Title: Distinct temporal trends in breast cancer incidence from 1997 to 2016 by molecular subtypes: A population-based study of Scottish cancer registry data (125/150 characters)

Running title: Breast cancer incidence by molecular subtypes (42/50 characters)

Authors: Ines Mesa-Eguiagaray[1], Sarah H Wild[1], Philip S. Rosenberg[2], Sheila M Bird[1], David H Brewster[1], Peter S Hall[1,4], David Cameron[4], David Morrison[5], Jonine D Figueroa[1,4]

Author affiliations:

[1]Usher Institute, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, UK; [2]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, USA; [3]Cambridge University's MRC Biostatistics Unit, Cambridge CB2 0SR [4]Cancer Research UK Edinburgh Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK; [5]NHS National Services and University of Glasgow, Glasgow, UK

Address for correspondence:

Jonine Figueroa

Usher Institute, University of Edinburgh

Old Medical School, Teviot Place, Edinburgh, EH8 9AG

CRUK Edinburgh Centre, Institute of Genetics and Molecular Medicine

jonine.figueroa@ed.ac.uk ; +44 (0)7478467698

Abstract (Word count: 187/200)

Background: Strategies for breast cancer prevention are informed by assessing whether incidence differs by tumour biology. We describe temporal trends of breast cancer incidence by molecular subtypes in Scotland.

Methods: Population-based cancer registry data on 72,217 women diagnosed with incident primary breast cancer from 1997 to 2016 were analysed. Age-standardised rates (ASR) and age-specific incidence were estimated by tumour subtype after imputing the 8% of missing oestrogen receptor (ER) status. Joinpoint regression and age- period- cohort models were used to assess whether significant differences were observed in incidence trends by ER status.

Results: ER positive tumour incidence steadily increased particularly for women of screening age 50 to 69 years from 1997 till around 2011 (1.6%/year, 95%CI: 1.2 to 2.1). ER negative incidence decreased among all ages at a consistent rate of -0.7%/year (95%CI: -1.5, 0) from around 2000-2016. Compared to the 1941-1959 central birth cohort, women born 1912-1940 had lower incidence rate ratios (IRR) for ER+ tumours and women born 1960- 1986 had higher IRR for ER- tumours.

Conclusions: We show evidence of aetiologic heterogeneity of breast cancer. Future incidence and survival reporting should be monitored by molecular subtypes.

Background

Breast cancer incidence is rising and it is the most common cancer among women worldwide (1). It is well established that breast cancer is not a single disease but comprises multiple subtypes, with oestrogen receptor (ER) expression a key marker of prognostic and aetiologic significance (2). ER+ tumours, which are amenable to targeted anti-oestrogenic therapies such as tamoxifen and aromatase inhibitors, are the most common type of breast cancers accounting for 65-75% of breast cancer cases in developed populations (3). Progesterone receptor (PR) is also a common marker of hormone responsiveness that is highly correlated with ER. Tumour over-expression of the human epidermal growth factor receptor 2 (HER2) was discovered two decades ago. This laid the foundation for biological therapies, which were shown to be clinically effective in treating tumours expressing HER2 and has been widely available in the UK since 2006 (4). ER- tumours do not express any of the hormone receptors described above and are rarer, have an earlier age of onset and worse prognosis than ER+ tumours, in part because fewer targeted treatments are available than for ER+ tumours. In addition to prognostic differences, epidemiologic studies have shown aetiologic differences by tumour subtypes (5, 6).

There are limited population cancer registries that collect ER, PR and HER2 data, the key distinguishing markers for molecular subtypes of breast cancer. Recent analyses support divergent incidence trends by ER status in the United States, Denmark and Ireland, with ER+ breast cancer incidence increasing and ER- breast cancer incidence decreasing (7-9). Data on combination of subtypes using ER, PR, and HER2 are even more limited, with few reports from the UK (10-12). ER, PR and HER2 molecular markers are used often as surrogates for the intrinsic subtypes of breast cancer defined by mRNA expression profiling (13) because, unlike genetic profiling, the markers have been measured routinely in recent years . In the age of precision medicine, quantifying and monitoring cancer incidence by molecular subtypes is important in optimising public health prevention programmes, the allocation of resources and availability of screening, diagnostic and therapeutic services, and for improving outcomes (14). An important aspect of assessing trends by ER status is to account for missing data, as completeness of marker data has improved over time, but imputation methods can be applied to address this limitation (7-9, 15).

Within Scotland's renowned, high-quality routine electronic health records, the Scottish cancer registry is an excellent resource to investigate temporal trends in cancer incidence. Data collection began for ER in 1997 and PR and HER2 in 2009 and so provides data almost a decade earlier than other UK national registries. While monitoring of breast cancer incidence in the UK is standard (16, 17), these data have not been presented by molecular subtypes. Further, recent changes in risk factors, such as, changes over time in reproductive factors and increasing obesity prevalence and alcohol consumption may have greater impact on ER+ specific breast cancer incidence rates compared to ER- given weaker/null associations with these subtypes (6, 18-20).

Here we report on breast cancer incidence trends in Scotland by ER and ER/HER2 combinations using several statistical methods: 1) Age-standardised and age-specific incidence rates, which are typically used to report cancer statistics (21); 2) Joinpoint regression models to determine whether significant changes occurred during 1997-2016, and the speed at which they have occurred (22); and 3) Age-Period-Cohort (APC) models (23-25) based on generalized linear model theory to enable description of age, period and birth cohort effects to provide possible clues to potential underlying factors contributing to incidence trends, and to thereby inform public health and NHS programmes.

Methods

Data and cohort definition

All primary invasive breast cancers (defined on the basis of the International Classification of Diseases, 10th revision code of C50) diagnosed in women aged 20+ years, between 1997 and 2016 were ascertained from the Scottish cancer registry held by Information Services Division (ISD) of NHS National Services Scotland. The Scottish cancer registry achieves 98% breast cancer case ascertainment and is over 99% complete (26). **Supplemental Figure 1** describes the creation of the final study population. Women with primary breast cancer are the basis for analysis, each characterised by her worst-prognosis tumour. Permission for use of the data was obtained from the Public Benefit and Privacy Panel (PBPP) of NHS Scotland (reference number 1718-0057) and analyses were conducted in the Scottish National Safe Haven.

