7) | 372 (59.1) | 122 (57.0) | 349 | 2,714   |

*P value for heterogeneity between subtypes using chi-square test: <2.2e-16

| Number of births | 0 | 1 | 2 | ≥ 3 |
|------------------|---|---|---|-----|
| 0                | 504 (30.5) | 291 (29.2) | 189 (30.0) | 45 (21.0) | 150 | 1,179   |
| 1                | 370 (22.4) | 215 (21.5) | 151 (24.0) | 53 (24.8) | 153 | 942     |
| 2                | 524 (31.8) | 337 (33.8) | 202 (32.1) | 77 (36.0) | 215 | 1,355   |
| ≥ 3               | 252 (15.3) | 155 (15.5) | 87 (13.8)  | 39 (18.2) | 99  | 632     |

*P value for heterogeneity between subtypes using chi-square test: 0.111

| Number of parous cases (%) | 1,146 (39.1) | 707 (24.1) | 440 (15.0) | 169 (5.8) | 467 (15.9) | 2,929 (100.0) |

*Significant at 5% level

**Total counts (N = 2,908) exclude parous women that did not have a pregnancy before diagnosis of breast cancer
| Age at first birth (years) | Luminal A-like (HER2-) | Luminal B-like (HER2-) | Luminal B-like (HER2+) | HER2-overexpressed | TNBC | Overall |
|---------------------------|------------------------|------------------------|------------------------|--------------------|------|---------|
| < 20                      | 144 (12.5)             | 89 (12.6)              | 44 (10.0)              | 18 (10.7)          | 73 (15.6) | 368 (12.6) |
| 20–24                     | 246 (21.5)             | 162 (22.9)             | 88 (20.0)              | 43 (25.4)          | 117 (25.1) | 656 (22.4) |
| 25–29                     | 348 (30.4)             | 195 (27.6)             | 149 (33.9)             | 49 (29.0)          | 117 (25.1) | 858 (29.3) |
| 30–34                     | 268 (23.4)             | 169 (23.9)             | 100 (22.7)             | 45 (26.6)          | 101 (21.6) | 683 (23.3) |
| ≥ 35                      | 140 (12.2)             | 92 (13.0)              | 59 (13.4)              | 14 (8.3)           | 59 (12.6) | 364 (12.4) |

*P value for heterogeneity between subtypes using chi-square test: 0.151*

| Time since last birth (years)** | Luminal A-like (HER2-) | Luminal B-like (HER2-) | Luminal B-like (HER2+) | HER2-overexpressed | TNBC | Overall |
|-------------------------------|------------------------|------------------------|------------------------|--------------------|------|---------|
| ≤ 2                           | 76 (6.7)               | 71 (10.1)              | 53 (12.1)              | 20 (11.9)          | 70 (15.1) | 290 (10.0) |
| 3–5                           | 120 (10.5)             | 116 (16.5)             | 70 (16.0)              | 30 (17.9)          | 73 (15.8) | 409 (14.1) |
| 6–10                          | 279 (24.5)             | 186 (26.5)             | 115 (26.3)             | 41 (24.4)          | 121 (26.1) | 742 (25.5) |
| > 10                          | 663 (58.3)             | 329 (46.9)             | 199 (45.5)             | 77 (45.8)          | 199 (43.0) | 1,467 (50.4) |

*P value for heterogeneity between subtypes using chi-square test: 5.61e-10*

*Significant at 5% level

**Total counts (N = 2,908) exclude parous women that did not have a pregnancy before diagnosis of breast cancer

women with TNBC were significantly more likely to have at least one (as against no births) in comparison to the women with luminal A-like tumours (Table 2). Although based on fewer cases, a similar association
was observed for women with HER2-overexpressed tumours who were more likely to have three or more births (relative to no birth records) when compared to women with luminal A-like tumours, in addition to a statistically significant test for trend across all sub types.

