Explaining variance in the *Cumulus* mammographic measures that predict breast cancer risk: a twins and sisters study

Tuong L. Nguyen[1], Daniel F. Schmidt[1], Enes Makalic[1], Gillian S. Dite[1], Jennifer Stone[1], Carmel Apicella[1], Minh Bui[1], Robert J. MacInnis[1,2], Fabrice Odefrey[3], Jennifer N. Cawson[4], Susan A. Treloar[5], Melissa C. Southey[3], Graham G. Giles[1,2], John L. Hopper[1]

1. Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Carlton VIC 3053, Australia
2. Cancer Epidemiology Centre, Cancer Council Victoria, Carlton VIC 3053, Australia
3. Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Parkville VIC 3010, Australia
4. Department of Medicine, St Vincent's Hospital, The University of Melbourne, Fitzroy VIC 3056, Australia
5. The University of Queensland, Centre for Military and Veterans' Health, Brisbane, Australia.

Running title

Variance in mammographic density measures that predict breast cancer

Key words

breast cancer, mammographic density, sisters, twins, variance components

Financial Support

This study was supported by the National Health and Medical Research Council (NHMRC) of Australia, the National Breast Cancer Foundation, the Victorian Breast Cancer Research Consortium (VBCRC), Cancer Australia, the Victorian Health Promotion Foundation and the NSW Cancer Council, the Cooperative Research Centre for Discovery of Genes for Common Human Diseases, Australia (1997-2004) under the Australian Government's Cooperative Research Centres program, and Cerylid Biosciences Ltd. John Hopper is a NHMRC Senior Principal Research Fellow, and Melissa Southey is a NHMRC Senior Research Fellow.

Corresponding Author

Prof John Hopper, Centre for Molecular, Environmental, Genetic, and Analytic (MEGA) Epidemiology, University of Melbourne, Level 3, 207 Bouverie Street, Carlton, VIC 3053 Australia. Telephone: +61 3 8344 0697. Fax: +61 3 9349 5815 Email: j.hopper@unimelb.edu.au

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Word count: 3,810

Number of tables: 5

Abstract

Background: Mammographic density (MD), the area of the mammographic image that appears white or bright, predicts breast cancer risk. We estimated the proportions of variance explained by questionnaire-measured breast cancer risk factors and by unmeasured residual familial factors.

Methods: For 544 MZ and 339 DZ twin pairs, and 1,558 non-twin sisters from 1,564 families, MD was measured using the computer-assisted method *Cumulus*. We estimated associations using multi-level mixed-effects linear regression and studied familial aspects using a multivariate normal model.

Results: The proportions of variance explained by age, body mass index (BMI), and other risk factors, respectively, were: 4%, 1% and 4% for dense area; 7%, 14% and 4% for percent dense area; and 7%, 40% and 1% for non-dense area. Associations with dense area and percent dense area were in opposite directions than for non-dense area. After adjusting for measured factors, the correlations of dense area with percent dense area and non-dense area were 0.84 and −0.46, respectively. The MZ, DZ and sister pair correlations were: 0.59, 0.28 and 0.29 for dense area; 0.57, 0.30 and 0.28 for percent dense area; and 0.56, 0.27 and 0.28 for non-dense area (standard error (SE) = 0.02, 0.04 and 0.03, respectively). **Conclusions**: Under the classic twin model, 50–60% (SE = 5%) of the variance of MD measures that predict breast cancer risk are due to undiscovered genetic factors, and the remainder to as yet unknown individual-specific, non-genetic factors. **Impact**: Much remains to be learnt about the genetic and environmental determinants of MD.

Introduction

Mammographic density (MD) is the area of the mammographic image of the breast that appears white or bright. The state-of-the-art method for measuring MD is a computer-assisted thresholding technique called *Cumulus*. This measures the total area of the breast image and the absolute area covered by dense tissue, as determined by the viewer, called dense area. From these measures, the absolute area of the breast image covered by non-dense tissue and the percentage of breast image covered by dense tissue is easily obtained.

Several case-control studies nested within cohorts of women attending mammographic screening services have shown that various measures of MD at recruitment (baseline) predict subsequent risk of breast cancer (1, 2). These studies have virtually always matched cases and controls on age at mammogram, and have adjusted for breast cancer risk factors measured at baseline. They have found that, for women of the same age, body mass index (BMI) and other measured risk factors for breast cancer, those with a greater dense area (either absolutely or as a percentage) are at greater risk of breast cancer.

Percent dense area is negatively associated with age, and even more so with BMI, yet age and BMI are positively associated with breast cancer risk in the age groups typically studied. Thus, when considering percent dense area as a risk factor for breast cancer, its associations with BMI and age must be properly taken into account. While it is often reported that women with high (≥75%) percent dense area have a 4-fold to 6-fold increased risk of breast cancer compared with women with primarily fatty breasts (percent dense area ≤10%) (3), it is rarely made explicit that these comparisons refer to *women of the same age and BMI*. Moreover, as we have shown (4), dense area and percent dense area adjusted for age and BMI are highly correlated ($r \sim 0.9$). Consequently, the MD measures that (best) predict breast cancer risk are those of dense area adjusted for age, BMI and other breast cancer risk factors.