Additional demographic and tumour data obtained were: age at diagnosis, NHS Scotland regions (North, South East and West), tumour grade (grade I-well differentiated to III-poorly differentiated), tumour size (less than 10mm, 10-20mm and more than 20mm), nodal involvement (yes or no), screen-detected tumour (yes or no) and the status of molecular markers ER, PR and HER2 (positive, negative or unknown). ER and PR status is measured using immunohistochemistry (IHC) and HER2 status was assessed using fluorescent in-situ hybridization. Previous studies have noted that assessment of ER status reliability is high with an error rate below 5% (27). ER/HER2 combinations were used as surrogates for the four intrinsic subtypes of breast cancer, the gold standard for which uses mRNA expression profiling. ER+/HER2- was used as surrogate for Luminal A tumours, ER+/HER2+ for Luminal B, ER-/HER2+ for HER2 enriched tumours and ER-HER2- for triple negative tumours. The high quality of these data has been previously described (28).

Statistical methods

Missing ER and ER/HER2 status were imputed conditioned on age and year of diagnosis, with the assumption that data were missing at random, using a validated method (7-9). Age-standardised incidence rates (ASR) per 100,000 women were calculated using the direct method, the European standard population (2013) (29) and mid-year estimates of the Scottish population for each age and year (30). Age-specific incidence rates were calculated for 5-year age groups (20-24 to 90+) and individual calendar years using two approaches: with number of tumours as the numerator for consistency with routine reporting and with one tumour per women as the numerator for all other analyses. ASR were calculated for all age groups combined and for three separate age-groups, with the middle group defined on the basis of eligibility for routine breast screening in Scotland, (20 to 49 years, 50 to 69 years and more than 70 years); and for each ER status and ER/HER2 combinations.

Joinpoint regression models were used to describe breast cancer incidence rates overall, by ER status and ER/PR/HER2 combinations for all women in the cohort and for three age groups (20-49, 50-69 and 70+ years). Joinpoint models describe if changes in the incidence trends occur and identify the time points at which a change is observed (referred to as joinpoints). The permutation test method, as described by Kim et al (22), was used iteratively: it starts by testing the null hypothesis of a simple model with zero joinpoints against the alternative hypothesis of a more complex model with the maximum number of

joinpoints previously specified (3 joinpoints for this study). The procedure continues until all possible numbers of joinpoints have been tested. In the final model, the estimated annual percentage change (EAPC) for each of the periods identified is calculated. The average annual percent change (AAPC) is also reported as a measure of the overall trend from 1997 to 2016. Joinpoint regression software was used for all the analysis (31).

APC models were fitted for age-standardised incidence of ER+ and ER- tumours. The APC model provides a unique set of best-fitting log incidence rates obtained by maximum likelihood estimators for period, age and cohort, which has been shown to provide similar rates to ASR, but allows investigation of differences by birth cohorts--with the middle cohort as referent-- which are not investigated in ASR or joinpoint regression analysis. As a consequence of small numbers in some strata, we restricted these models to women aged 30 to 85 years and used 28 two-year age groups (from 30-31 to 84-85) and 10 two-year periods (from 1997-1998 to 2015-2016) of calendar year of diagnosis, which covered birth cohorts from 1912 to 1986. The net drift, similar to the EAPC and AAPC estimate, is reported with 95% confidence intervals (CI). Local drifts were also estimated and describe the annual percentage change for each age-specific rate over time (32). In addition, period and cohort rate ratios are also presented to compare the age-specific rates in each period or cohort with the reference points in the middle of the study period and distribution of the birth cohort (2006 for period and 1949 for cohort). Along with cohort rate ratios, a combination test of significance for the complete cohort deviations is reported. This new combination test (32) aims to determine if there is an association of the observed rates with the birth cohorts above the linear influences represented by the net drift . The test provides a more robust method than the traditional Wald test while correcting for multiple testing. All statistical tests were two-sided and deemed statistically significant at the five percent level (p<0.05). With the exception of joinpoint regression, performed using specific software, all analyses were carried out using R (33).

Cumulative epidemiologic data supports a causal relationship between menopausal hormone therapy (MHT) and breast cancer risk (34, 35). To determine the possible proportion of women exposed, data from the Prescribing Information System in Scotland, which includes all items dispensed in Scotland was obtained from eDRIS to determine the

number of MHT prescriptions and estimated proportions of prescriptions of women who might be exposed from 1996-2016.

Results

Characteristics of cohort by ER status

Between 1997 and 2016, 72,217 women of 20 years of age or older were diagnosed with at least one invasive breast cancer in Scotland (**Table 1**). Seventy six percent of these tumours were ER+, 16% were ER- and 8% had unknown ER status. However, the percentage of missing ER status decreased over time from 20% in 1997 to 2% in 2016. Proportions with unknown ER status differed by region and age: was higher in the West compared to the North and Southeast of Scotland and in women aged 70 years or older compared to women aged less than 70 years (14% missing vs 5%). Almost half of breast cancers were diagnosed among women of 50 to 69 years of age, the range for eligibility for routine breast cancer screening.

Tumour characteristics differed by ER status, with ER- tumours having characteristics associated with more advanced/aggressive disease. ER- tumours had higher grade, were larger, more likely to have positive lymph node status and less likely to be screen detected than ER+ tumours. The patterns of other molecular markers also differed by ER status, with ER- tumours more likely to be PR- and HER2+ than ER+ tumours. In contrast, ER+ tumours were more likely to be PR+ and HER2- than ER- tumours.

The combinations of ER/HER2 status **Figure 1a** show that among women with known ER/HER2 status (N=27,580) most tumours were ER+/HER2-, with ER-/HER2+ tumours being the least common combination. ER-/HER2- tumours, the most aggressive subtype, was the second most common at 11%. Cross-sectional age specific curves for the ER/HER2 combinations (**Figure 1b**) show incidence of all subtypes increasing rapidly with age until the approximate age of menopause, age 50 years, after which the increase continues more gradually up to 70 years for ER+/HER2- tumours but did not increase further for ER- tumours or ER+/HER2+ tumours.

Age standardised incidence rates with EAPCs from joinpoint regression

As would be expected, increases in breast cancer incidence over time are less marked when using worst tumour characteristics per person (so that subjects are only counted once) rather than tumours as the numerator for incidence. Age-standardised incidence of ER+ tumours increased from 98 per 100,000 women in 1997 to 113 per 100,000 women in 2016 (**Figure 2**), with an average annual percentage change (AAPC) of 0.4% (95% CI: -0.1 to 1%). Incidence of breast cancer was higher for ER+/HER2- tumours than for the rest of the subtypes (**Figure 2**) and the time trend for incidence this subtype was similar to that of ER+ tumours with increases observed up to 2011. ER+/HER2+ and ER-/HER2+ tumour incidence decreased over time and there appeared to be an increase ER-/HER2- tumour incidence from 2011, although the latter finding was based on relatively small numbers.