Table 2
Association of number of births among women born after 1965 and diagnosed with primary invasive breast cancer between 2009 and 2016 in Scotland by molecular subtypes adjusted for age at diagnosis

|                      | Luminal B-like (HER2-) (n = 998) | Luminal B-like (HER2+) (n = 629) | HER2-overexpressed (n = 214) | TNBC (n = 617) |
|----------------------|----------------------------------|---------------------------------|-------------------------------|---------------|
| **OR**               | 1.00 (reference)                 | 1.00 (reference)                | 1.00 (reference)              | 1.00 (reference) |
| **95% CI**           | 0.82–1.27                        | 0.86–1.43                       | 1.07–2.49                     | 1.09–1.84     |
| **Number of births**|                                  |                                 |                               |               |
| 0                    |                                  |                                 |                               |               |
| 1                    | 1.02                             | 1.10                            | 1.63                          | 1.41          |
| 2                    | 1.16                             | 1.08                            | 1.74                          | 1.46          |
| ≥ 3                  | 1.10                             | 0.96                            | 1.82                          | 1.39          |
| **P-value for trend**| 0.83                             | 0.17                            | 0.02                          | 0.05          |
| **Age at Diagnosis**|                                  |                                 |                               |               |
| **(years)**          |                                  |                                 |                               |               |
| ≤ 40                 | 1.00 (reference)                 | 1.00 (reference)                | 1.00 (reference)              | 1.00 (reference) |
| 41–50                | 0.54                             | 0.47                            | 0.42                          | 0.42          |
| **95% CI**           | 0.46–0.64                        | 0.39–0.57                       | 0.31–0.56                     | 0.34–0.51     |

*Significant at 5% level

Referent subtype used here is Luminal A-like breast cancer.

Table 3 shows case-case analysis for age at first birth by subtype. There was no evidence of an association between age at first birth and tumour subtype.
Table 3
Association of age at first birth among parous women born after 1965 and diagnosed with primary invasive breast cancer between 2009 and 2016 in Scotland by surrogate molecular subtypes adjusted for age at diagnosis

|                        | Luminal B-like (HER2-) (n = 707) | Luminal B-like (HER2+) (n = 440) | HER2-overexpressed (n = 169) | TNBC (n = 467) |
|------------------------|----------------------------------|----------------------------------|-------------------------------|---------------|
|                        | OR 95% CI                         | OR 95% CI                        | OR 95% CI                     | OR 95% CI     |
| **Age at first birth** |                                  |                                  |                               |               |
| (years)                |                                  |                                  |                               |               |
| < 20                   | 1.00 (reference)                  | 1.00 (reference)                 | 1.00 (reference)              | 1.00 (reference) |
| 20–24                  | 1.09 0.78–1.53                    | 1.20 0.79–1.82                   | 1.44 0.80–2.60                | 0.97 0.67–1.39 |
| 25–29                  | 0.90 0.65–1.24                    | 1.39 0.94–2.06                   | 1.12 0.63–1.99                | 0.66 0.46–0.94 |
| 30–34                  | 0.99 0.71–1.38                    | 1.19 0.79–1.80                   | 1.30 0.73–2.34                | 0.72 0.50–1.04 |
| ≥ 35                   | 1.17 0.80–1.71                    | 1.52 0.96–2.40                   | 0.90 0.43–1.88                | 0.94 0.62–1.43 |
| **P-value for trend*** | 0.76                             | 0.09                             | 0.64                          | 0.05          |

| **Age at Diagnosis**  |                                  |                                  |                               |               |
| (years)               |                                  |                                  |                               |               |
| ≤ 40                  | 1.00 (reference)                  | 1.00 (reference)                 | 1.00 (reference)              | 1.00 (reference) |
| 41–50                 | 0.46 0.37–0.56                    | 0.46 0.37–0.59                   | 0.41 0.29–0.58                | 0.39 0.31–0.49 |

*Significant at 5% level

Referent subtype used here is Luminal A-like breast cancer.