Recent studies have identified that non dense area might also be associated with risk even after adjusting for dense area or percent dense area, age and BMI (non-dense area is highly correlated with BMI), but in the opposite direction to dense area and percent dense area. The correlation of non-dense

area with dense area is around −0.3, which raises the possibility that the risk associations in opposite directions with dense area and non-dense area might be at least in part a consequence of the same underlying phenomenon (5). The issue is also complicated by the fact that the association of BMI with breast cancer risk is not constant with respect to age at baseline, or age at diagnosis. After adjusting for BMI as a function of age, Baglietto and colleagues found that the linear combination of dense area and percent dense area that best predicted breast cancer risk was dense area −0.24 non dense area (5).

Given that various MD measures predict future occurrences of breast cancer, it is important to identify the factors that determine their mean values and quantify how much they explain their variation. In this regard, it has been found from twin and family studies that MD measures are correlated in relatives (6-8), so part of their variances must be due to familial, if not genetic, factors.

Here we have conducted a large cross-sectional study of female twin pairs, both genetically identical (monozygotic, MZ) and fraternal (dizygotic, DZ), and their sisters. We have estimated the means of the MD measures as functions of the breast cancer risk factors measured by questionnaire, taking into account that the women are from families. The adjusted measures are therefore the MD measures that predict breast cancer risk, independent of the other risk factors. We have then used this powerful study design to obtain insights about, and estimates of, the roles of both genetic and non-genetic factors in explaining the variances of the MD measures that predict breast cancer risk.

Methods

Participants

Participants were from the Australian Mammographic Density Twins and Sisters Study (AMDTSS), details of which are provided in Odefrey *et al*., (4), the Genes Behind Endometriosis Study (GBES; see Treloar *et al*. (9), the Australian Breast Cancer Family Study (ABCFS) (10) and volunteers from the Breast Cancer Network Australia (BCNA) and other sources. Briefly, female twin pairs aged 40–70 years without a prior diagnosis of breast cancer were recruited through the Australian Twin Registry. Participating twins completed a questionnaire and gave permission to access their mammograms. They were also asked to seek the permission from any eligible sisters to be invited to participate in the study. We recruited 3,324 twins and sisters from 1,564 families, including 544 MZ and 339 DZ twin pairs and 1,558 non-twin sisters. Of these, 2,345 were from the AMDTSS, 788 from the GBES, 71 from the ABCFS, and 120 from the BCNA and other sources. The study was approved by the Human Research Ethics Committee (HREC) of the University of Melbourne.

Epidemiology Questionnaire

Telephone administered questionnaires were used to recorded demographic information and self-reported weight, height, smoking history, alcohol consumption, reproductive history, cessation of menstruation, use of oral contraceptives, use of hormone replacement therapy and personal and family history of cancer. A woman was defined as postmenopausal if she: had had a hysterectomy, both ovaries removed, or radiation; was not on hormone replacement therapy at the time of the mammogram and had not menstruated 12 months prior; or was on hormone replacement therapy at the time of mammogram and had not menstruated 12 months prior and was not menstruating before commencing hormone replacement therapy. Subjects not fitting these criteria were considered premenopausal. For twin pairs, zygosity was determined by a standard question that describes the differences between MZ and DZ twin pairs and has been shown to give 95% agreement with true zygosity (11).

MD Measurements

All available episodes of mammograms were retrieved with the participants' written consent, mostly from Australian *BreastScreen* services, but also from private clinics, and private hospitals. We also retrieved mammograms from the participants themselves. The cranio-caudal views for left and right breasts were selected and digitized by using the Lumysis 85 scanner at Australian Mammographic Density Research Facility. For each woman, the most recent right breast cranio-caudal view was selected for MD measurement and the left breast cranio-caudal view was selected if the right breast mammogram was missing or unavailable. Mammographic measurements of total area and dense area were performed using Cumulus 4.0, a computer-assisted thresholding technique, after randomization and blind to information, by three independent operators (J.S., F.O., and T.L.N.) with high repeatability (4). Non dense area and percent dense area were calculated from these measures.

Statistical Methods

Associations between variables measured by questionnaire and the means of the MD measures were estimated using ordinary linear regression modeling under the assumption that the residuals were normally distributed, though correlated within families. The Box-Cox procedure was used to test the normality of the distributions of the MD measures and, if necessary, select an appropriate power transformation. Consequently, dense area was cube root transformed, while percent dense area and non-dense area were square root transformed. All questionnaire measures were inspected for missing or invalid values which were replaced with the average for continuous exposure variables and the most common value for binary or categorical exposure variables. The percentage of missing values was <1% for all variables except DCIS, for which it was 8% and all unknowns were coded as "no" given <1% of responders answered "yes".

We estimated the regression coefficients, β, for the associations of predictors with mean MD measures using multilevel mixed effects regression analysis and the XT-MIXED option in the Stata software package (12) as it accounted for the correlations between twins and sisters. We log transformed BMI because the associations were approximately linear with log BMI.

As in Stone et al. (1), the Bayesian information criterion (13) score was used to select the best-fitting model (not present in the table). Given the multiple factors being fitted in models, we took p = 0.005 as our nominal threshold for statistical inference.