Estimates from the joinpoint analysis (**Table 2**) show that the increase in incidence of ER+ tumours was reasonably constant (1.2% increase annually, 95% CI: 0.8 to 1.5) from 1997 till around 2012 after which incidence dereased by approximately 2.2% annually (95% CI: -4.7 to 0.4%). By contrast, ER- tumour incidence decreased consistently over the study period by approximately 2.5% per year (95% CI: -3.9 to -1.1%). Joinpoint regression results by age-group in **Table 2** revealed that women of 50 to 69 years of age had the highest increases in ER+ incidence at a similar period as noted overall (**Figure 3a**); followed by women aged 20 to 49 years where ER+ tumour incidence increased by 1.1% annually. For older women of 70 years or more rates were stable. The decreases observed in ER- tumours were consistent across the three age groups (**Figure 3b**).

Additional joinpoint regression analysis by the ER/HER2 combinations of breast cancer (**Supplemental Table 1**), revealed that ER+/HER2- and ER-/HER2+ breast cancer incidence remained constant from 2009 to 2016. In contrast, ER+/HER2+ incidence decreased for all women, and that decrease was predominantly among women aged more than 50 years old. ER-/HER2- breast cancer incidence slightly increased from 2011 to 2016 and was most marked among women aged 20 to 49 years.

Age-Period-Cohort models

Results from APC models were consistent with those observed from joinpoint regression, with net drifts suggesting increases in the overall incidence of ER+ tumours by 0.8% per year (95% CI: 0.6 to 1.0%/year) from 1997 to 2016, and ER- tumour incidence decreasing by 1.4%

(95% CI: -1.8 to -1.1%/year). After adjusting for period and cohort effects, local drifts showed that the highest increase in incidence of ER+ tumours was observed in women around 70 years of age (2% per year, 95%CI: 1.6 to 2.4%) (**Supplemental Figure 2a**). The greatest drop in incidence of ER- tumours was observed in women between 53 and 61 years of age (**Supplemental Figure 2b**).

Compared to the women born in 1949, ER+ tumour incidence was higher among more recent birth cohorts. In contrast, ER- incidence was lower for more recent birth cohorts compared to the cohort born in 1949. Cohort rate ratios (CRRs) compared to women born in 1949 ranged from 0.7 for women born in 1913 to 1.8 for women born in 1985 for ER+ tumours and from 1.5 for women born in 1913 to 0.5 for women born in 1985 for ER- tumours (**Figure 4**). The combination test for ER+ tumours revealed cohort effects beyond the log-linear trend shown by the net drift (p value=6.3×10^{-8}) but the test for ER- tumours failed to reach significance (p value=0.14).

MHT use in Scotland

Supplemental Table 2 shows that MHT utilisation has changed dramatically since 2003 in Scotland and use steadily declined following the results of the Women's Health Initiative trial indicating increased risks of cardiovascular disease and breast cancer among women randomised to MHT (36). Data from the Prescribing Information System in Scotland, which includes all items dispensed in Scotland, show that numbers of MHT-prescriptions and estimated proportions of women who might be MHT-exposed at 40-79 years of age as high as 16.5% in 1999 but as low as 4.7% by 2015.

Discussion

This study demonstrates that In Scotland, temporal trends of breast cancer incidence were distinct by molecular subtypes, with increases for ER+ and decreases for ER- tumours between 1997 and 2016. With respect to ER+ tumours, their incidence increased for all ages for the study period but particularly among women of screening ages 50 to 69 years, with the largest increases occurring from around 1997-2011 followed by modest declines. In contrast, the incidence of ER- cancers decreased among all ages till the early 2000's. Finally we noted cohort effects such that in comparison to women born around 1950, women of older generations (those born in the 1910s to 1940s) had a lower risk of ER+ tumours,

whereas there was no significant evidence for cohort effects for ER- tumours. Further analysis of the incidence trends by subtype (as defined by ER/HER2 combinations) showed similar results to those observed by ER status only. ER+/HER2- (surrogate for luminal A) tumours followed the same pattern as ER+ tumours, and ER-/HER2- (surrogate for Triple Negative) similar to ER- tumours.

Consistent with other reports from the United States, Denmark and Ireland (7-9), our data show for the first time in a UK national cancer registry, contrasting temporal trends of breast cancer incidence by ER status and suggest the presence of aetiologic heterogeneity with distinct patterns by period, age at diagnosis and birth cohort. Previous studies have shown estimated annual increases in the age-standardised rate of breast cancer from early 90's to the 2010 for ER+ ranging from 0.1 to 3% and declines for ER- ranging from -1.9 to -3.4% (7-9). The Scottish Cancer Registry's detailed tumour hormone receptor data have been used to describe trends in incidence patterns of breast cancer. Specifically, it was previously reported that there were declines in ER positive tumours among women 50–64 years of age that was statistically significant by 2005 (12). These findings were attributed to reduction in MHT use, which had been shown to be associated with increased risk of breast cancer (34, 36). Unlike the previous analysis we excluded women with a previous malignancy, imputed missing ER status, used individuals rather than tumours as the numerator for incidence rates but the findings were similar for comparable years. The declines in breast cancer incidence observed in Scottish data have been also shown in the United States (7, 37-39), Sweden (8) and France (40). We observed consistent increases over time for ER+ tumour incidence beyond 2002 after which MHT use declined. Based on recent reports on the association of MHT use and breast cancer risk (35) MHT has been estimated to contribute to 1 in 20 breast cancers diagnosed worldwide from 1990. In more recent years when MHT use has declined in use, it has been estimated to have an approximate 5 year lag time to breast cancer incidence and contribute to 2.3% of breast cancers in Scotland in recent years (41). Despite reductions in HRT use between 2005-2011 incidence of breast cancer continued to increase. In addition to long-term effects of previous HRT use, other factors, such as screening, obesity and alcohol consumption (41) are likely also to contribute to time trends in breast cancer incidence.

