Novel mammogram-based measures improve breast cancer risk prediction beyond an established measure of mammographic density

Tuong L. Nguyen,[1#] Daniel F. Schmidt,[2#] Enes Makalic,[1#] Gertraud Maskarinec,[3] Shuai Li,[1] Gillian Dite,[1] Ye K. Aung,[1] Christopher F. Evans,[1] Ho N. Trinh,[1] Laura Baglietto,[4] Jennifer Stone,[5] Yun-Mi Song,[6] Joohon Sung,[7,8] Robert J. MacInnis,[1,9] Pierre-Antoine Dugué,[1,9,10] James G. Dowty,[1] Mark A. Jenkins,[1] Roger L. Milne,[1,9,10] Melissa C. Southey,[9,10] Graham G. Giles,[1,9,10] John L. Hopper[1*]

1. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, 207 Bouverie Street, Parkville, Victoria 3010, Australia.
2. Faculty of Information Technology, Monash University, Clayton, Victoria, Australia.
3. University of Hawaii Cancer Center, Honolulu, Hawaii, USA.
4. Department of Clinical and Experimental Medicine, University of Pisa, Italy.
5. Curtin UWA Centre for Genetic Origins of Health and Disease, Curtin University and the University of Western Australia, Perth, Western WA, 6009, Australia.
6. Department of Family Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Irwon-ro 81, Gangnamgu, Seoul, South Korea.
7. Department of Epidemiology School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, 151-742 Seoul, South Korea.
8. Institute of Health and Environment, Seoul National University, 1 Gwanak-ro, Gwanak-gu, 151-742 Seoul, South Korea.
9. Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia
10. Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia

# These authors contributed equally to this work.

* **Corresponding author**: Professor John Hopper, Centre for Epidemiology and Biostatistics, The University of Melbourne, 207 Bouverie Street, Carlton, Victoria, 3053, Australia (email: j.hopper@unimelb.edu.au)

Key Words: breast cancer, Cirrus, Cirrocumulus, interval breast cancer, mammographic density, screen-detected breast cancer

Grant support

This research was supported by the National Health and Medical Research Council (251533, 209057, and 504711), the Victorian Health Promotion Foundation, Cancer Council Victoria, Cancer Council NSW, Cancer Australia, and the National Breast Cancer Foundation. It has also been supported by the Breast Cancer Network Australia, the National Breast Cancer Foundation, Victoria Breast Cancer Research Consortium and was further supported by infrastructure provided by the Cancer Council Victoria and the University of Melbourne. We thank the Victorian Cancer Registry, BreastScreen Victoria, the Australian Mammographic Density Research Facility. TLN has been supported by Cure Cancer Australia Foundation through Cancer Australia Priority-Driven Collaborative Cancer Research Scheme (1159399). TLN and SL have been supported by Victorian Cancer Council Post-Doctoral Fellowships and grants from the Picchi Foundation, Victorian Comprehensive Cancer Centre. JLH is a NHMRC Senior Principal Research Fellow. MAJ and MCS are NHMRC Senior Research Fellows.

Disclosure of Potential Conflicts of Interest

GSD receives funding from Genetic Technologies Ltd for work unrelated to this study.

Number of Tables: 4

Number of Figures: 1

Word count: 2,790

ABSTRACT

Background: Mammograms contain information that predicts breast cancer risk. We recently discovered two novel mammogram-based breast cancer risk measures based on image brightness (*Cirrocumulus*) and texture (*Cirrus*). It is not known whether these measures improve risk prediction when fitted together, and with an established measure of mammographic density (*Cumulus*).

Methods: We used three studies consisting of: 168 interval cases and 498 matched controls; 422 screen-detected cases and 1,197 matched controls; and 354 younger-diagnosis cases and 944 frequency-matched controls. We conducted conditional and unconditional logistic regression analyses of individually- and frequency-matched studies, respectively. We reported risk gradients as change in odds ratio per standard deviation of controls after adjusting for age and body mass index (OPERA). For models involving multiple measures, we calculated the OPERA equivalent to the area under the receiver operating characteristic curve.