To quantify the amount of variance explained by the questionnaire-measured variables, all the MD measurements and questionnaire-measured variables were standardized by the formula: $\frac{(\mathbf{x} - \bar{\mathbf{x}})}{\mathbf{sd}}$, where $\bar{\mathbf{x}}$ is the mean and sd is the standard deviation. Consequently, for each estimated regression coefficient, β^*, $(\beta^*)^2$ approximates the amount of variance explained by fitting that variable with all other variables in the model held constant.

To estimate the correlation between pairs of relatives and to fit a variance components model, we applied multivariate Gaussian regression using the software FISHER with inference based on asymptotic likelihood theory (14, 15). This approach assumes that, after adjusting the mean for specified measured variables, the family residuals follow a multivariate normal distribution with a covariance structure that can be parameterised. It allows estimation of correlations separately for MZ and DZ twin pairs, or for pairs of non-twin sisters (including a twin and her sister), and statistical comparisons.

We also fitted models estimating independent genetic and environmental components of variance that represent additive genetic factors (A), environment factors shared by twins and sisters (C), and individual specific environmental factors and measurement error (E), where A+C+E = total residual variance (V). MZ pairs share all their genes while DZ pairs and sister pairs share on average half their genes, so the correlation in additive genetic factors is 1.0 for MZ pairs and 0.5 for DZ and sister pairs (16). Under the assumption that the effects of environmental factors shared by twins and sisters are independent of zygosity and the same for twins and sisters, the correlation between a pair will be $(2\phi_{ij}A+C)/V$ where $2\phi_{ij} = 1$ if MZ else 0.5.

Results

Table 1 shows the characteristics of the 3,324 participants (544 MZ and 399 DZ twin pairs, and 1,558 of their non-twin sisters) based on the questionnaire, and their MD measures. There was no evidence that, after adjusting for the covariates below, the means or proportions differed depending on whether the woman was an MZ or DZ twin, or a non-twin. The absolute within-pair difference in age or time between mammograms was 1.34 years for MZ and 1.67 years for DZ twins pairs, and there was no significant difference between MZ and DZ pairs (all P>.05).

Table 2 shows that, 22% of the families had one member, 54% had two members, 18% had three members, 5% had four members, and the remainder had five, six or seven members. The majority of families (57%) contained one twin pair. There were a total 1,483 sister-sister pairings (including sister-twin pairings) that were not independent within families.

Table 3 shows that, univariately, cube root dense area was negatively associated with age at mammogram (6.8%), log BMI (1.7%), age at menopause (3.6%), number of live births (2.9%; the percentage of variance explained by each factor, $(\beta^*)^2$ is shown in brackets). When fitted concurrently, the associations with age at mammogram, BMI, age at menopause and number of live births remained nominally significant but, given that these factors were correlated with one another, the percentages of variance explained was approximately halved to 4.0%, 1.0%, 1.0% and 1.0%, respectively.

After adjusting for the above negative associations, cube root dense DA was positively associated with current use of HRT, years of alcohol consumption, having a benign breast lump removed and having ductal carcinoma *in situ*, and negatively associated with years of oral contraceptive use, explaining 0.3%, 0.4%, 0.5%, 0.3% and 0.4% of variance, respectively. Overall, these measured factors explained ~9% of total variance.

Table 4 shows that, univariately, square root percent dense area was negatively associated with age at mammogram (12.3%), log BMI (16.8%), age at menopause (6.8%), number of live births (4.0%). When fitted concurrently, the associations with age at mammogram, BMI, age at menopause and number of live births remained nominally significant and the percentages of variance explained reduced to 7.3%, 14.4%, 0.8% and 1.0%, respectively.

After adjusting for the above negative associations, square root percent dense area was positively associated with current use of HRT, years of alcohol consumption, having a benign breast lump removed and having ductal carcinoma *in situ*, and negatively associated with years of oral contraceptive use and current smoking, explaining 0.3%, 0.4%, 0.4%, 0.2%, 0.3% and 0.2% of variance, respectively. Overall, these measured factors explained ~25% of total variance.

Table 5 shows that, univariately, square root non- dense area was positively associated with age at mammogram (8.4%), log BMI (42.2%), age at menopause (4.8%), number of live births (2.3%). When fitted concurrently, the associations with age at mammogram, BMI, and number of live births remained nominally significant and the percentages of variance explained reduced to 7.3%, 39.7%, and 0.3%, respectively.

After adjusting for the above positive associations, square root non- dense area was negatively associated with years of use of hormone replacement therapy and years of alcohol consumption, and positively associated with ever smoking and having ovaries removed, explaining 0.4%, 0.2%, 0.2%, and 0.3% of variance, respectively. Overall these measured factors explained ~48% of total variance.

After adjusting for the factors above, the correlation between dense area residuals and percent dense area residuals was 0.84, and between dense area residuals and non-dense area residuals was -0.46. There was no evidence that any of the associations above, or the correlations between residuals, depended on whether the woman was an MZ or DZ twin, or a non-twin.

After adjusting for the factors, the MZ, DZ and sister pair correlations were: 0.59, 0.28 and 0.29 for dense area; 0.57, 0.30 and 0.28 for percent dense area; and 0.56, 0.27 and 0.28 for percent dense area for non-dense area, respectively. The standard errors (SE) were 0.02, 0.04 and 0.03, respectively, for all three measures. Clearly, the MZ correlations were greater than the DZ and sister pair correlations (all $p < 0.001$), and the DZ and sister pair correlations were not significantly different from one another.