Mammographic screening is likely to be an important contributing factor to the increased incidence of ER+ tumours we observed from 1997-2011. In Scotland, the breast screening programme was established in 1988 with full national coverage attained in 1991 (42). Scotland's breast screening programme was introduced earlier than in other countries that have evaluated breast cancer incidence trends by ER status (ie. 2000 in Ireland and 2010 in Denmark, in the US there are no national screening programmes). In Scotland, from 1994-2003 women of 50-64 years of age were invited for screening, which was extended to include women of 65 to 70 years of age in 2003. Over the course of the entire study period in Scotland, the mammographic screening programme had around 75% uptake. Our data showing ER+ tumours are more likely to be screen detected than ER- tumours (31% vs 17%) and our APC model results showing incidence of ER+ tumours greatest for those of screening ages between 65 and 72 years, suggest that some of the increases observed in ER+ tumours is likely to be due to detection of prevalent disease in these older women. A similar pattern was also observed in the previous report (12). In contrast, screening is probably not a major contributor to the more recent declines in incidence that we observed for ER+ tumours starting in 2011 among women aged 50 to 69 years. Some studies have suggested that decreases in incidence may coincide with the saturation of the screening programme (7, 38), however, given that uptake in Scotland remained approximately stable for the whole study period with a slight fall (0.6%) for the last three-year period (42) this seems unlikely to be a contributing factor.

Obesity prevalence and alcohol intake have increased over time, although declines for alcohol in particular have been noted since 2003. Recent analysis estimates these two factors account for 18% of all breast cancers diagnosed in Scotland (41). The relationships of obesity and alcohol use are complicated and seem to be modified by menopausal status, but previous analysis suggest increased alcohol use to be associated with increased risk of ER+ tumours. Obesity in premenopausal women is particularly associated with a higher risk of ER-negative compared with ER-positive cancer (43-45). Secular trends in increasing obesity prevalence may contribute to the divergent trends in breast cancer incidence by ER status observed in younger women. Our analysis although showing overall increases in incidence for ER+ breast cancers for the entire study period, suggests for the first-time incidence of ER+ tumours not rising or slightly declining, particularly for women aged 50 to

69 years of age that started around 2011. This recent decrease in ER+ tumours incidence parallels results in Ireland showing incidence of ER+ tumours stabilizing or slightly decreasing from 2008 to 2013 (9). From national survey data in Scotland, obesity prevalence increased between 1995 and 2008 and then remained approximately stable until 2012 after which there was some evidence of declines especially among the most affluent groups of women (46). Future work is required to clarify the contribution of changes in obesity prevalence to trends in incidence of breast cancer and the lag time in its contribution to breast cancer incidence, including short and long term effects.

In addition to obesity, other factors seem to have a differential effect on risk of ER+ and ER- tumours. For example, reproductive factors such as younger age at menarche, older age at first birth, nulliparity, lack of breastfeeding and older age at menopause are associated with an increased risk of ER+ breast cancers (3, 47). Parous women have an increased risk of triple negative or basal-like tumours compared to nulliparous women (48-50). Breastfeeding has been consistently associated with a reduced risk particularly more aggressive subtypes of breast cancer, namely ER- (48) and triple negative tumours (47, 48). In Scotland, reproductive patterns have changed considerably over time: the average age of a woman at first birth has increased from 26 years in 1975 to 30.5 years in 2017, fertility rates have decreased and the number and proportion of women not having children has also increased in recent decades (51, 52). Our study provides evidence of changes in incidence of breast cancer between birth cohorts by ER status with higher year of birth positively associated with risk of ER+ breast cancer and inversely associated with risk of ER-. The reproductive factor changes that have occurred in Scotland may be associated with the increasing ER+ breast cancer incidence in more recent birth cohorts. Breastfeeding practice at 6-8 weeks has increased in Scotland from 36% in 2001 to 42% in 2018 (53), and this may have also contributed to some of the declines observed for ER- tumours. Current, estimates suggest 5.2% of breast cancer attributable to not breastfeeding in Scotland (41).

The strengths of our study are the high quality of the longitudinal data collected within the Scottish cancer registry, the first one in the UK and fifth in the world that routinely started recording molecular marker data (ER status from 1997 and PR and HER2 status from 2009). Marker data can be used to monitor and describe incidence trends in the future and for other types of cancer that also display heterogeneity. Further, monitoring breast cancer

incidence by molecular subtypes can help the NHS allocate resources for treatment and prevention and lead to the identification of high-risk groups of women for which to implement future prevention programmes and treatments.

Limitations of our study include the absence of individual level risk factor data. However, future studies should be possible to identify some key factors using linked data. Scotland is renowned for its high-quality longitudinal data and ability to perform linkage studies using a unique identifier. Hence, we envision future analysis using the Scottish cancer registry linked to other datasets, including community prescription drug records, maternity records and hospital records to provide more detailed information on the role of patterns of key risk factors in breast cancer incidence trends. Another limitation of the study is the lack of mRNA expression assays for the classification of the molecular subtypes of breast cancer. In our study, markers measured by IHC are used as surrogates for the molecular subtypes, which are reasonably good proxies but mRNA profiling data would be considered a gold-standard for intrinsic-subtype classification (54).

In conclusion, incidence trends of breast cancer in Scotland differ by ER status and are consistent with trends observed in other countries. Additional data are needed to establish whether HER2+ tumours, which are ER-, remain low since trastuzumab is one of the costliest breast cancer treatments used by the NHS. Further research should be focused on monitoring incidence trends by subtype given notable risk and treatment differences for breast cancer subtypes.

Additional information

Ethics approval and data Availability:

The data used in this study can be accessed through application to electronic Data Research and Innovation Service (eDRIS), a part of the Information Services Division of NHS Scotland. Approval from the Public Benefit and Privacy Panel for Health (PBPP) and Social Care is a requirement for data access. Our project was approved by PBPP reference number 1718-0057.

Conflict of interest:

SMB holds shares in GlaxoSmithKline. Other authors declare no competing interest.

Funding:

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

Authors' contributions:

Conception & Design of the study: JDF, SW, IME, SB

Interpretation of data: All authors

Drafting of the manuscript: JDF, IME, SW

Revised work and provided important intellectual content: All authors

Final approval of the manuscript: All authors

Acknowledgements**:**

We thank NHS Scotland and Information Services Division for collection of data used for this analysis. We specifically would like to acknowledge Andrew Deas, Suhail Iqbal, Rita Nogueira and Ross Murdoch of NHS National Services Scotland.

References

1.   Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.