When compared to the luminal A-like subtype, TNBC and luminal B-like (HER2+) cases were significantly less likely to have last given birth > 10 years ago (relative to ≤ 2 years ago) with a significant linear trend being evident (Table 4). Other subtypes did not show a clear association for time since last birth.
Table 4
Association of time since most recent birth with among parous women born after 1965 and diagnosed with primary invasive breast cancer between 2009 and 2016 in Scotland who had their last birth prior to diagnosis by surrogate molecular subtypes adjusted for age at diagnosis

|                      | Luminal B-like (HER2-) (n = 702) | Luminal B-like (HER2+) (n = 437) | HER2-overexpressed (n = 168) | TNBC (n = 463) |
|----------------------|----------------------------------|----------------------------------|-----------------------------|----------------|
| **Time since last birth (years)** | OR 95% CI | OR 95% CI | OR 95% CI | OR 95% CI |
| ≤ 2                  | 1.00 (reference)                 | 1.00 (reference)                 | 1.00 (reference)            | 1.00 (reference) |
| 3–5                  | 1.15 0.76–1.75                   | 0.92 0.58–1.46                   | 1.07 0.56–2.03              | 0.74 0.47–1.15 |
| 6–10                 | 0.97 0.66–1.43                   | 0.78 0.50–1.19                   | 0.79 0.43–1.46              | 0.65 0.43–0.97 |
| > 10                 | 0.84 0.57–1.24                   | 0.65 0.42–0.99                   | 0.75 0.41–1.38              | 0.53 0.35–0.79 |
| **P-value for trend*** | 0.06                             | 0.01                             | 0.10                        | 6e-6           |

**Age at Diagnosis (years)**

|                      | Luminal B-like (HER2-) (n = 702) | Luminal B-like (HER2+) (n = 437) | HER2-overexpressed (n = 168) | TNBC (n = 463) |
|----------------------|----------------------------------|----------------------------------|-----------------------------|----------------|
| ≤ 40                 | 1.00 (reference)                 | 1.00 (reference)                 | 1.00 (reference)            | 1.00 (reference) |
| 41–50                | 0.51 0.40–0.64                   | 0.55 0.41–0.73                   | 0.46 0.31–0.69              | 0.49 0.37–0.65 |

*Significant at 5% level

Referent subtype used here is Luminal A-like breast cancer.

**Discussion**

Using Scottish cancer registry data linked to maternity health records, we show that parity, number of births and time since last birth to diagnosis of breast cancer differ by IHC-defined molecular subtypes of breast cancer among women ≤ 50 years of age at diagnosis of breast cancer. Breast cancer aetiology in younger women is not fully understood as few risk factors have been identified. Furthermore, few opportunities for early detection of breast cancer are available for younger women beyond genetic counseling for high-risk families.
Multiple reports and pooled analyses have recently evaluated IHC and mRNA expression profiling defined molecular subtypes of breast cancer and consistently show a positive association with parity for triple-negative or basal-like breast tumors [22–27]. Interestingly, significant differences in the incidence of breast cancer exist for different ethnic and racial groups that also frequently have different reproductive histories [28]. Consistent with these data, we also found evidence of heterogeneity in reproductive history across IHC-defined molecular subtypes of breast cancer in this Scottish cohort. Women with ER- tumours (HER2-overexpressed and TNBC) were more likely to have a higher number of births compared to women with luminal A-like subtype. Unlike ER- cancers, we did not observe heterogeneity in number of births between luminal B-like (HER2+) and luminal A-like, which concurs with other consortium efforts [29–32]. Time since last birth showed differential associations by subtype, where women with TNBC or luminal B-like (HER2+) were less likely than women with luminal A-like tumours to have a longer time between their most recent birth and diagnosis of breast cancer. Findings for TNBC correspond well with the existing studies [33, 34].

Parity confers a dual effect on the risk of breast cancer with an augmented risk observed in the initial years following pregnancy (3–5 years, or even up to 10–15 years) [35–37], possibly by stimulating the growth of cells that have undergone initial stages of malignant change and also due to the immunosuppressive effects of pregnancy [35, 38]. It is only subsequent to this phase that the protective effect of parity sets in [39, 40] owing to the differentiation of normal breast cells that have the potential to undergo malignant transformation. While this has been observed for ER + breast cancers (luminal A-like) [9, 29, 41], an increased risk of ER- breast cancer continues to persist even in the longer term [32, 34, 42]. Our results revealed no significant difference across subtypes for age at first birth. However, TNBC cases were more likely to have a younger age at first birth when compared to luminal A-like cases (approximately 16% versus 12.5% patients for age at first birth < 20 years). A similar, statistically significant association has been reported by other studies [29, 30, 42–44]. Luminal B-like (HER2-) cases showed no statistically significant difference from luminal A-like for either of the three risk factors of interest even though studies have reported an inverse association with number of births and a positive association with age at first birth for this subtype [45, 46].