Results: For interval, screen-detected and younger-diagnosis cancer, the best fitting models (OPERAs [95% confidence intervals]) were: *Cumulus* (1.81 [1.41 to 2.31]) and *Cirrus* (1.7 [1.38 to 2.14]); *Cirrus* (1.49 [1.32 to 1.67]) and *Cirrocumulus* (1.16 [1.03 to 1.31]); and *Cirrus* (1.70 [1.48 to 1.94]) and *Cirrocumulus* (1.46 [1.27 to 1.68]), respectively. Their OPERA equivalents were: 2.35, 1.58, and 2.28, respectively.

Conclusions: Our mammogram-based measures improved risk prediction beyond and, except for interval cancers, negated the influence of conventional mammographic density. Combined, these new mammogram-based risk measures are at least as accurate as the current polygenetic risk scores (OPERA ~ 1.6) in predicting, on a population basis, women who will be diagnosed with breast cancer.

248 words

Historically, mammographic density has been defined as the light or bright areas on a mammogram (we call this *Cumulus*) and has well-established risk associations with breast cancer overall and with both interval and screen-detected cancers [1]. But there has been debate about the extent to which these associations are due to existing tumours being missed at mammographic screening, especially for interval cancers. It is also not clear if this measure is the only, let alone the best, mammogram-based predictor of breast cancer.

We addressed these issues by trying to discover aspects of a mammographic image that differ between women with and without breast cancer. First, we redefined mammographic density at, in effect, a higher pixel brightness threshold to encompass just the brightest regions to create *Cirrocumulus* [2-5]. Second, we applied machine learning to textural patterns to create *Cirrus* [6].

We previously considered each new measure separately with the established measure. Both *Cirrus* and *Cirrocumulus* were correlated with *Cumulus*; r~0.4 and 0.6, respectively. Except for interval cancer, we found that, when fitted together, the *Cumulus* risk gradient substantially decreased compared with being fitted alone. On the other hand, the *Cirrocumulus* and *Cirrus* risk gradients both remained similar to what they were when fitted alone [2-6]. We concluded that "conventional mammographic density predicts interval cancer due to its role in masking, while the new mammogram-based risk measures could have a causal effect on both interval and screen-detected breast cancer" [7].

We assessed the strength of risk prediction, in terms of the ability to differentiate cases from controls on a population basis, using the odds per adjusted standard deviation (OPERA) [8]. Here we used the standard deviation of the residuals for controls after adjusting for age and body mass index, not the standard deviation of the cross-sectional distribution of the measure itself. This is because age and body mass index confound the associations of mammogram-based risk measures with breast cancer risk and need to be adjusted for. After doing this, the resulting estimated risk gradient estimate relates to the adjusted measure, not to the raw measure.

OPERA also allows risk factors to be compared and put into perspective in terms of risk discrimination in ways that are not possible using, for example, *change* in the area under the receiver operating characteristic curve (AUC), which depends on the order in which the factors are included.

The OPERA approach showed that, individually, our new mammogram-based risk measures are among the strongest of all currently known breast cancer risk factors [6]. But it is not known what risk prediction is obtained when these are fitted together with the conventional mammographic density measure.

In this paper we aimed to determine the extent to which our new measures are correlated with one another and the extent to which risk prediction is improved when our new measures are fitted together and with the established mammographic density measure. In doing so, we sought to find out how risk prediction obtained from combining our new measures compares with that from other breast cancer risk factors.

Methods

We used data from three independent studies: (i) a nested case-control study of 168 cases with interval breast cancer (those diagnosed within two years of a negative screen) and 498 matched controls within the Melbourne Collaborative Cohort Study (MCCS) [5, 9-11]; (ii) a nested case-control study of 422 cases with screen-detected breast cancer and 1,197 matched controls within the MCCS [5, 9-10]; and (iii) a case-control study of 354 cases with on average younger-diagnosis breast cancers and 944 controls from the Australian Breast Cancer Family Study and the Australian Mammographic Density Twins and Sisters Study, both over-sampled and frequency matched for family history [3].

For all studies, the average ages at diagnosis of the cases and the average ages at interview of the controls were similar. For studies (i) and (ii), the average times between mammogram and diagnosis were 6 and 5 years (standard deviation 3) for screen-detected and interval cancers, respectively. Mean (standard deviation) age at diagnosis was 62.3 (7.3) years for the interval cancers, 64.3 (8.2) years for the screen-detected cancers, and 48.5 (10.7) years for the younger-diagnosis cases. For study (iii), by design 30% of cases and 29% of controls had a family history of breast cancer compared with 10% of controls in studies (i) and (ii).