2.   Polyak K. Heterogeneity in breast cancer. The Journal of Clinical Investigation. 2011;121(10):3786-8.

3.   Brinton LA, Gaudet MM, Gierach GL. Chapter 45- Breast Cancer. In: Thun M, Linet MS, Cerhan JR, Haiman CA, Schottenfeld D, editors. Cancer Epidemiology and prevention: Oxford University Press; 2017.

4.   Mayor S. NICE approves trastuzumab for early stage breast cancer. BMJ (Clinical research ed). 2006;332(7555):1409-.

5.   Gaudet MM, Press MF, Haile RW, Lynch CF, Glaser SL, Schildkraut J, et al. Risk factors by molecular subtypes of breast cancer across a population-based study of women 56 years or younger. Breast Cancer Res Treat. 2011;130(2):587-97.

6.   Tamimi RM, Colditz GA, Hazra A, Baer HJ, Hankinson SE, Rosner B, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. Breast Cancer Research and Treatment. 2012;131(1):159-67.

7.   Anderson WF, Katki HA, Rosenberg PS. Incidence of Breast Cancer in the United States: Current and Future Trends. Journal of the National Cancer Institute. 2011;103(18):1397-402.

8.   Anderson WF, Rosenberg P, Petito L, Katki HA, Ejlertsen B, Ewertz M, et al. Divergent estrogen receptor positive and negative breast cancer trends and etiologic heterogeneity in Denmark. Cancer Res. 2013;73(8):1.

9.   Mullooly M, Murphy J, Gierach GL, Walsh PM, Deady S, Barron TI, et al. Divergent oestrogen receptor-specific breast cancer trends in Ireland (2004-2013): Amassing data from independent Western populations provide etiologic clues. European Journal of Cancer. 2017;86:326-33.

10.   Agboola AJ, Musa AA, Wanangwa N, Abdel-Fatah T, Nolan CC, Ayoade BA, et al. Molecular characteristics and prognostic features of breast cancer in Nigerian compared with UK women. Breast Cancer Research and Treatment. 2012;135(2):555-69.

11.   Jack RH, Davies EA, Renshaw C, Tutt A, Grocock MJ, Coupland VH, et al. Differences in breast cancer hormone receptor status in ethnic groups: A London population. European Journal of Cancer. 2013;49(3):696-702.

12.   Sharpe KH, McClements P, Clark DI, Collins J, Springbett A, Brewster DH. Reduced risk of oestrogen receptor positive breast cancer among peri- and post-menopausal women in Scotland following a striking decrease in use of hormone replacement therapy. European Journal of Cancer. 2010;46(5):937-43.

13.   Perou CM, Sørlie T, Eisen MB, Van De Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747.

14.   Bray F, Møller B. Predicting the future burden of cancer. Nature Reviews Cancer. 2006;6(1):63-74.

15.   Howlader N, Noone AM, Yu M, Cronin KA. Use of imputed population-based cancer registry data as a method of accounting for missing information: application to estrogen receptor status for breast cancer. Am J Epidemiol. 2012;176(4):347-56.

16.   ISD- NHS Scotland. National Cancer Registration Data Definitions Version 14.4 2017 [Available from: https://www.isdscotland.org/Health-Topics/Cancer/Scottish-Cancer-Registry/Cancer-Metadata/_docs/Cancer-Registration-Definitions-v14-4.pdf.

17.   Cancer Research UK. Breast cancer incidence trends over time 2019 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#ref-.

18.   Ellingjord-Dale M, Vos L, Hjerkind KV, Hjartaker A, Russnes HG, Tretli S, et al. Alcohol, Physical Activity, Smoking, and Breast Cancer Subtypes in a Large, Nested Case-Control Study from

the Norwegian Breast Cancer Screening Program. Cancer Epidemiol Biomarkers Prev. 2017;26(12):1736-44.

19.	Hirko KA, Chen WY, Willett WC, Rosner BA, Hankinson SE, Beck AH, et al. Alcohol consumption and risk of breast cancer by molecular subtype: Prospective analysis of the nurses' health study after 26 years of follow-up. Int J Cancer. 2016;138(5):1094-101.

20.	Nattenmüller CJ, Kriegsmann M, Sookthai D, Fortner RT, Steffen A, Walter B, et al. Obesity as risk factor for subtypes of breast cancer: results from a prospective cohort study. BMC Cancer. 2018;18(1):616-.

21.	Esteve J, Benhamou E, Raymond L. Statistical methods in cancer research. Volume IV. Descriptive epidemiology. IARC Sci Publ. 1994(128):1-302.

22.	Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med. 2000;19(3):335-51.

23.	Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. Stat Med. 1987;6(4):469-81.

24.	Holford TR. The estimation of age, period and cohort effects for vital rates. Biometrics. 1983;39(2):311-24.

25.	Rosenberg PS, Anderson WF. Age-period-cohort models in cancer surveillance research: ready for prime time? Cancer Epidemiol Biomarkers Prev. 2011;20(7):1263-8.

26.	Brewster DH, Stockton DL. Ascertainment of breast cancer by the Scottish Cancer Registry: an assessment based on comparison with five independent breast cancer trials databases. Breast. 2008;17(1):104-6.

27.	Brewster DH, Stockton D, Harvey J, Mackay M. Reliability of cancer registration data in Scotland, 1997. European journal of cancer (Oxford, England : 1990). 2002;38(3):414-7.

28.	Gray E, Marti J, Brewster DH, Wyatt JC, Hall PS. Independent validation of the PREDICT breast cancer prognosis prediction tool in 45,789 patients using Scottish Cancer Registry data. Br J Cancer. 2018;119(7):808-14.

29.	European Commission. Revision of the European Standard Population- Report of Eurostat's task force. Luxembourg: Publications Office of the European Union; 2013.

30.	National records of Scotland. Mid-Year Population Estimates 2019 [Available from: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates.

31.	Joinpoint Regression Program, Version 4.6.0.0 - Statistical Methodology and Applications Branch, Surveillance Research Program, National Cancer Institute.; April 2018.

32.	Rosenberg PS. A new age-period-cohort model for cancer surveillance research. Stat Methods Med Res. 2018:962280218801121.

33.	R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. ; 2017.

34.	Beral V, Million Women Study C. Breast cancer and hormone-replacement therapy in the Million Women Study.[Erratum appears in Lancet. 2003 Oct 4;362(9390):1160]. Lancet. 2003;362(9382):419-27.

35.	Collaborative Group on Hormonal Factors in Breast Cancer. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. The Lancet. 2019;394(10204):1159-68.