ER- breast cancers are less likely than ER + breast cancers to be detected through screening [47], and predictive modelling of breast cancer risk has been proposed as possible solution for personalised medicine and risk stratified screening [48–50]. Modelling studies using UK data suggest such risk stratified screening approaches could reduce overdiagnosis, improve cost-effectiveness, while maintaining the benefits of screening [51].

The key strengths of our study are the high-quality longitudinal data collected within the Scottish Cancer Registry for the entire population, and the availability and high level of completeness of molecular marker and tumour grade data (10% missing data). Another strength of the study is the inclusion of women diagnosed at age 50 years or below. Although breast cancer is less common within this age range, the tumours are more aggressive with poor prognosis making it important to identify and implement effective approaches to prevention amongst this age group [52]. Moreover, breast cancer incidence appears to be
Increasing in younger age groups in recent years in Scotland [12] and other populations such as the United States [53].

Although this is one of the largest studies of breast cancer among young women, a limitation is the modest number of cases for rarer tumour subtypes, especially HER2-overexpressed (5% of all cases), potentially reducing the statistical power of analyses for these tumour subtypes. Future work including a comparison cohort of women not diagnosed with breast cancer would add further updated information about the role of reproductive history as a risk factor for breast cancer, including, in due course for whose breast cancer is diagnosed at older ages. Other limitations of our study were the potential for incomplete maternity records for women whose children were born outside Scotland, lack of availability of data for other factors such as breastfeeding as well as for a more detailed mRNA expression or mutation profiling of the cancers.

In conclusion, our data highlight the value of integrating molecular data from tumours with routinely collected health records data for understanding cancer epidemiology. There is scope for future analysis using the cancer registry linked to other datasets, including community prescription records, mammographic imaging, and primary care records, to provide more detailed information on the role and patterns of key risk factors and possible new aetiologic or prognostic factors for subtypes of breast and other cancers.

Declarations

Data availability

All data used in the present study can be accessed by submitting an application to electronic Data Research and Innovation Service (eDRIS), a part of the Information Services Division of Public Health Scotland. More information on how to request access is available at https://www.isdscotland.org/eDRIS.

Competing interests

No competing interests were disclosed.

Funding

This study was funded by Wellcome Trust grant 207800/Z/17/Z

Author contributions

This study was conceived and designed by JDF. Formal analysis was conducted by JDF, IME and AC. JDF and AC prepared the original draft. SHW, DC and IME contributed to data interpretation and critical
review of manuscript drafts. PSH and AHS edited and reviewed the drafts.

Anushri Chitkara
Roles: Conceptualization, Data Curation, Formal Analysis, Investigation, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Ines Mesa-Eguiagaray
Roles: Conceptualization, Data Curation, Formal Analysis, Investigation, Software, Writing – Review & Editing

Sarah H Wild
Roles: Investigation, Writing – Review & Editing

Peter S Hall
Roles: Writing – Review & Editing

David Cameron
Roles: Investigation, Writing – Review & Editing

Andrew H Sims
Roles: Writing – Review & Editing

Jonine D Figueroa
Roles: Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Resources, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing
Acknowledgements

We are grateful to NHS Scotland and the Information Services Division of Public Health Scotland for collecting data and conducting the linkage of Scottish Morbidity Records used for analyses in this study. We would particularly like to acknowledge Suhail Iqbal of NHS National Services Scotland for enabling access to pseudonymised data in the National Safe Haven. We dedicate this manuscript to Andy Sims who unfortunately passed away prior to the publication of this work. He was a colleague and friend and is sorely missed. His passing reinforces why we conduct research—to prevent and improve the lives of cancer patients and we will continue the work he started providing a lasting important legacy that we hope he will be proud of.

