TITLE PAGE

Smoking and risk of breast cancer in the Generations Study cohort

AUTHORS

Michael E Jones[1]
Minouk J Schoemaker[1]
Lauren B Wright[1]
Alan Ashworth[2,3,4,5]
Anthony J Swerdlow[1,2]

AUTHOR AFFILIATIONS

1) Division of Genetics and Epidemiology, The Institute of Cancer Research, London, United Kingdom, SW7 3RP
2) Division of Breast Cancer Research, The Institute of Cancer Research, London, United Kingdom, SW7 3RP
3) Breakthrough Breast Cancer Research Centre at The Institute of Cancer Research, London, United Kingdom, SW7 3RP
4) Division of Molecular Pathology, The Institute of Cancer Research, London, United Kingdom, SW7 3RP
5) Current Address: UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, California, USA, 94158

AUTHOR E-MAIL ADDRESS

Michael E Jones — Michael.Jones@icr.ac.uk
Minouk J Schoemaker — Minouk.Schoemaker@icr.ac.uk
Lauren B Wright — Lauren.Wright@icr.ac.uk
Alan Ashworth — Alan.Ashworth@ucsf.edu
Anthony J Swerdlow — Anthony.Swerdlow@icr.ac.uk

CORRESPONDING AUTHOR

Dr Michael E Jones
Division of Genetics and Epidemiology
The Institute of Cancer Research
London
United Kingdom
SM2 5NG
Tel: +44 (0)20 8722 4254
Fax: +44 (0)20 8722 4019
E-mail: Michael.Jones@icr.ac.uk

ABSTRACT

Background: Plausible biological reasons exist why smoking could affect breast cancer risk but epidemiological evidence is inconsistent.

Methods: We used serial questionnaire information from the Generations Study cohort (United Kingdom) to estimate hazard ratios (HRs) for breast cancer in relation to smoking adjusted for potentially confounding factors including alcohol intake.

Results: Among 102,927 women recruited 2003–2013, with 7.7 years average follow-up, 1815 developed invasive breast cancer. The HR (reference group: never smoker) was 1.14 (95% confidence interval (CI): 1.03–1.25; *P*=0.010) for ever-smoking, 1.24 (95% CI: 1.08–1.43; *P*=0.002) for starting smoking at ages <17 years, and 1.23 (1.07–1.41; *P*=0.004) for starting smoking 1–4 years after menarche. Breast cancer risk was not statistically associated with interval from initiation of smoking to first birth (*P-trend*=0.97). Women with a family history of breast cancer (ever smoker vs never smoker HR=1.35; 95% CI: 1.12–1.62; *P*=0.002) had significantly larger HR in relation to ever-smoking (interaction: *P*=0.039) than women without (ever smoker vs never smoker HR=1.07; 95% CI: 0.96–1.20; *P*=0.22); the interaction was prominent for age started smoking (*P*=0.003) and starting smoking relative to age at menarche (*P*=0.0001).

Conclusions: Smoking was associated with a modest but significantly increased risk of breast cancer, particularly among women who started smoking at adolescent or peri-menarcheal ages. The relative risk of breast cancer associated with smoking was greater for women with a family history of the disease.

KEYWORDS:

smoking, breast neoplasms, cohort studies

BACKGROUND

The carcinogenic potential of tobacco smoke is unarguable [1, 2] and there are plausible biological reasons why smoking could affect breast cancer risk [2-5]. Reviews of the association between cigarette smoking and breast cancer up to 2004 did not, however, generally find conclusive evidence for a causal relationship in humans [5-7]. More recent epidemiological analyses have reported modest raised risks with current [8-19] or former [8-15, 20] smoking, but questions remain about the extent to which this association is a consequence of confounding by alcohol use, whether risk is increased if smoking starts in adolescence or before first childbirth, and whether risk is modified by family history of breast cancer [1, 2]. We therefore examined risk of invasive breast cancer in relation to smoking in a large cohort study using detailed questionnaire information at recruitment and during follow-up, with adjustment for alcohol consumption and other potentially confounding factors.

METHODS

The Generations Study is a cohort study of over 113,700 women aged 16 or older from the United Kingdom, from whom questionnaire information and informed consent was gained at recruitment since 2003 [21]. Initial recruits to the cohort were from women involved in the breast cancer charity that funded the study, and women who responded to publicity about the study. Women who joined the study were asked to nominate female friends and family members, who were then contacted about joining the study. This referral method continued with subsequent recruits [21]. The first follow-up questionnaire (2½ years after recruitment) was completed by 99% of non-deceased participants, a second (six years after recruitment) by 96%, and a third (9½ years after recruitment) by 94% (of those recruited long enough ago to have entered this round of follow-up). The study was approved by the South East Multi-Centre Research Ethics Committee.

Breast and other cancers occurring in the cohort were identified from recruitment and follow-up questionnaires, spontaneous reports to the study centre, and from 'flagging' (see below) for those lost to questionnaire follow-up. Confirmation of diagnosis was obtained from cancer

registries in the United Kingdom, 'flagging' at the National Health Service Central Registers (virtually complete registers of the populations of England and Wales, and of Scotland, to which study participants can be linked and on which deaths, cancer registrations, and emigrations are 'flagged' and then periodically reported to authorized medical researchers), pathology reports, and correspondence with patients' general practitioners.