The estimates for A and C, as a percentage of total residual variance, were: for dense area, 0.56 and 0.01; for percent dense area, 0.52 and 0.04; and for non-dense area, 0.64 and −0.06. The correlations between estimates of A and C were −0.80, −0.81 and −0.86, respectively. The SEs of

these estimates were all ~0.05, so the estimates of C were not significant. By a *post hoc* power calculation, we had 80% power at the 0.05 level of significance to detect values of C > 0.13. For non-dense area, the estimate of A when C is constrained to be non-zero was 0.58.

Discussion

We found that dense area and percent dense area have the same determinants. The amounts of variance explained by BMI, and to a lesser extent age at mammogram, are substantially less for dense area (4% and 1%, respectively) than for percent dense area (7% and 14%). This is an important issue because the associations with these factors are in the opposite direction to the relationship of these factors to breast cancer risk (especially for age and, for BMI, at least for post-menopausal disease and for post-menopausal women), and more so for percent dense area than dense area. After adjusting dense area and percent dense area for age and BMI, the associations with other risk factors are almost identical, and explain ~4% of variance. This is consistent with the fact that, after adjusting for all measured risk factors, the dense area and percent dense area residuals are highly correlated with one another (0.85).

In using mammographic measures to create a breast cancer risk factor, dense area and percent dense area are very similar once adjusted for age and BMI, but percent dense area is more problematic due to its much stronger association with BMI. Each step in calculating percent dense area and adjusting it for BMI and age has the potential to introduce more measurement error.

We also found that non-dense area has very similar determinants to dense area and percent dense area, but mostly in the opposite direction. The associations with age, and especially BMI, are much greater, explaining 7% and 40% of variance, respectively. After adjusting for measured factors, dense area and non- dense area are substantially, though negatively, correlated. This is interesting because, as a linear combination of dense area and non-dense area, it has been found that the best predictor of breast cancer risk is F = (dense area − 0.24 non-dense area) ; i.e. each cm^2 of dense area is four times more predictive than each cm^2 of non-dense area, and in the opposite direction. There is an intrinsic collinearity between dense area and non-dense area, whose sum is constrained to be equal to the total breast area, especially after adjusting for age and BMI. The factor F above could be representing a single phenomenon that is more common in what is considered by the observer to be dense area and therefore less common in what is considered to be non-dense area.

After adjusting for measured factors, we then considered the roles of unmeasured factors in explaining the residual MD measures. By studying MZ twin pairs we were able to estimate the maximum amount of residual variance due to familial factors, and found this was almost 60%. By studying DZ and sister pairs, we were able to test if the familial sources of variance were independent of genetic similarity, and were able to reject the null hypothesis. Note that this does not prove that a difference in correlation by zygosity is only due to the differences in shared genes by zygosity, as the MZ pairs could have shared non-genetic factors to a greater degree. In this regard, we found no evidence that DZ and sister pair correlations differed from each other, implying that the degree to which these two types of first-degree relatives share non-genetic factors relevant to the MD measures is not (substantially) different.

One can always find a non-genetic explanation for familial correlations, and in this case it would be that the MZ pairs share such factors twice as strongly as do DZ and non-twin sister pairs. But this two-fold difference is also highly consistent with the theoretical model first proposed by R.A. Fisher in 1918 (16), which predicted that this pattern would be observed if the reason why the relatives were correlated was solely due to the presence of 'additive' genetic factors.

Applying the classic twin model to our data, we predicted that about 50–60% of residual variance would be due to genetic factors. The remainder would be due to unmeasured individual specific non-familial (and therefore non-genetic) factors. The latter would include measurement error, which for these mammographic measures is not large and ~5%; e.g. a UK study found the repeatability was 0.94 for dense area, 0.91 for percent dense area and 0.96 for non-dense area (1). The former would include variants in and around genes such as *LSP1* (4, 17), *ZNF365(18)* which have been found to be associated with both dense area and percent dense area adjusted for risk factors, and with breast cancer risk itself. These recently discovered variants, however, explain in the order of ~1% or less of the residual variance.

In terms of the MD measures themselves (dense area, percent dense area and non- dense area), the likely genetic component of variance is much greater for dense area due to the fact that far less variance is explained by measured factors, mostly BMI and age. But in terms of the MD measures that predict breast cancer risk, the genetic variances are almost identical.

The fact that the breakdown of residual variance was so similar for dense area and non- dense area is not surprising, given their high correlation. But the fact that the same applied to dense area and non- dense area is intriguing, and supports the notion that – in terms of predicting breast cancer risk – dense area and non-dense area (after adjusting for age, BMI and other breast cancer risk factors) are 'two sides of the same coin'; see discussion about factor F above.