36.	Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. Jama. 2002;288(3):321-33.

37.	Hou NQ, Huo DZ. A trend analysis of breast cancer incidence rates in the United States from 2000 to 2009 shows a recent increase. Breast Cancer Research and Treatment. 2013;138(2):633-41.

38.	Jemal A, Ward E, Thun MJ. Recent trends in breast cancer incidence rates by age and tumor characteristics among U. S. women. Breast Cancer Res. 2007;9(3):6.

39.   Keegan TH, Chang ET, John EM, Horn-Ross PL, Wrensch MR, Glaser SL, et al. Recent changes in breast cancer incidence and risk factor prevalence in San Francisco Bay area and California women: 1988 to 2004. Breast Cancer Res. 2007;9(5):12.

40.   Fontenoy AM, Leux C, Delacour-Billon S, Allioux C, Frenel JS, Campone M, et al. Recent trends in breast cancer incidence rates in the Loire-Atlantique, France: A decline since 2003. Cancer epidemiology. 2010;34(3):238-43.

41.   Brown KF, Rumgay H, Dunlop C, Ryan M, Quartly F, Cox A, et al. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. Br J Cancer. 2018;118(8):1130-41.

42.   ISD- NHS Scotland. Scottish breast screening pogramme statistics 2015/2016 2017 [Available from: https://www.isdscotland.org/Health-Topics/Cancer/Publications/2017-04-25/2017-04-25-SBSP-Cancer-Report.pdf.

43.   Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor–defined breast cancer: a systematic review of the literature. Cancer Epidemiology and Prevention Biomarkers. 2004;13(10):1558-68.

44.   Kerlikowske K, Gard CC, Tice JA, Ziv E, Cummings SR, Miglioretti DL. Risk factors that increase risk of estrogen receptor–positive and–negative breast cancer. JNCI: Journal of the National Cancer Institute. 2017;109(5).

45.   Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. Journal of the National Cancer Institute. 2011;103(3):250-63.

46.   Scottish Government. Obesity Indicators 2018. 2018.

47.   Lambertini M, Santoro L, Del Mastro L, Nguyen B, Livraghi L, Ugolini D, et al. Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. Cancer Treat Rev. 2016;49:65-76.

48.   Fortner RT, Sisti J, Chai B, Collins LC, Rosner B, Hankinson SE, et al. Parity, breastfeeding, and breast cancer risk by hormone receptor status and molecular phenotype: results from the Nurses' Health Studies. Breast Cancer Res. 2019;21(1):40.

49.   Gaudet MM, Carter BD, Brinton LA, Falk RT, Gram IT, Luo J, et al. Pooled analysis of active cigarette smoking and invasive breast cancer risk in 14 cohort studies. Int J Epidemiol. 2017;46(3):881-93.

50.   Palmer JR, Viscidi E, Troester MA, Hong CC, Schedin P, Bethea TN, et al. Parity, lactation, and breast cancer subtypes in African American women: results from the AMBER Consortium. Journal of the National Cancer Institute. 2014;106(10).

51.   ISD- NHS Scotland. Births in Scottish Hospitals 2018 [Available from: https://www.isdscotland.org/Health-Topics/Maternity-and-Births/Publications/2018-11-27/2018-11-27-Births-Report.pdf.

52.   National Records of Scotland. Scotland's Population- The registrar general's annual review of demographic trends 2018 [Available from: https://www.nrscotland.gov.uk/files/statistics/rgar/2017/rgar17.pdf.

53.   ISD- NHS Scotland. Infant Feeding Statistics Scotland 2018 [Available from: https://www.isdscotland.org/Health-Topics/Child-Health/Publications/2018-10-30/2018-10-30-Infant-Feeding-Report.pdf.

54.   Tang P, Tse GM. Immunohistochemical Surrogates for Molecular Classification of Breast Carcinoma: A 2015 Update. Arch Pathol Lab Med. 2016;140(8):806-14.

Figure Legends

Figure 1: Distribution of the breast cancer subtypes by ER/HER2 status and their age-specific incidence in Scotland for 2009-2016(N=31,099)

Panel (a) shows a pie chart and panel b) shows age specific incidence on the log scale by subtype (b) Data are for 31,099 breast cancer cases with ER/HER2 status recorded and missing status imputed for analysis. Dotted lines in graph denote ages 50 to 70 years, the age-group invited for screening in Scotland every 3-years.

Figure 2: Age-standardised breast cancer incidence rates by ER and HER2 status in Scotland for 1997-2016

ER status was available for the entire period and data are for N=72,217 breast cancer cases. ER/HER2 combinations were available from 2009 to 2016, N=31,099. Dotted line at year 2009 denotes when HER2 status started to be collected in the Scottish cancer registry. Rates are age-standardised to the 2013 European population and missing ER and HER2 marker status were imputed based on age and year of diagnosis (see methods).

Figure 3: Age-specific trends in breast cancer incidence stratified by age groups and ER status in Scotland for 1997-2016

ER positive (a) and ER negative (b) age-specific trends for age groups 20-49 (red), 50 to 69 (blue) and 70 years old (green). Shaded areas surrounding lines indicate 95% CI of rates.

Figure 4: Birth cohort rate ratios (CRR) for breast cancer incidence rates in Scotland by ER status

CRR describe the incidence rates for each birth cohort relative to the 1949 birth cohort

Table 1: Descriptive characteristics by ER status for all women with an invasive breast cancer diagnosed between 1997 and 2016 in Scotland

Characteristics		ER-		ER+		ER unknown	
		n	%	n	%	n	%
		11726	[16]	55144	[76]	5347	[8]
Age at diagnosis	<50 years	3196	(27)	10550	(19)	695	(13)
	50 to 69 years	5668	(48)	28441	(52)	1580	(30)
	70 years or older	2862	(24)	16153	(29)	3072	(57)
Grade	I-Well differentiated	195	(2)	8288	(15)	232	(4)
	II- Moderately differentiated	1714	(15)	25734	(47)	602	(11)
	III- Poorly differentiated	8308	(71)	14586	(26)	642	(12)
	Unknown	1509	(13)	6536	(12)	3871	(72)
Nodal Status	Uninvolved/negative	6194	(53)	29400	(53)	869	(16)
	Involved/positive	4110	(35)	17369	(31)	415	(8)
	Unknown	1422	(12)	8375	(15)	4063	(76)
Screen detected	No	9622	(82)	37400	(68)	4047	(76)
	Yes	1943	(17)	17119	(31)	490	(9)
	Unknown	161	(1)	625	(1)	810	(15)
Tumour size	Less than 10mm	1017	(9)	6470	(12)	202	(4)
	10 to 20 mm	3428	(29)	20449	(37)	478	(9)
	More than 20mm	4960	(42)	18168	(33)	512	(10)
	Unknown	2321	(20)	10057	(18)	4155	(78)
PR status*	Negative	3803	(79)	3036	(12)	<10	(<1)
	Positive	226	(5)	15869	(62)	<10	(<1)
	Unknown	764	(16)	6489	(26)	901	(99)
HER2 status*	Negative	2761	(66)	18709	(84)	36	(5)
	Positive	1210	(29)	2553	(11)	10	(1)
	Unknown	184	(4)	1129	(5)	725	(94)