Information on risk factors for breast cancer was obtained from recruitment and follow-up questionnaires. In relation to smoking, women were asked if they had "ever smoked regularly (i.e. most days for at least 6 months)", if they still smoked regularly, age started and stopped, and number of cigarettes smoked per day at different periods of their lives (during ages: 16–24, 25–49, 50+ years). For analysis, we defined the period of 'current smoking' to include both current smokers and the year immediately after stopping, to avoid potential 'reverse-causation' bias from women who may have stopped smoking during the work-up to a formal breast cancer diagnosis. For alcohol use we asked women if they had been a regular drinker "in the sense of drinking at least one glass of alcohol per week on average", ages started and stopped, and quantity consumed at different periods of life (during ages: 18–24, 25–49, 50+ years). We converted the quantity of alcohol consumed at each period of life into daily grams of alcohol. We split into three groups the women who reported current drinking (<60g/day, 60+g/day, and amount unknown), and we classified women who had reported stopping drinking as former drinkers. For some women we did not know their current drinking status during follow-up, but we knew they had consumed alcohol in the past and these women were classified as 'ever-drinkers'. Because we had collected ages or dates at which certain events or changes in lifestyle occurred we were able to update smoking status, alcohol use, parity, oral contraceptive (OC) use, menopausal hormone therapy (MHT) use, and menopausal status, at the ages these episodes occurred through to the second follow-up questionnaire. We updated duration of smoking for current smokers, and time since cessation for former smokers, in yearly increments, using smoking start and stop ages from the recruitment and second follow-up

questionnaire. We updated cigarettes smoked per day, pack-years smoked, alcohol consumption, and post-menopausal body mass index (BMI), at the date of the second follow-up questionnaire.

Statistical analysis

The current analytic cohort is based on all women who were recruited to the study during June 2003–December 2013 without prior invasive or *in-situ* breast cancer or other malignant cancer (except non-melanoma skin cancer), or prior mastectomy. The recruitment cut-off at December 2013 was selected because at the time of analysis the second follow-up was practically complete for this group of recruits, two-thirds of the cohort had reached the third follow-up, and we had 'flagging' information to June 2017. Women entered risk at their date of recruitment and were censored at the earliest date of: invasive breast cancer or *in-situ* breast cancer; other malignancy (except non-melanoma skin cancer); death; most recent follow-up questionnaire (depending on date of recruitment) if completed, or the date most recent follow-up questionnaire was due if cancer and vital status was known from 'flagging'; or previously completed questionnaire if lost to follow-up. We censored follow-up at *in-situ* breast cancer or other malignancy because we reasoned that if smoking is related to risk of *in-situ* breast cancer or other malignancy, and ensuing treatments or their consequences alter risk of subsequent invasive breast cancer, including subsequent follow-up may obscure associations between smoking and invasive breast cancer.

Left-truncated and right censored Cox proportional hazards regression [22] using attained age as the implicit time scale was used to estimate hazard ratios (HR) and 95% confidence intervals (CI) for smoking and risk of first invasive breast cancer. We adjusted for: time since recruitment to cohort (0, 1–2, 3+ years); birth cohort (1908–39, 1940–49, 1950–59, 1960–69, 1970–96); benign breast disease (yes, no); family history of breast cancer in 1st degree relatives (yes, no); socio-economic score (ACORN score as trend, or missing indicator); age at menarche (trend, or missing indicator); age at first pregnancy (trend, or missing indicator); parity (trend, or missing indicator); duration of breastfeeding (trend, or missing indicator); current OC use during follow-up, before

menopause (yes, no); alcohol consumption (trend for current drinker 1– <60g/day, indicator variables for never regular, current drinker 60+g/day, past drinker, drinker with unknown details); physical activity (log(metabolic equivalent) trend, missing indicator); pre-menopausal BMI at age 20 years (trend, or missing indicator); post-menopausal BMI (trend, or missing indicator); MHT use (never used, ex-user, current estrogen only user, current estrogen plus progestogen user, current user of other types, missing indicator); menopausal status (pre- or post-menopausal) and age at menopause (trend, or missing indicator). BMI was used to create two separate variables: pre-menopausal BMI (potentially available for all women) and post-menopausal BMI (only available at post-menopausal ages). We used BMI at age 20 to represent pre-menopausal BMI. Separately, if a woman was post-menopausal at entry to the cohort we used her BMI at entry for her post-menopausal BMI (and if she was pre-menopausal at this time, post-menopausal BMI was unknown). If a woman was post-menopausal at the time of the follow-up questionnaire we updated from this point in time her post-menopausal BMI with the value from this follow-up questionnaire. Statistical trends were evaluated using continuous values, except for duration and time since cessation of smoking which were based on discrete time-varying annually updated values. For trend analyses where there was an unexposed group (e.g. never smokers in analyses of smoking duration) the unexposed group was not assigned a zero magnitude but was treated as a separate categorical term, as was any missing value group. In particular we adjusted our analyses of smoking and breast cancer for alcohol using daily current alcohol consumption as a continuous measure if within the range 1– <60g/day, and categorical terms for non-drinkers, for those with consumption 60+g/day (because we did not want a minority of women who reported very high consumption to influence unduly the trend with daily consumption), past drinkers, and those for whom details of consumption were missing, by fitting appropriate interaction terms in the Cox regression model. Heterogeneity in HRs by sub-type of breast cancer defined by estrogen receptor (ER) status or morphology was assessed using a data augmentation method [23] and Wald chi-square tests [24]. All statistical tests were two-sided and analyses were conducted using Stata/IC version 14.0 [25].