The statistical analysis approach we used is optimal in that it provides asymptotically unbiased estimates without subdividing the data into pairs, and uses all the information in the all the families, including isolated individuals. This is the strength of the likelihood approach, which produces estimates of standard errors that take into account the fact that the pairings within a family are not independent (19-21)**.** Therefore, information on the correlation between sister pairs comes from sibling pairs in which one was a twin and the other not, as well as from pairs of non-twin sisters. Information on the means (main effects) comes from all women in the data set. Comparison with data from the population-based sample of unaffected women in the ABCFS of the same age did not reveal any major differences in the general characteristics of the participants in this study. As is the case for all studies, it is difficult to exclude the possibility that participants in this twin and sister study are different from the general population in terms of lifestyle factors such as smoking, alcohol consumption, etc. However, those factors are not, or are at most weakly, associated with the mammographic density measures that predict breast cancer risk. They therefore explain at most a very small proportion of variation in these risk-predicting measures, the topic of interest for this paper.

This study supports on-going research to discover the genetic and environmental determinants of MD. The mammographic density measures that predict breast cancer risk (i.e. adjusted for age and covariates) are highly stable with age/time. The correlations are >0.8 for measures even 10 years apart (22). Therefore these familial associations are likely established at a young age, and we are currently studying the MD measures of younger adult women and their relatives, including their mothers, to gain greater insights into the genetic and environmental determinants of the MD measures that predict breast cancer.

The quest to find more genetic variants associated with MD measures that predict breast cancer risk is on-going, with two major international consortia MODE and DENSNPS (17, 18). Two

approaches are being applied. The first involves testing if the common variants being found to be associated with breast cancer risk are also associated with MD measures that predict breast cancer firs. The second involves conducting genome-wide association studies of the MD measures themselves.

The other major challenge is to find the non-familial (which implies non-genetic) factors, other than the established breast cancer risk factors measured here by questionnaire, that explain the substantial remainder of variation. This could involve new thinking about breast cancer risk as we have found that the factors measured by conventional questionnaires usually administered in mid-life explain little variance. Issues to be considered could include epigenetics, and measures of early-life environment and growth using cohorts that collected relevant measures in the past.

Acknowledgements

We would like to thank the twins and sisters who participated in this study, and the Australian Twin Registry. We would also like to thank Avis McPhee and all the volunteers from the Breast Cancer Network Australia.

References

1. Stone J, Ding J, Warren RM, Duffy SW, Hopper JL. Using mammographic density to predict breast cancer risk: dense area or percentage dense area. Breast Cancer Res 2010;12:R97.

2. Boyd NF, Martin LJ, Yaffe MJ, Minkin S. Mammographic density and breast cancer risk: current understanding and future prospects. Breast Cancer Res 2011;13:223.

3. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomark Prev 2006;15:1159-69.

4. Odefrey F, Stone J, Gurrin LC, Byrnes GB, Apicella C, Dite GS, et al. Common genetic variants associated with breast cancer and mammographic density measures that predict disease. Cancer Res 2010;70:1449-58.

5. Baglietto L. Associations of mammographic dense and non-dense area and body mass index by age at diagnosis with risk of breast cancer. Am J Epidemiol; Accepted. 2013.

6. Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, McCredie MR, et al. Heritability of mammographic density, a risk factor for breast cancer. N Engl J Med 2002;347:886-94.

7. Stone J, Dite GS, Gunasekara A, English DR, McCredie MR, Giles GG, et al. The heritability of mammographically dense and nondense breast tissue. Cancer Epidemiol Biomark Prev 2006;15:612-7.

8. Ursin G, Lillie EO, Lee E, Cockburn M, Schork NJ, Cozen W, et al. The relative importance of genetics and environment on mammographic density. Cancer Epidemiol Biomark Prev 2009;18:102-12.

9. Treloar SA, Wicks J, Nyholt DR, Montgomery GW, Bahlo M, Smith V, et al. Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26. Am J Hum Genet. 2005;77:365-76.

10. Dite GS, Jenkins MA, Southey MC, Hocking JS, Giles GG, McCredie MR, et al. Familial risks, early-onset breast cancer, and BRCA1 and BRCA2 germline mutations. J Natl Cancer Inst 2003;95:448-57.

11. Torgersen S. The determination of twin zygosity by means of a mailed questionnaire. Acta Genetic Med Gemellol 1979;28:225-36.

12. StataCorp. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP; 2009.

13. Schwarz G. Estimating the dimension of a model. Annal Stat 1978;Vol. 6:461-4.

14. Lange K, Boehnke M. Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. Am J Med Genet. 1983;14:513-24.

15. Lange K, Boehnke M, Weeks D. Programs for pedigree analysis. Los Angeles: Department of Biomathematics, University of California; 1987.

16. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. . Trans Roy Soc Edinb. 1918;52:399-433.

17. Vachon CM, Scott CG, Fasching PA, Hall P, Tamimi RM, Li J, et al. Common breast cancer susceptibility variants in LSP1 and RAD51L1 are associated with mammographic density measures that predict breast cancer risk. Cancer Epidemiol Biomark Prev 2012;21:1156-66.

18. Lindstrom S, Vachon CM, Li J, Varghese J, Thompson D, Warren R, et al. Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. Nat Genet 2011;43:185-7.

19. Lange K, Westlake J, Spence MA. Extensions to pedigree analysis. III. Variance components by the scoring method. Ann Hum Genet 1976;39:485-91.

20. Hopper JL, Mathews JD. Extensions to multivariate normal models for pedigree analysis. Ann Hum Genet 1982;46:373-83.

21. Hopper JL. On analysis of path models by the multivariate normal model for pedigree analysis. Genet Epidemiol 1986;3:279-81.