Brackets [ ] indicate row percentages and parenthesis ( ) indicate column percentages for that category. *Denotes markers that were recorded from 2009 to 2016, and number of cases for those years=31,099. Differences by known ER status for all characteristics were significantly different with χ^2 p<0.001.

Table 2: Joinpoint regression analysis stratified by age groups and ER status from 1997 to 2016.

ER status	Age groups	Rate in 1997 per 100,000 women	Rate in 2016 per 100,000 women	Change in rate from 1997 to 2016 N per 100,000 (%)	Average Annual Percentage Change (95%CI)	N for complete case analysis	N for estimated counts corrected for missing ER status	Years before joinpoint	EAPC (95%CI) for period before joinpoint	Years after joinpoint	EAPC (95%CI) for period after joinpoint
Positive	20-49	41.9	52.1	10.2 (20%)	1.1% (0.7, 1.5)	10,550	11,083	No significant change point identified from 1997-2016			
	50-69	192.3	237.4	45.1 (19%)	0.7% (0.2, 1.3)	28,441	29,758	1997-2011	1.6% (1.2, 2.1)	2011-2016	-1.8 (-3.7, 0.1)
	70+	235.9	234.5	-1.4 (0.6%)	0.1% (-0.3, 0.5)	16,153	18,763	No significant change point identified from 1997-2016			
	All ages	97.7	112.8	15.1 (13%)	0.4% (-0.1, 1.0)	55,144	59,604	1997-2012	1.2% (0.8, 1.5)	2012-2016	-2.2 (-4.7, 0.4)
Negative	20-49	23.8	15.2	-8.6 (36%)	-2.2% (-3.9, -0.6)	3,196	3,358	1997-2001	-10% (-17.0, -3.0)	2001-2016	0% (-1.1, 1.2)
	50-69	64.1	45.5	-18.6 (29%)	-1.6% (-2.5, -0.8)	5,668	5,931	No significant change point identified from 1997-2016			

70+	71.8	41.2	-30.6 (43%)	-2.4% (-4.2, -0.7)	2,862	3,324	1997-2003	-7% (-11.0, -2.0)	2003-2016	-0.3% (-1.9, 1.5)
All ages	35.5	23.1	-12.4 (35%)	-2.5% (-3.9, -1.1)	11,726	12,613	1997-2000	-11% (-19.0, -3.0)	2000-2016	-0.7% (-1.5, 0)

EAPC estimated annual percentage change. AAPC estimated average annual percentage change. Joinpoint regression was performed using the estimated counts corrected for missing ER status, and analysis corrects for multiple testing using Bonferroni correction (See methods section).