RESULTS

During 2003–2013 a recruitment questionnaire was completed by 102,940 women who had no previous invasive or *in-situ* breast cancer or other malignancy (except non-melanoma skin cancer). At censoring date 1.1% of women had died. Of the remainder, cancer and vital status was known for 96.5% who had completed the relevant follow-up questionnaire, and a further 2.4% from 'flagging' at the National Health Service Central Registers. The remaining 1.1% were lost to follow-up at an earlier date. Thirteen women (including one with breast cancer) were excluded from subsequent analyses because of self-contradictory information for parity or smoking, leaving 102,927.

Table 1 presents descriptive characteristics at recruitment of the cohort eligible for analysis. The median age at recruitment was 47 years (Inter-Quartile Range (IQR): 36–57). A majority of participants, 64.1%, reported never smoking but only 10.3% were never-regular consumers of alcohol. In relation to alcohol consumption, 12.5% of never-smokers were non-drinkers in contrast to 6.4% of ever-smokers. Among those who reported drinking <60g/day the median alcohol consumption (g/day) was 14.2 (IQR: 8.7–22.1) among never smokers and 19.0 (IQR: 11.9–29.2) among ever smokers. Supplementary Table 1 provides further descriptive characteristics of the cohort in relation to age started smoking, thelarche, parity, menopausal status, and BMI.

[TABLE 1 here]

During 788,361 person-years (median 6.6 years; mean 7.7 years) of follow-up 1815 invasive breast cancers were diagnosed, of which 1813 were confirmed through national cancer registration or medical records, and the remaining two were self-reported with treatments that imply breast cancer. ER-status data were available for 99.3%, and of these 83.7% were ER-positive. Invasive ductal carcinoma accounted for 78.8%, and lobular 16.4%, of tumours. Further descriptive characteristics of the breast cancer cases are given in Supplementary Table 2.

185 The HR for invasive breast cancer in relation to ever smoking was 1.17 (95% CI: 1.07–1.29;
186 *P*=0.0009) when adjusted only for attained age, 1.13 (95% CI: 1.03–1.24; *P*=0.012) when also
187 adjusted for alcohol consumption, and 1.14 (95% CI: 1.03–1.25; *P*=0.010) when further adjusted for
188 other potentially confounding variables (see Methods and Table 2). All subsequent results are
189 adjusted for attained age, alcohol consumption and the potentially confounding variables, unless
190 otherwise stated.

191 Table 2 presents results for breast cancer overall and by ER status. The HR for ever-smoking
192 was raised for ER-positive (HR=1.12; 95% CI: 1.01–1.24; *P*=0.035) and ER-negative (HR=1.25; 95% CI:
193 0.99–1.58; *P*=0.063) breast cancer, and the difference between the HRs was not significant (*P*=0.40).
194 Breast cancer risk increased significantly with number of cigarettes smoked per day for all breast
195 cancer (*P-trend*=0.0060) and for ER-positive tumours (*P-trend*=0.023). Breast cancer risks were
196 raised significantly after 10+ years duration of smoking (10+ years vs never-smoking: *P*=0.0004).
197 Breast cancer risks did not further rise beyond 10 years duration and because of this non-linear
198 relationship there was no significant linear trend with duration of smoking (*P-trend*=0.24), nor was
199 there significant heterogeneity in the trend by ER status. Pack-years of smoking was associated with
200 breast cancer risk overall (*P-trend*=0.0069) and ER-positive breast cancer (*P-trend*=0.024) but not for
201 ER-negative (*P-trend*=0.16) tumours; there was no significant heterogeneity of the pack-years trend
202 by ER status (*P*=0.66).

203 [TABLE 2 here]

204 The HR within the year after smoking cessation was 2.68 (95% CI: 1.60–4.46), based on 15
205 cases, but for reasons described in Methods this risk period was assigned for further analyses to the
206 'current-smoker' group. On this basis risk of breast cancer was raised in current (HR=1.12; 95% CI:
207 0.89–1.39; *P*=0.34) and former (HR=1.14; 95% CI: 1.03–1.26; *P*=0.011) smokers although only the
208 latter reached statistical significance; there was no significant heterogeneity by ER status. Breast
209 cancer risks were significantly raised within the first 20 years after cessation of smoking and

decreased with greater time since cessation although the trend was not significant (*P-trend*=0.071) and there was no significant heterogeneity in this trend by ER status.

There was significant variation in risk of breast cancer by age at start of smoking (Table 3) (*P-heterogeneity*=0.018; not presented in Table 3). Breast cancer risk was significantly increased if smoking started at ages <17 (HR= 1.24; 95% CI: 1.08–1.43; *P*=0.0023) or 17–19 (HR= 1.15; 95% CI: 1.01–1.31; *P*=0.030) years relative to non-smokers, but not if it started at older ages. The risk was significantly increased for ER-positive, only for smokers starting at ages <17 years, and no significant risk increase was noted for ER-negative breast cancer. When adjusted for pack-years the breast cancer risk for starting smoking at ages <17 years was (HR= 1.12; 95% CI: 0.96–1.32; *P*=0.14), and when adjusted for duration of smoking it was (HR=1.16; 95% CI: 0.96–1.40; *P*=0.11) (not presented in Table 3).