22. Stone J, Dite GS, Giles GG, Cawson J, English DR, Hopper JL. Inference about causation from examination of familial confounding: application to longitudinal twin data on mammographic density measures that predict breast cancer risk. Cancer Epidemiol Biomark Prev 2012;21:1149-55.

Table 1. Characteristics of the 3,324 participants; mean and standard deviation (SD) for continuous variables, and number with the characteristic (N) and percentage (%) for categorical variables

Characteristic	**Mean**	**SD**
Mammographic density measures		
Dense area (cm^2)	29.67	23.06
Non-dense area (cm^2)	99.35	57.49
Percent dense area	26.13	18.21
Age at interview *(years)*	54.63	8.24
Mammogram age *(years)*	54.54	8.44
Weight *(kg)*	69.60	14.22
Height *(cm)*	162.88	6.87
BMI *(kg/m^2)*	26.25	5.24
Number of live births *(n=2,880)*	2.33	1.41
Age at menarche *(years)*	13.05	1.58
Years of hormone replacement therapy use *(n=1,333)*	2.51	4.96
Years of oral contraceptive use *(n=2,915)*	6.87	7.38
Years of smoking *(n=1,271)*	4.31	8.19
Years of alcohol consumption *(n=1,913)*	12.38	13.90
	N	**%**
Live birth ever *(yes)*	2,880	87
Breast feeding *(yes)*	2,557	77
Menopausal status *(post-menopausal)*	2,299	70
Hormone replacement therapy use		
Ever (yes)	1,333	40
Current (yes)	527	16
Oral contraceptive use		
Ever (yes)	2,915	88
Current (yes)	212	7
Smoking status		
Ever (yes)	1,271	38
Current (yes)	366	11

Alcohol consumption		
Previous (yes)	1,913	58
Current (yes)	1,585	48
Number of ovaries removed		
None	2,671	80
1	267	8
2	386	12
Benign breast lump removed *(yes)*	469	14
Ductal carcinoma *in situ (yes)*	26	1
Other cancer *(yes)*	244	7
Country of birth		
Australia	3,016	91
Other	308	9
Ethic background		
Caucasian	3,286	99
Other	38	1

Table 2. Breakdown of families by number and type of participants within the family

Number of participants	Type of participants	Number of families
1	Non-twin	337
2	MZ twin pair	379
2	DZ twin pair	242
2	Non-twin sisters	219
3	MZ twin pair and non-twin sister	123
3	DZ twin pair and non-twin sister	75
3	3 non-twin sisters	83
4	MZ twin pair and 2 non-twin sisters	36
4	DZ twin pair and 2 non-twin sisters	18
4	4 non-twin sisters	24
5	MZ twin pair and 3 non-twin sisters	3
5	DZ twin pair and 3 non-twin sisters	5
5	5 non-twin sisters	10
6	MZ twin pair and 4 non-twin sisters	3
6	6 non-twin sisters	5
7	7 non-twin sisters	2

Table 3. Regression coefficients (SE), standardized regression coefficients, and p-values for cube root of dense area

	Univariate				Multivariate[b]			
Variables	β	(SE)	β* [a]	p-value	β	(SE)	β* [a]	p-value
Mammogram age (years)	−0.026	(0.002)	−0.26	<0.001	**−0.020**	**(0.002)**	**−0.20**	**<0.001**
Log of BMI (kg/m^2)	−0.568	(0.077)	−0.13	<0.001	**−0.452**	**(0.074)**	**−0.10**	**<0.001**
Menopausal status	−0.348	(0.031)	−0.19	<0.001	**−0.162**	**(0.037)**	**−0.09**	**<0.001**
Number of live births	−0.098	(0.010)	−0.17	<0.001	**−0.059**	**(0.010)**	**−0.10**	**<0.001**
Breast feeding	−0.067	(0.033)	−0.03	0.04	*0.075*	*(0.037)*	*0.04*	*0.04*
Age at menarche (years)	−0.002	(0.009)	−0.004	0.8	*−0.010*	*(0.009)*	*−0.02*	*0.3*
Hormone replacement therapy use								
Never	Ref		ref		ref		ref	
Ever	−0.098	(0.029)	−0.06	0.001	*0.012*	*(0.034)*	*0.01*	*0.7*
Current	0.060	(0.038)	0.03	0.1	**0.104**	**(0.036)**	**0.05**	**0.004**
Years of hormone replacement therapy use	−0.006	(0.003)	−0.03	0.04	*0.003*	*(0.003)*	*0.02*	*0.3*
Oral contraceptive use								
Never	Ref		ref		ref		ref	
Ever	0.116	(0.043)	0.05	0.007	*0.022*	*(0.044)*	*0.01*	*0.6*
Current	0.088	(0.056)	0.03	0.1	*−0.093*	*(0.059)*	*−0.03*	*0.1*
Years of oral contraceptive use	0.002	(0.002)	0.02	0.4	**−0.007**	**(0.002)**	**−0.06**	**<0.001**
Smoking status								
Never	Ref		ref		ref		ref	
Ever	0.069	(0.029)	0.04	0.02	*−0.004*	*(0.029)*	*−0.002*	*0.9*
Current	−0.001	(0.044)	−0.0002	0.99	*−0.091*	*(0.042)*	*−0.03*	*0.03*
Years of smoking	0.003	(0.002)	0.03	0.1	*0.002*	*(0.002)*	*0.02*	*0.2*
Alcohol consumption								
Never	Ref		ref		ref		ref	