Figure 1: Distribution of the breast cancer subtypes by ER/HER2 status and their age-specific incidence in Scotland for 2009-2016(N=31,099)
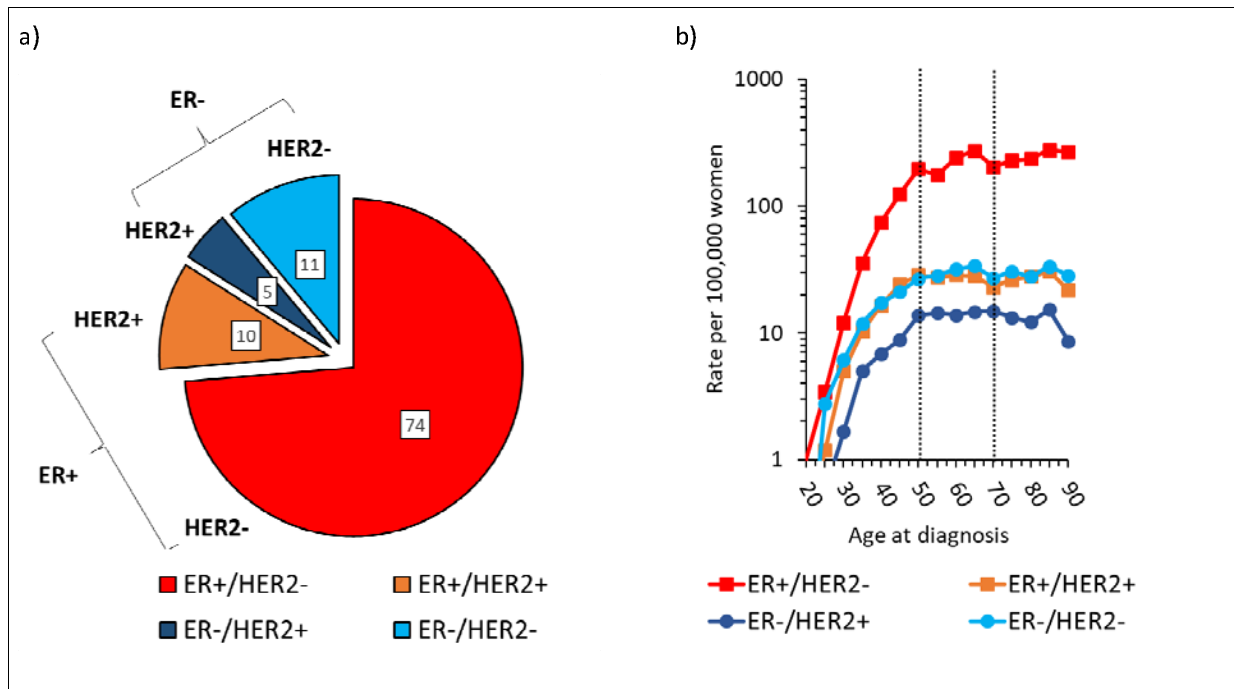
Figure 2: Age-standardised breast cancer incidence rates by ER and HER2 status in Scotland for 1997-2016
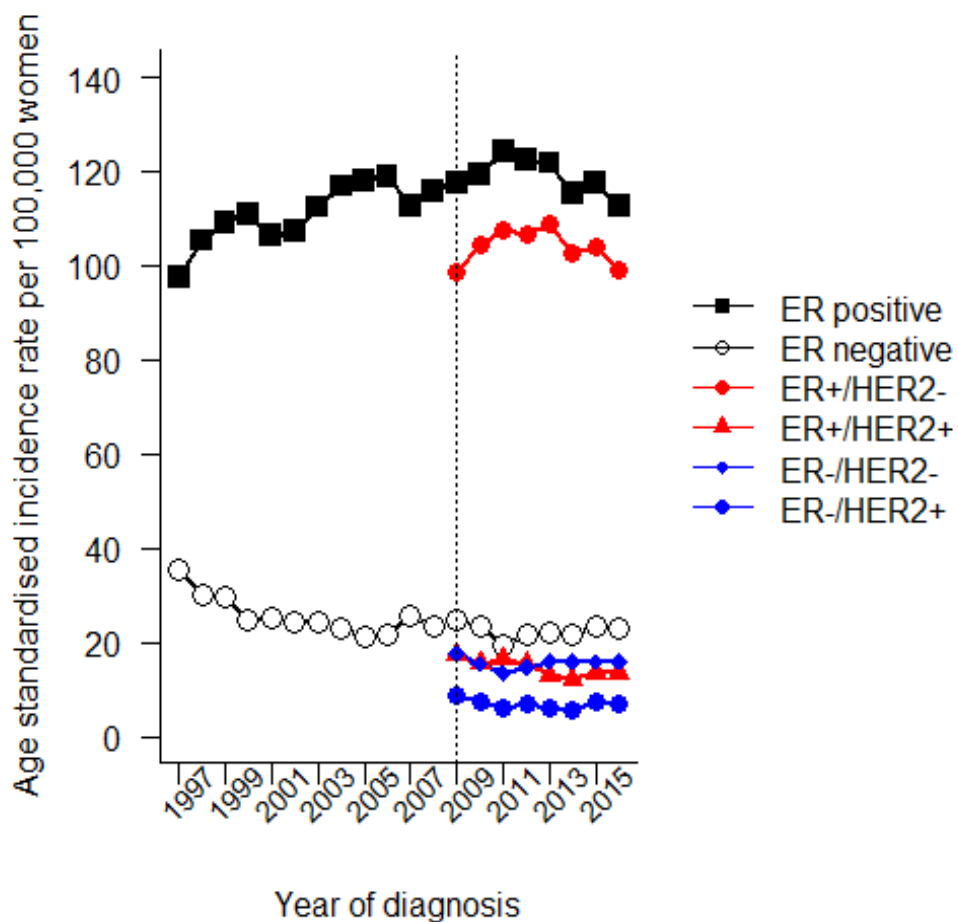
Figure 3: Age-specific trends in breast cancer incidence stratified by age groups and ER status in Scotland for 1997-2016
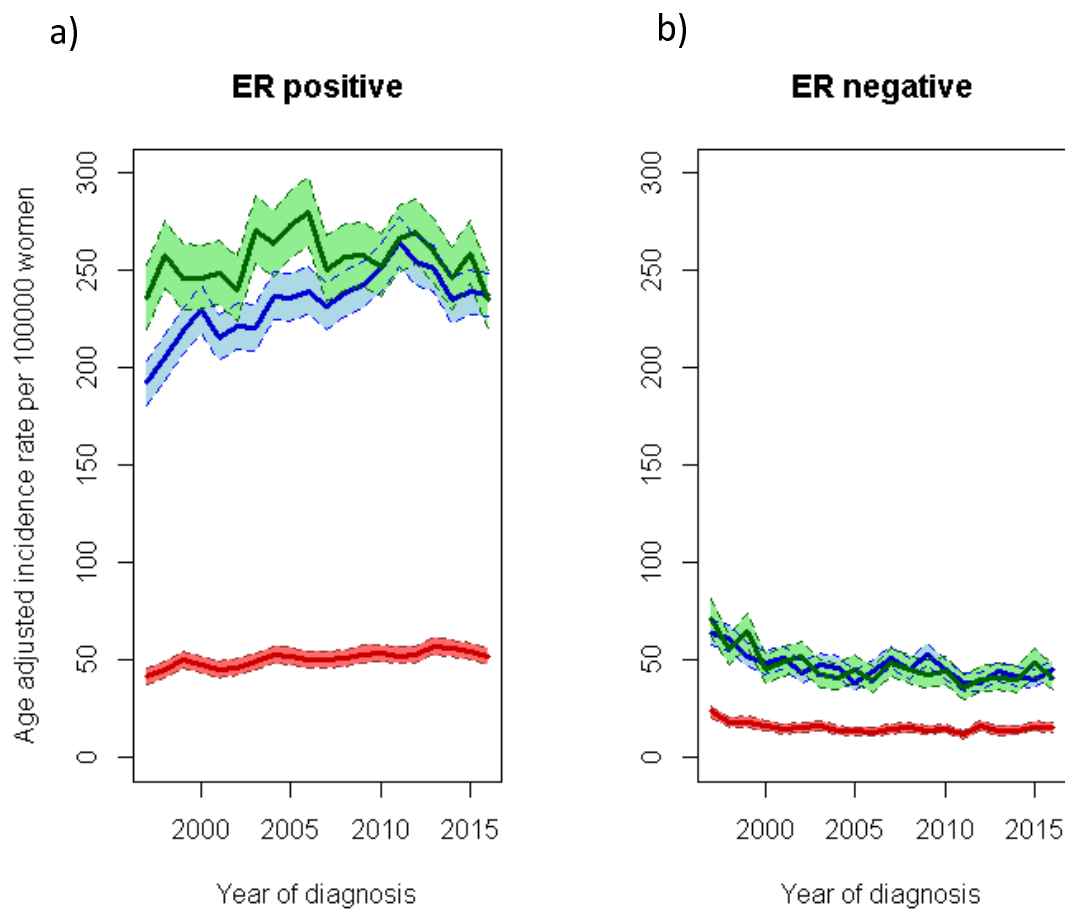
Figure 4: Birth cohort rate ratios (CRR) for breast cancer incidence rates in Scotland by ER status