[TABLE 3 here]

In our questionnaire we asked women only about the amount they smoked per day beginning at age 16; therefore we could not examine smoking intensity at younger ages. There was no significant trend in breast cancer risk, however, in relation to cigarettes smoked per day at ages 16–24 years. Relative to age at menarche, breast cancer risks were highest if smoking started at or before menarche (HR=1.40; 95% CI: 0.98–1.99; *P*=0.061) or 1–4 years after (HR=1.23; 95% CI: 1.07–1.41; *P*=0.0040), with a significant downward trend in breast cancer risk with increasing interval from age at menarche to age at starting smoking (*P*=0.031). A similar pattern was seen for ER-positive, but was less clear for ER-negative, breast cancer. A weaker relationship was seen with age at thelarche (e.g. 1–4 years after thelarche (HR=1.17; 95% CI: 1.00–1.37; *P*=0.056)). When adjusted for pack-years of smoking the HRs for age started smoking 1-4 years after menarche (HR=1.12; 95% CI: 0.96–1.31; *P*=0.15) or thelarche (HR=1.05; 95% CI: 0.88–1.25; *P*=0.59) were attenuated (not presented in Table 3). There was a comparable attenuation after adjusting for duration of smoking. Among parous women there was a significant trend in breast cancer risk with

interval from starting smoking to birth of first child (*P-trend*=0.013); for an interval of 15+ years the HR was 1.46 (95% CI:1.18–1.81; *P*=0.0005). However, these results were not adjusted for age at first child birth and parity (not in Tables), and when we adjusted (as shown in Table 3) there were no significantly raised HRs, or trends for all breast cancer or by ER status.

When analysed by morphological type (Supplementary Table 3) we found significant associations for ductal breast cancer similar to the results for breast cancer overall, and generally non-significant results for lobular breast cancer. There were no significant interactions by morphological type in the risk of breast cancer with smoking.

There was no raised risk of breast cancer with ever-smoking in non-drinkers (HR=0.97; 95% CI: 0.61–1.52; *P*=0.89) but a significantly raised breast cancer risk in those who had ever been drinkers (HR=1.18; 95% CI: 1.07–1.30; *P*=0.0010) although the difference in HRs was not significant (*P-interaction*=0.41) (Table 4). When further stratified by amount of alcohol consumed the HRs for ever-smoking among current drinkers remained raised. Results were similar when we examined breast cancer risk by drinking status for former smokers relative to never smokers (Supplementary Table 4).

[TABLE 4 here]

We examined further potential risk factor interactions with smoking but found no significant interactions with parity (*P*=0.095) although for nulliparous ever smoking women there was a statistically significantly increased risk of breast cancer (*P*=0.012) (Supplementary Table 5), or menopausal status (*P*=0.73) (Supplementary Table 6), although while the hazard ratio of pre-menopausal ever smokers was somewhat larger than for post-menopausal ever smokers, the former did not reach statistical significance (*P*=0.088), while the latter did (*P*=0.040). Nor did we find significant interactions with birth cohort (*P*=0.092), or BMI at age 20 (*P*=0.55) or post-menopausal ages (*P*=0.26), but we did see a significant interaction with family history of breast cancer (*P*=0.038). There were significant interactions between family history and age at starting smoking (*P*=0.0029)

and starting smoking relative to age at menarche (*P*=0.0001) in relation to risk of breast cancer (Table 5). In particular, among women with a family history of breast cancer, HRs were raised if smoking started at age 20+ years (HR=1.56; 95% CI: 1.17–2.10; *P*=0.0028) or <20 years (HR=1.26; 95% CI: 1.02–1.56: *P*=0.029), and if started 5+ years after menarche (HR=1.53; 95% CI: 1.22–1.91; *P*=0.0002), and we note these were somewhat different to the results among women without a family history of breast cancer.

[TABLE 5 here]

DISCUSSION

In the Generations Study cohort we found significant but modestly raised risk of invasive breast cancer in ever and former smokers, in women who smoked more than five cigarettes a day, had 10+ pack-years of use, or had stopped for less than 20 years. Some previous studies have reported similar associations with smoking [8-17, 20], cigarettes per day [9-11, 19], pack-years [9-13, 17-19, 26-29], and cessation [8, 12, 19, 26, 28], but not all studies find these associations [10, 11, 13, 15-17, 19, 20, 29, 30]. We saw significantly raised risk with 10+ years duration of smoking, but no increasing trend beyond 10+ years. Increased risks at long durations (or significant trends) have previously been reported in some studies [8-13, 18-20, 26-28], although some classified non-smokers as smokers with zero duration [12, 20, 26, 28] and this may artefactually produce a significant trend which partly or wholly reflects the difference in risk between non-smokers and smokers (but this may not be the only reason for an association with 20+ years (long duration) of smoking).

We found risk was significantly raised in former smokers, as has been previously reported [8-15, 20]. Risk was also raised with current smoking but the numbers of current smokers in our cohort was small and this result did not reach statistical significance, although some other studies have reported significantly raised risks in this group [8-19]. The raised risks for current and former smokers were similar (HR 1.12 and 1.14) and the confidence intervals overlapped, suggesting, within

285 our cohort, no material difference between current and former smokers in relation to breast cancer
286 risk.