Previous	0.118	(0.029)	0.07	<0.001	*−0.048*	*(0.042)*	*−0.03*	*0.3*
Current	0.134	(0.028)	0.08	<0.001	*0.023*	*(0.041)*	*0.01*	*0.6*
Years of alcohol consumption	0.004	(0.001)	0.07	<0.001	**0.004**	**(0.001)**	**0.06**	**<0.001**
Ovary removed	−0.098	(0.035)	−0.05	0.005	*−0.050*	*(0.035)*	*−0.02*	*0.2*
Number of ovaries removed	−0.057	(0.021)	−0.05	0.006	*−0.025*	*(0.021)*	*−0.02*	*0.2*
None	*Ref*		*ref*		*ref*		*ref*	
1	−0.087	(0.050)	−0.1	0.08	*−0.062*	*(0.048)*	*−0.07*	*0.2*
2	−0.107	(0.043)	−0.13	0.01	*−0.040*	*(0.043)*	*−0.05*	*0.4*
Benign breast lump removed	0.133	(0.039)	0.06	0.001	**0.161**	**(0.037)**	**0.07**	**<0.001**
Ductal carcinoma *in situ*	0.539	(0.148)	0.06	<0.001	**0.461**	**(0.141)**	**0.05**	**0.001**

[a]standardized standard error = 0.02

[b]Adjusted for: age at mammogram, menopausal status, number of live births, breastfeeding, log-transformed current BMI, current hormone replacement therapy, number of years of oral contraceptive use, current smoking, number years of smoking, number years of alcohol use, benign breast lump removed, and ductal carcinoma *in situ*.

Table 4. Regression coefficients (SE), standardized regression coefficients, and p-values for square root of percent dense area

	Univariate				Multivariate[b]			
Variables	β	(SE)	β* [a]	p-value	β	(SE)	β* [a]	p-value
Mammogram age (years)	−0.082	(0.004)	−0.35	<0.001	**−0.064**	**(0.005)**	**−0.27**	**<0.001**
Log of BMI (kg/m^2)	−4.316	(0.165)	−0.41	<0.001	**−4.037**	**(0.154)**	**−0.38**	**<0.001**
Menopausal status	−1.106	(0.072)	−0.26	<0.001	**−0.374**	**(0.077)**	**−0.09**	**<0.001**
Number of live births	−0.284	(0.022)	−0.20	<0.001	**−0.142**	**(0.020)**	**−0.10**	**<0.001**
Breast feeding	−0.231	(0.076)	−0.05	0.002	*0.083*	*(0.076)*	*0.02*	*0.3*
Age at menarche (years)	0.058	(0.021)	0.05	0.006	*−0.006*	*(0.018)*	*−0.005*	*0.7*
Hormone replacement therapy use								
Never	Ref		ref		ref		ref	
Ever	−0.360	(0.067)	−0.09	<0.001	*0.020*	*(0.071)*	*0.005*	*0.8*
Current	0.151	(0.088)	0.03	0.09	**0.266**	**(0.075)**	**0.05**	**<0.001**
Years of hormone replacement therapy use	−0.019	(0.007)	−0.05	0.003	*0.011*	*(0.007)*	*0.03*	*0.1*
Oral contraceptive use								
Never	Ref		ref		ref		ref	
Ever	0.322	(0.100)	0.05	0.001	*0.006*	*(0.091)*	*0.001*	*0.95*
Current	0.521	(0.131)	0.06	<0.001	*−0.139*	*(0.122)*	*−0.02*	*0.3*
Years of oral contraceptive use	0.015	(0.005)	0.06	0.001	**−0.013**	**(0.004)**	**−0.05**	**0.001**
Smoking status								
Never	Ref		ref		ref		ref	
Ever	0.102	(0.069)	0.03	0.1	*−0.028*	*(0.066)*	*−0.01*	*0.7*
Current	0.110	(0.103)	0.02	0.3	**−0.246**	**(0.088)**	**−0.04**	**0.005**
Years of smoking	0.001	(0.004)	0.01	0.7	*0.005*	*(0.004)*	*0.02*	*0.2*
Alcohol consumption								
Never	Ref		ref		ref		ref	
Previous	0.372	(0.067)	0.09	<0.001	*−0.143*	*(0.086)*	*−0.04*	*0.1*

Current	0.460	(0.066)	0.12	<0.001	*0.089*	*(0.086)*	*0.02*	*0.3*
Years of alcohol consumption	0.011	(0.002)	0.08	<0.001	**0.008**	**(0.002)**	**0.06**	**<0.001**
Ovary removed	−0.409	(0.081)	−0.08	<0.001	*−0.137*	*(0.073)*	*−0.03*	*0.06*
Number of ovaries removed	−0.260	(0.048)	−0.09	<0.001	*−0.084*	*(0.044)*	*−0.03*	*0.05*
None	*Ref*		*ref*		*ref*		*ref*	
1	−0.251	(0.116)	−0.13	0.03	*−0.105*	*(0.100)*	*−0.05*	*0.3*
2	−0.522	(0.100)	−0.26	<0.001	*−0.163*	*(0.090)*	*−0.08*	*0.07*
Benign breast lump removed	0.279	(0.091)	0.05	0.2	**0.334**	**(0.077)**	**0.06**	**<0.001**
Ductal carcinoma *in situ*	1.189	(0.347)	0.05	0.001	**0.965**	**(0.294)**	**0.04**	**0.001**