References

1. Sung H et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin 71(3):209–249
2. The Scottish Public Health Observatory (ScotPHO). Breast Cancer: key points (2021) [cited 25 September 2021; Available from: https://www.scotpho.org.uk/health-wellbeing-and-disease/cancer-breast/key-points/
3. Perou CM et al (2000) Molecular portraits of human breast tumours. Nature 406(6797):747–752
4. Cheang MCU et al (2009) Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer. J Natl Cancer Inst 101(10):736–750
5. Goldhirsch A et al (2011) Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol 22(8):1736–1747
6. Tao L et al (2016) Occurrence and outcome of de novo metastatic breast cancer by subtype in a large, diverse population. Cancer Causes Control 27(9):1127–1138
7. O'Brien KM et al (2010) Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. Clinical cancer research: an official journal of the American Association for Cancer Research 16(24):6100
8. Newcomb PA et al (2011) Late age at first full term birth is strongly associated with lobular breast cancer. Cancer 117(9):1946–1956
9. Ma H et al (2010) Use of four biomarkers to evaluate the risk of breast cancer subtypes in the women's contraceptive and reproductive experiences study. Can Res 70(2):575–587
10. Nichols HB et al (2019) Breast cancer risk after recent childbirth: A pooled analysis of 15 prospective studies. Ann Intern Med 170(1):22–30
11. Tavani A et al (1999) Risk factors for breast cancer in women under 40 years. Eur J Cancer 35(9):1361–1367
12. Mesa-Eguiagaray I et al., Distinct temporal trends in breast cancer incidence from 1997 to 2016 by molecular subtypes: a population-based study of Scottish cancer registry data. Br J Cancer, 2020
13. Martínez ME et al., *What can we learn about disease etiology from case-case analyses? Lessons from breast cancer*. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2010. 19(11): p. 2710

14. The Scottish Public Health Observatory (ScotPHO). *ISD linked database* (2016) [cited 25 September 2021; Available from: https://www.scotpho.org.uk/publications/overview-of-key-data-sources/scottish-national-data-schemes/isd-linked-database]

15. Public Health Scotland. *Scottish Cancer Registry* (2020) [cited 24 September 2021; Available from: https://www.isdscotland.org/Health-Topics/Cancer/Scottish-Cancer-Registry/]

16. Brewster DH, Stockton DL (2008) Ascertainment of breast cancer by the Scottish Cancer Registry: An assessment based on comparison with five independent breast cancer trials databases. The Breast 17(1):104–106

17. Information Services Division (ISD) Scotland. *Breast Cancer Quality Performance Indicators*. 2019; Available from: https://www.isdscotland.org/Health-Topics/Quality-Indicators/Publications/2019-04-09/2019-04-09-Breast-QPI-Report.pdf

18. Information Services Division (ISD) Scotland. *National Data Catalogue* (2021) [cited 24 September 2021; Available from: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=8

19. Information Services Division (ISD) Scotland. *National Cancer Registration Data Definitions version 14.4* (2017) [cited 24 September 2021; Available from: https://www.isdscotland.org/Health-Topics/Cancer/Scottish-Cancer-Registry/Cancer-Metadata/_docs/Cancer-Registration-Definitions-v14-4.pdf

20. Stoltzfus J (2011) Logistic Regression: A Brief Primer. Acad Emerg Med 18(10):1099–1104

21. Core Team R (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

22. Anderson KN, Schwab RB, Martinez ME (2014) Reproductive risk factors and breast cancer subtypes: a review of the literature. Breast Cancer Res Treat 144(1):1–10

23. Fortner RT et al (2019) Parity, breastfeeding, and breast cancer risk by hormone receptor status and molecular phenotype: results from the Nurses’ Health Studies. Breast Cancer Res 21(1):40

24. Gaudet MM et al (2018) Pooled Analysis of Nine Cohorts Reveals Breast Cancer Risk Factors by Tumor Molecular Subtype. Cancer Res 78(20):6011–6021

25. Ma H et al (2017) Reproductive factors and the risk of triple-negative breast cancer in white women and African-American women: a pooled analysis. Breast Cancer Res 19(1):6

26. Lambertini M et al (2016) Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. Cancer Treat Rev 49:65–76

27. John EM et al (2018) Reproductive history, breast-feeding and risk of triple negative breast cancer: The Breast Cancer Etiology in Minorities (BEM) study. Int J Cancer 142(11):2273–2285
28. Figueroa JD et al., *Reproductive factors and risk of breast cancer by tumor subtypes among Ghanaian women: A population-based case-control study*. Int J Cancer, 2020