287 **Breast cancer sub-types**. We found significant raised risks for ER-positive and ductal breast
288 cancer, which were the most common types in our study, but no significant heterogeneity by ER-
289 status or morphological type of the breast cancer in relation to smoking. The statistical power to
290 examine differences by ER-status or morphology was low in our cohort because of the relative
291 uncommonness of ER-negative and non-ductal type tumours. Some studies have tended to find
292 stronger risks for ER-positive breast cancer [12, 16, 20, 31] but none have found significant
293 interactions and the literature is inconclusive [2]. We observed larger HRs for smoking and pre-
294 menopausal, relative to post-menopausal, breast cancer but the former did not reach statistical
295 significance, and although the literature is variable it does in general suggest a greater relative risk
296 among pre-menopausal women [1, 2]. However, we found no evidence for a significant interaction
297 with menopausal status, similar to other studies [8, 11, 32].

298 **Confounding by alcohol**. Alcohol consumption was associated with smoking and is itself a
299 known risk factor for breast cancer [7]. We adjusted for alcohol intake and although this reduced
300 the strength of the association between smoking and breast cancer (from HR=1.17 to 1.14) the
301 association remained raised and significant. There is, however, concern that statistical adjustment
302 using self-reported alcohol consumption may not be adequate to control fully for confounding by
303 alcohol [7] so to explore further the extent of potential confounding we stratified by alcohol
304 consumption (Table 4). Within each stratum of consumption (<20g/day, 20–40g/day, and 40–
305 <60g/day) the difference in self-reported alcohol intake between never and ever smokers was
306 ≈1g/day, and we calculate this difference in consumption would be associated with <1% change in
307 relative risk of breast cancer (using the alcohol-breast cancer estimate of relative risk from a large
308 collaborative re-analysis [7]). Within each of these strata it would require ever-smokers to be
309 drinking 20g/day more than never-smokers to produce a difference of ≈15% (similar to the 12–17%

we saw). This implies the association we observed between ever-smoking and breast cancer may be too large to be explained by differences in alcohol intake alone.

We saw no significant association between smoking and breast cancer risk among non-drinkers, in concordance with a collaborative re-analysis of 43 case-control and 10 cohort studies [7], the American Cancer Society's CPS II cohort [16], and a subsequent pooled analysis of 14 cohort studies [8]. It is possible there may be synergistic interaction between ever-smoking and alcohol consumption, and risk of breast cancer, although only one study has reported the interaction as statistically significant [8]. There is some precedent to invoke synergism between smoking and alcohol because, for example, there is an established positive interaction between these two exposures and the aetiology of head and neck cancers [33]. However, non-drinking may occur for cultural or religious reasons, or because of underlying illness or other health issues, and in the UK at least non-drinkers are a minority group; therefore this potential interaction could be a reflection of a particular distribution of breast cancer risk factors among non-drinkers (and inadequate control for confounding among drinkers). Conversely, three other cohort studies found significantly raised risk among non-drinkers [18, 26, 29], although in two the increased raised risks were only in subgroups [26, 29].

Smoking in adolescence. Based on epidemiological considerations and animal studies the period from puberty to first birth may represent a window of particular susceptibility to breast cancer [34-37]. At puberty the breast is made up of mainly undifferentiated terminal ductal and lobular structures which animal studies show are sensitive to chemical carcinogenesis [34]. At these young ages ionizing radiation exposure also increases risk of breast cancer [37], especially if exposure is within six months of menarche [38]. We found risk of breast cancer in ever-smokers was greatest if smoking started at ages <17 years, or started at peri-menarcheal or, more weakly, peri-thelarcheal ages. A number of other studies have also found raised risks if smoking started in adolescence [8-13, 16-18, 20, 26, 28, 29, 32] or around menarche [11, 16, 26]. However, when we adjusted for pack-years of smoking the raised risks for starting smoking close to age at menarche or

thelarche were somewhat attenuated suggesting over-adjustment (because of possible correlation between age starting smoking and pack-years) or confounding by pack-years. Previous studies have not made this adjustment so the relative importance of early initiation or pack-years of use remains unclear.

Smoking before first childbirth . Young age at first birth and increasing parity confer long-term protection against breast cancer [34, 35] and animal models point to terminal differentiation of breast tissue at full term pregnancy being important in this process [34-36]. Increased risks have been reported for invasive breast cancer if smoking started before first childbirth [8-11, 16, 17, 20, 26, 28, 29, 32] but we found the association was only significant if we did not adjust for age at first pregnancy. A number of previous studies have adjusted for age at first pregnancy and still found significant associations with interval to first birth [8, 9, 11-13, 16-18, 20, 26, 28, 29] however it is difficult to determine the adequacy of adjustment. For example, in a large pooled analysis of 14 cohort studies there was a strong trend with smoking interval before first birth after adjustment for potential confounders that included age at first birth and number of live births (P=0.0000002) whereas after stratification by age at first birth the trends in each strata were weaker (P=0.12, 0.02, and 0.28) [8], which is suggestive of confounding.