[a]standardized standard error = 0.02

[b]Adjusted for: age at mammogram, menopausal status, number of live births, log-transformed current BMI, current hormone replacement therapy, number of years of hormone replacement therapy use, number of years of oral contraceptive use, current smoking, current alcohol consumption, previous alcohol consumption, number of years of alcohol use, number of ovaries removed, benign breast lump removed, and ductal carcinoma *in situ*.

Table 5. Regression coefficients (SE), standardized regression coefficients, and p-values for square root of non-dense area

	Univariate				Multivariate[b]			
Variables	β	(SE)	β* [a]	p-value	β	(SE)	β* [a]	p-value
Mammogram age (years)	0.095	(0.006)	0.29	<0.001	**0.089**	**(0.005)**	**0.27**	**<0.001**
Log of BMI (kg/m^2)	9.789	(0.195)	0.65	<0.001	**9.462**	**(0.185)**	**0.63**	**<0.001**
Menopausal status	1.335	(0.103)	0.22	<0.001	*0.204*	*(0.094)*	*0.03*	*0.03*
Number of live births	0.296	(0.032)	0.15	<0.001	**0.089**	**(0.024)**	**0.05**	**<0.001**
Breast feeding	0.239	(0.109)	0.04	0.03	*0.038*	*(0.091)*	*0.01*	*0.7*
Age at menarche (years)	−0.173	(0.030)	−0.10	<0.001	*−0.022*	*(0.022)*	*−0.01*	*0.3*
Hormone replacement therapy use								
Never	Ref		ref		ref		ref	
Ever	0.415	(0.095)	0.07	0.001	*−0.052*	*(0.087)*	*−0.01*	*0.5*
Current	−0.178	(0.126)	−0.02	0.1	*−0.207*	*(0.101)*	*−0.03*	*0.04*
Years of hormone replacement therapy use	0.018	(0.009)	0.03	0.06	**−0.034**	**(0.007)**	**−0.06**	**<0.001**
Oral contraceptive use								
Never	Ref		ref		ref		ref	
Ever	−0.360	(0.142)	−0.04	0.01	*0.089*	*(0.104)*	*0.01*	*0.4*
Current	−0.981	(0.187)	−0.09	<0.001	*−0.036*	*(0.137)*	*−0.003*	*0.8*
Years of oral contraceptive use	−0.038	(0.006)	−0.10	<0.001	*−0.003*	*(0.005)*	*−0.01*	*0.6*
Smoking status								
Never	Ref		ref		ref		ref	
Ever	0.026	(0.098)	0.005	0.8	**0.238**	**(0.072)**	**0.04**	**0.001**
Current	−0.262	(0.147)	−0.03	0.08	*0.214*	*(0.116)*	*0.02*	*0.07*
Years of smoking	0.006	(0.006)	0.02	0.3	*−0.003*	*(0.005)*	*−0.01*	*0.5*
Alcohol consumption								
Never	Ref		ref		ref		ref	
Previous	−0.530	(0.095)	−0.09	<0.001	*0.219*	*(0.104)*	*0.04*	*0.04*

Current	−0.724	(0.094)	−0.13	<0.001	*−0.108*	*(0.104)*	*−0.02*	*0.3*
Years of alcohol consumption	−0.015	(0.003)	−0.08	<0.001	**−0.009**	**(0.003)**	**−0.04**	**0.001**
Ovary removed	0.729	(0.115)	−0.10	<0.001	*-0.078*	*(0.242)*	*−0.01*	*0.7*
Number of ovaries removed	0.474	(0.068)	0.11	<0.001	**0.189**	**(0.052)**	**0.05**	**0.001**
None	*Ref*		*ref*		*ref*		*ref*	
1	0.391	(0.164)	0.14	0.02	*0.155*	*(0.118)*	*0.05*	*0.2*
2	0.971	(0.143)	0.35	<0.001	**0.388**	**(0.109)**	**0.14**	**<0.001**
Benign breast lump removed	−0.156	(0.130)	−0.02	0.2	*−0.155*	*(0.093)*	*−0.02*	*0.1*
Ductal carcinoma *in situ*	−0.941	(0.496)	−0.03	0.06	*−0.709*	*(0.352)*	*−0.02*	*0.05*

[a]standardized standard error = 0.02

[b]Adjusted for: age at mammogram, menopausal status, number of live births, log-transformed current BMI, current hormone replacement therapy, number years of hormone replacement therapy use, previous smoking, current alcohol consumption, previous alcohol consumption, number years of alcohol use, number of ovaries removed, and ductal carcinoma *in situ*.

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

Explaining variance in the Cumulus mammographic measures that predict breast cancer risk: a twins and sisters study.

Tuong L. Nguyen, Daniel F. Schmidt, Enes Makalic, et al.

Cancer Epidemiol Biomarkers Prev Published OnlineFirst October 15, 2013.