29. Yang XR et al (2011) Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J Natl Cancer Inst 103(3):250–263

30. Brouckaert O et al (2017) Reproductive profiles and risk of breast cancer subtypes: a multi-center case-only study. Breast Cancer Res 19(1):119

31. Devi CR, Tang TS, Corbex M (2012) Incidence and risk factors for breast cancer subtypes in three distinct South-East Asian ethnic groups: Chinese, Malay and natives of Sarawak, Malaysia. Int J Cancer 131(12):2869–2877

32. Kwan ML et al., *Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors*. Breast Cancer Research, 2009. 11(3)

33. De Mulder H et al (2018) Breast cancer subtype and survival by parity and time since last birth. Breast Cancer Research Treatment 169(3):481–487

34. Trivers KF et al (2009) The epidemiology of triple-negative breast cancer, including race. Cancer Causes Control 20(7):1071–1082

35. Lambe M et al (1994) Transient increase in the risk of breast cancer after giving birth. N Engl J Med 331(1):5–9

36. Liu Q et al (2002) Transient increase in breast cancer risk after giving birth: postpartum period with the highest risk (Sweden). Cancer Causes Control 13(4):299–305

37. Albrektsen G et al (2005) Breast cancer risk by age at birth, time since birth and time intervals between births: exploring interaction effects. British journal of cancer 92(1):167–175

38. Forneatti J et al (2012) Emerging targets for the prevention of pregnancy-associated breast cancer. Taylor & Francis

39. Kelsey JL, Gammon MD, John EM (1993) Reproductive factors and breast cancer. Epidemiol Rev 15(1):36–47

40. Rosner B, Colditz GA, Willett WC (1994) Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. Am J Epidemiol 139(8):819–835

41. Setiawan VW et al (2009) Breast cancer risk factors defined by estrogen and progesterone receptor status: the multiethnic cohort study. Am J Epidemiol 169(10):1251–1259

42. Millikan RC et al (2008) Epidemiology of basal-like breast cancer. Breast Cancer Research Treatment 109(1):123–139

43. Chen L et al (2016) Reproductive Factors and Risk of Luminal, HER2-Overexpressing, and Triple-Negative Breast Cancer Among Multiethnic Women. Cancer Epidemiol Biomarkers Prev 25(9):1297–1304

44. Martinez ME et al (2013) Reproductive factors, heterogeneity, and breast tumor subtypes in women of mexican descent. Cancer Epidemiol Biomarkers Prev 22(10):1853–1861
45. Ellingjord-Dale M et al (2017) Parity, hormones and breast cancer subtypes - results from a large nested case-control study in a national screening program. Breast Cancer Res 19(1):10
46. Horn J et al (2014) Reproductive history and the risk of molecular breast cancer subtypes in a prospective study of Norwegian women. Cancer Causes Control 25(7):881–889
47. Yen AM et al (2017) Initiators and promoters for the occurrence of screen-detected breast cancer and the progression to clinically-detected interval breast cancer. J Epidemiol 27(3):98–106
48. Howell A et al (2014) Risk determination and prevention of breast cancer. Breast Cancer Res 16(5):446
49. Lee A et al (2019) BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med 21(8):1708–1718
50. Pal Choudhury P et al (2020) Comparative Validation of Breast Cancer Risk Prediction Models and Projections for Future Risk Stratification. J Natl Cancer Inst 112(3):278–285
51. Pashayan N et al (2018) Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. JAMA Oncol 4(11):1504–1510
52. Brandt J et al (2015) Age at diagnosis in relation to survival following breast cancer: a cohort study. World J Surg Oncol 13(1):33
53. Lima SM et al., Trends in Parity and Breast Cancer Incidence in US Women Younger Than 40 Years From 1935 to 2015. JAMA Network Open, 2020. 3(3)

Figures

**Figure 1**

Distribution of breast cancer subtypes defined by immunohistochemistry and tumour grade among 4,108 women born after 1965 who had breast cancer diagnosed in Scotland between 2009 and 2016. Luminal
A-like (n=1,650), Luminal B-like (HER2-) (n=998), Luminal B-like (HER2+) (n=629), HER2-overexpressed (n=214), Triple negative (n=617).