Interaction with family history. We found the association between smoking and breast cancer was significantly larger among women with a family history of the disease than those without. Five previous studies have reported on this interaction with family history. Two studies reported no significant interaction but did not present stratified results so we cannot determine if the direction of interaction support or contradict our finding [16, 19]. Three studies reported significant interactions, with one showing increased breast cancer risk with smoking only among those with a positive family history [39], whereas two found breast cancer risk was raised only among those with no family history [15, 18]. Increased risk of breast cancer with smoking has also been seen in some [40, 41], but not all (see review [1] and a large meta-analysis [41]), studies of *BRCA1/2* carriers (or by proxy, women with three or more first degree relatives with breast or

362 ovarian cancer[42]). There are also reports of significant interactions with smoking and
363 polymorphisms in carcinogen metabolism genes *NAT2* [43] and *CYP1A1* [44, 45] and breast cancer
364 susceptibility SNPs [46, 47]. Moreover, BRCA1 and BRCA2 proteins are involved in the repair of DNA
365 damage and it is therefore possible that BRCA1/2 carriers may be more sensitive to effects of
366 carcinogens in cigarette smoke. Thus, despite the limited and inconsistent literature, is it possible
367 there are gene-smoking interactions in relation to breast cancer risk (as there is, for instance, with
368 bladder cancer [48]) and studies may benefit from focusing on more detailed measures and timing of
369 exposure (e.g. peri-menarcheal smoking or pack-years of use) rather than just ever/never smoking.

370 As in previous studies we excluded from analysis women with prevalent breast or other
371 malignant cancer [11-13, 15-17, 20, 28, 32] or prevalent *in-situ* breast cancer [13] at recruitment,
372 restricted the analysis to invasive breast cancer [7-18, 20, 26, 28, 30], and adjusted for menopausal
373 status and BMI [8, 10, 11, 13, 16, 18-20, 26, 30, 31], potential confounding variables that may also be
374 influenced by smoking. There was little scope for bias from unascertained mortality or exits, or
375 erroneous reporting of breast cancer, because follow-up for vital and breast cancer status was
376 obtained for 99% of participants and confirmation of reported breast cancers for over 99%. Our
377 smoking information was gained at recruitment and from follow-up questionnaire six years later,
378 and we were able to update smoking status, so that women who gave up smoking were classified as
379 former smokers from that point in time. Only a small number of other cohort studies [13, 16, 20]
380 have been able to update smoking exposure through follow-up. One limitation of our study is that
381 we have no direct information on passive (second hand) smoking and therefore our risk estimates
382 might be underestimated if never-smokers were exposed to passive smoking and if this exposure
383 affects risk of breast cancer [49].

384 If our results are not due to chance, residual confounding, or unidentified bias, they suggest
385 certain biologic mechanisms deserve further attention, e.g., those involving exposure at peri-
386 menarcheal ages, and gene-environment interactions, either of which may be the direct result of

chemical carcinogenesis or an indirect consequence on hormonal pathways during this susceptible period of breast development.

CONCLUSIONS

We found that smoking was associated with a modest but significantly increased risk of breast cancer, particularly among those who started at adolescent or peri-menarcheal ages, and the relative risk of breast cancer associated with smoking was significantly greater for women with a family history of the disease.

LIST OF ABBREVIATIONS

Body Mass Index, BMI

Confidence Interval, CI

Estrogen receptor, ER

Hazard Ratio, HR

Inter-Quartile Range, IQR

Menopausal Hormone Therapy, MHT

Oral Contraceptive, OC

DECLARATIONS

Ethics approval and consent to participate:

The study was approved by the South Thames Multicentre Research Ethics Committee (ref: MREC 03/01/014) and participants provided informed consent.

Consent for publication:

Not applicable.

Availability of data and material:

The datasets generated during and/or analysed during the current study are not publicly available due to confidentiality reasons but anonymised versions may be available from the corresponding author on reasonable request.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This work was funded by Breast Cancer Now and The Institute of Cancer Research. We acknowledge National Health Service (NHS) funding to the Royal Marsden NHS Foundation Trust and The Institute of Cancer Research National Institute for Health Research (NIHR) Biomedical Research Centre. The funders did not participate in the design of the study, analysis, or interpretation of data, nor in writing or approving the manuscript.

Authors' contributions:

AJS and AA designed and obtained funding for the Generations Study. AJS, MEJ and MJS set up and collected data in the Generations Study. MEJ, MJS and LW collected and prepared data for the analysis. MEJ conducted the analyses and drafted the manuscript. All authors contributed to data interpretation and preparation of the final manuscript. All authors read and approved the final manuscript.

Acknowledgements:

We thank Breast Cancer Now and The Institute of Cancer Research for support of the Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study.

Authors' information (optional):

Not applicable.

REFERENCES

1. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans: **Personal habits and indoor combustions. Volume 100 E. A review of human carcinogens**. *IARC monographs on the evaluation of carcinogenic risks to humans* 2012, **100**(Pt E):1-538.
2. US Department of Health and Human Services. In: *The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General.* edn. Atlanta (GA); 2014.
3. Morabia A: **Smoking (active and passive) and breast cancer: epidemiologic evidence up to June 2001**. *Environ Mol Mutagen* 2002, **39**(2-3):89-95.
4. Johnson KC, Miller AB, Collishaw NE, Palmer JR, Hammond SK, Salmon AG, Cantor KP, Miller MD, Boyd NF, Millar J *et al*: **Active smoking and secondhand smoke increase breast cancer**

risk: the report of the Canadian Expert Panel on Tobacco Smoke and Breast Cancer Risk (2009). *Tob Control* 2011, **20**(1):e2.
5. Terry PD, Rohan TE: **Cigarette smoking and the risk of breast cancer in women: a review of the literature**. *Cancer Epidemiol Biomarkers Prev* 2002, **11**(10 Pt 1):953-971.
6. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans: **Tobacco smoke and involuntary smoking**. *IARC monographs on the evaluation of carcinogenic risks to humans* 2004, **83**:1-1438.
7. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., Coates RJ, Liff JM, Talamini R, Chantarakul N *et al*: **Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease**. *Br J Cancer* 2002, **87**(11):1234-1245.
8. Gaudet MM, Carter BD, Brinton LA, Falk RT, Gram IT, Luo J, Milne RL, Nyante SJ, Weiderpass E, Beane Freeman LE *et al*: **Pooled analysis of active cigarette smoking and invasive breast cancer risk in 14 cohort studies**. *Int J Epidemiol* 2016.
9. Bjerkaas E, Parajuli R, Weiderpass E, Engeland A, Maskarinec G, Selmer R, Gram IT: **Smoking duration before first childbirth: an emerging risk factor for breast cancer? Results from 302,865 Norwegian women**. *Cancer Causes Control* 2013, **24**(7):1347-1356.
10. Catsburg C, Miller AB, Rohan TE: **Active cigarette smoking and risk of breast cancer**. *Int J Cancer* 2015, **136**(9):2204-2209.
11. Dossus L, Boutron-Ruault MC, Kaaks R, Gram IT, Vilier A, Fervers B, Manjer J, Tjonneland A, Olsen A, Overvad K *et al*: **Active and passive cigarette smoking and breast cancer risk: results from the EPIC cohort**. *Int J Cancer* 2014, **134**(8):1871-1888.
12. Luo J, Margolis KL, Wactawski-Wende J, Horn K, Messina C, Stefanick ML, Tindle HA, Tong E, Rohan TE: **Association of active and passive smoking with risk of breast cancer among postmenopausal women: a prospective cohort study**. *BMJ* 2011, **342**:d1016.
13. Xue F, Willett WC, Rosner BA, Hankinson SE, Michels KB: **Cigarette smoking and the incidence of breast cancer**. *Arch Intern Med* 2011, **171**(2):125-133.
14. Macacu A, Autier P, Boniol M, Boyle P: **Active and passive smoking and risk of breast cancer: a meta-analysis**. *Breast Cancer Res Treat* 2015, **154**(2):213-224.
15. Nyante SJ, Gierach GL, Dallal CM, Freedman ND, Park Y, Danforth KN, Hollenbeck AR, Brinton LA: **Cigarette smoking and postmenopausal breast cancer risk in a prospective cohort**. *Br J Cancer* 2014, **110**(9):2339-2347.
16. Gaudet MM, Gapstur SM, Sun J, Diver WR, Hannan LM, Thun MJ: **Active smoking and breast cancer risk: original cohort data and meta-analysis**. *J Natl Cancer Inst* 2013, **105**(8):515-525.
17. Gram IT, Little MA, Lund E, Braaten T: **The fraction of breast cancer attributable to smoking: The Norwegian women and cancer study 1991-2012**. *Br J Cancer* 2016, **115**(5):616-623.
18. Reynolds P, Hurley S, Goldberg DE, Anton-Culver H, Bernstein L, Deapen D, Horn-Ross PL, Peel D, Pinder R, Ross RK *et al*: **Active smoking, household passive smoking, and breast cancer: evidence from the California Teachers Study**. *J Natl Cancer Inst* 2004, **96**(1):29-37.
19. Cui Y, Miller AB, Rohan TE: **Cigarette smoking and breast cancer risk: update of a prospective cohort study**. *Breast Cancer Res Treat* 2006, **100**(3):293-299.
20. Al-Delaimy WK, Cho E, Chen WY, Colditz G, Willet WC: **A prospective study of smoking and risk of breast cancer in young adult women**. *Cancer Epidemiol Biomarkers Prev* 2004, **13**(3):398-404.
21. Swerdlow AJ, Jones ME, Schoemaker MJ, Hemming J, Thomas D, Williamson J, Ashworth A: **The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology**. *Br J Cancer* 2011, **105**(7):911-917.
22. Cox DR: **Regression Models and Life-Tables**. *J R Stat Soc B* 1972, **34**(2):187-220.
23. Lunn M, McNeil D: **Applying Cox regression to competing risks**. *Biometrics* 1995, **51**(2):524-532.

24. Buse A: **The Likelihood Ratio, Wald, and Lagrange Multiplier Tests - an Expository Note**. *Am Stat* 1982, **36**(3):153-157.
25. StataCorp: **Stata Statistical Software: Release 14**. In., 14.0 edn. College Station, TX: StataCorp LP; 2015.
26. Gram IT, Braaten T, Terry PD, Sasco AJ, Adami HO, Lund E, Weiderpass E: **Breast cancer risk among women who start smoking as teenagers**. *Cancer Epidemiol Biomarkers Prev* 2005, **14**(1):61-66.
27. Li CI, Malone KE, Daling JR: **The relationship between various measures of cigarette smoking and risk of breast cancer among older women 65-79 years of age (United States)**. *Cancer Causes Control* 2005, **16**(8):975-985.
28. Luo J, Horn K, Ockene JK, Simon MS, Stefanick ML, Tong E, Margolis KL: **Interaction between smoking and obesity and the risk of developing breast cancer among postmenopausal women: the Women's Health Initiative Observational Study**. *Am J Epidemiol* 2011, **174**(8):919-928.
29. Gram IT, Park SY, Kolonel LN, Maskarinec G, Wilkens LR, Henderson BE, Le Marchand L: **Smoking and Risk of Breast Cancer in a Racially/Ethnically Diverse Population of Mainly Women Who Do Not Drink Alcohol: The MEC Study**. *Am J Epidemiol* 2015, **182**(11):917-925.
30. Catsburg C, Kirsh VA, Soskolne CL, Kreiger N, Rohan TE: **Active cigarette smoking and the risk of breast cancer: a cohort study**. *Cancer epidemiology* 2014, **38**(4):376-381.
31. London SJ, Colditz GA, Stampfer MJ, Willett WC, Rosner BA, Speizer FE: **Prospective study of smoking and the risk of breast cancer**. *J Natl Cancer Inst* 1989, **81**(21):1625-1631.
32. Egan KM, Stampfer MJ, Hunter D, Hankinson S, Rosner BA, Holmes M, Willett WC, Colditz GA, Nurses' Health S: **Active and passive smoking in breast cancer: prospective results from the Nurses' Health Study**. *Epidemiology* 2002, **13**(2):138-145.
33. Boffetta P, Hashibe M: **Alcohol and cancer**. *The Lancet Oncology* 2006, **7**(2):149-156.
34. Russo J, Mailo D, Hu YF, Balogh G, Sheriff F, Russo IH: **Breast differentiation and its implication in cancer prevention**. *Clin Cancer Res* 2005, **11**(2 Pt 2):931s-936s.
35. Colditz GA, Bohlke K, Berkey CS: **Breast cancer risk accumulation starts early: prevention must also**. *Breast Cancer Res Treat* 2014, **145**(3):567-579.
36. Colditz GA, Frazier AL: **Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus**. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 1995, **4**(5):567-571.
37. Carmichael A, Sami AS, Dixon JM: **Breast cancer risk among the survivors of atomic bomb and patients exposed to therapeutic ionising radiation**. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* 2003, **29**(5):475-479.
38. Cooke R, Jones ME, Cunningham D, Falk SJ, Gilson D, Hancock BW, Harris SJ, Horwich A, Hoskin PJ, Illidge T *et al*: **Breast cancer risk following Hodgkin lymphoma radiotherapy in relation to menstrual and reproductive factors**. *Brit J Cancer* 2013, **108**(11):2399-2406.
39. Suzuki T, Matsuo K, Wakai K, Hiraki A, Hirose K, Sato S, Ueda R, Tajima K: **Effect of familial history and smoking on common cancer risks in Japan**. *Cancer* 2007, **109**(10):2116-2123.
40. Breast Cancer Family Registry, Kathleen Cuningham Consortium for Research into Familial Breast Cancer, Ontario Cancer Genetics Network: **Smoking and risk of breast cancer in carriers of mutations in BRCA1 or BRCA2 aged less than 50 years**. *Breast Cancer Res Treat* 2008, **109**(1):67-75.
41. Friebel TM, Domchek SM, Rebbeck TR: **Modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: systematic review and meta-analysis**. *J Natl Cancer Inst* 2014, **106**(6):dju091.
42. Couch FJ, Cerhan JR, Vierkant RA, Grabrick DM, Therneau TM, Pankratz VS, Hartmann LC, Olson JE, Vachon CM, Sellers TA: **Cigarette smoking increases risk for breast cancer in high-**

risk breast cancer families. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2001, **10**(4):327-332.

43. Ambrosone CB, Kropp S, Yang J, Yao S, Shields PG, Chang-Claude J: **Cigarette smoking, N-acetyltransferase 2 genotypes, and breast cancer risk: pooled analysis and meta-analysis**. *Cancer Epidemiol Biomarkers Prev* 2008, **17**(1):15-26.

44. Ishibe N, Hankinson SE, Colditz GA, Spiegelman D, Willett WC, Speizer FE, Kelsey KT, Hunter DJ: **Cigarette smoking, cytochrome P450 1A1 polymorphisms, and breast cancer risk in the Nurses' Health Study**. *Cancer Res* 1998, **58**(4):667-671.

45. Saintot M, Malaveille C, Hautefeuille A, Gerber M: **Interactions between genetic polymorphism of cytochrome P450-1B1, sulfotransferase 1A1, catechol-o-methyltransferase and tobacco exposure in breast cancer risk**. *Int J Cancer* 2003, **107**(4):652-657.

46. Barrdahl M, Rudolph A, Hopper JL, Southey MC, Broeks A, Fasching PA, Beckmann MW, Gago-Dominguez M, Castelao JE, Guenel P *et al*: **Gene-environment interactions involving functional variants: Results from the Breast Cancer Association Consortium**. *Int J Cancer* 2017.

47. Rudolph A, Milne RL, Truong T, Knight JA, Seibold P, Flesch-Janys D, Behrens S, Eilber U, Bolla MK, Wang Q *et al*: **Investigation of gene-environment interactions between 47 newly identified breast cancer susceptibility loci and environmental risk factors**. *Int J Cancer* 2015, **136**(6):E685-696.

48. Garcia-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, Hein DW, Tardon A, Serra C, Carrato A, Garcia-Closas R *et al*: **NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses**. *Lancet* 2005, **366**(9486):649-659.

49. Johnson KC: **Accumulating evidence on passive and active smoking and breast cancer risk**. *Int J Cancer* 2005, **117**(4):619-628.