Mammogram-based measures

We used digitised film mammograms. The *Cumulus* and *Cirrocumulus* measures had been created using the computer-assisted threshold software CUMULUS [12]. The *Cirrus* measures were those created previously [6]. All measures were transformed to approximate normality, adjusted for age and body mass index, and scaled by the standard deviation of the residuals for controls. For *Cirrocumulus* we used the absolute measure because it has less measurement error. For *Cumulus* we used the percentage measure because it was the better predictor of interval cancer [10].

Statistical Methods

All analyses were conducted using the Stata software [13]. For descriptive purposes, we presented the numbers of cases and controls for each pair of measures, and the estimated risks relative to the population average for different tertile-by-tertile categories (based on controls) using the cci option. The control data was used to investigate the joint distributions of the pairs of measures and to determine the proportion of the population in different categories.

Risk gradients were estimated using conditional logistic regression for the two nested matched case-control studies and using unconditional logistic regression for the frequency-matched case-control study. To compare fits, we used the likelihood ratio criterion [14] with $P<0.05$ considered to be the threshold for nominal statistical significance.

The risk gradient, and hence the ability to differentiate cases from controls on a population basis, was reported as the change in OPERA for which we used adjustment for age at mammogram and body mass index [8]. We present OPERA estimates in the tables for ease of interpretation and log(OPERA) in the text because it is the natural scale on which to assess risk gradients.

Under the assumptions of a multiplicative risk model for a normally distributed risk factor, $\log(\text{OPERA}) = \sqrt{2}\Phi^{-1}(\text{AUC})$, where Φ is the normal (0,1) distribution function (see Supplementary Material in [6]), so that the AUC is approximately linearly related to log(OPERA), at least in the range of AUC from 0.5 to 0.7 (OPERA from 1 to 2). Under the multiplicative risk model, log(OPERA) is equal to the difference in means between cases and controls divided by the standard deviation of the adjusted risk factor; the inter-quartile risk ratio is approximately $\text{OPERA}^{2.5}$. When we fitted multiple risk measures together, we used the AUC and the formula above to calculate the corresponding (equivalent) log(OPERA) as if we had fitted one combined measure.

Results

Combining data from the three control samples, the correlation between *Cirrocumulus* and *Cirrus* was 0.3, while the correlations of percentage *Cumulus* with *Cirrocumulus* and *Cirrus* were both 0.5; see Supplementary Figures. The standard errors of these correlations were ~0.02.

Interval breast cancer

Table 1 shows that when *Cumulus* and *Cirrus*, or *Cirrocumulus* and *Cirrus*, were fitted together, their risk associations both attenuated but remained significant. When all three were fitted together, *Cirrocumulus* was not significant (P=0.6). The best-fitting model involved *Cumulus* and *Cirrus* with log OPERAs of 0.59 (95% confidence interval [CI] 0.34 to 0.84) and 0.54 (95% CI, 0.32 to 0.76), respectively. The AUC was equivalent to log(OPERA) = 0.85 (95% CI, 0.68 to 1.04).

Women in the highest tertiles of both *Cumulus* and *Cirrus* are at 2.51 (95% CI, 1.72 to 3.64) times population risk (P<0.001), and this group comprised 17% of controls. At the other extreme, women in the lowest tertiles of both *Cumulus* and *Cirrus* are at 0.22 (95% CI, 0.08 to 0.49) times population risk (P<0.001), and this group comprised 19% of controls; see Supplementary Table 1. Similar findings were obtained when stratifying women by tertiles of both *Cirrocumulus* and *Cirrus*; see Table 2.

Screen-detected breast cancer

Table 3 shows that, when *Cumulus* was included with *Cirrus* or *Cirrocumulus*, there was no improvement in fit and the *Cumulus* estimate was no longer significant. When *Cirrocumulus* and *Cirrus* were fitted together, their risk associations both attenuated but remained significant. The best fitting model involved *Cirrus* and *Cirrocumulus* with log OPERAs of 0.40 (95% CI, 0.28 to 0.51)

and 0.15 (95% CI, 0.03 to 0.27), respectively. For the combined measures, the AUC was equivalent to log(OPERA) = 0.46 (95% CI, 0.34 to 0.58).

Women in the highest tertiles of both *Cirrocumulus* and *Cirrus* are at 2.01 (95% CI, 1.54 to 2.61) times population risk ($P<0.001$), and this group comprised 15% of controls. At the other extreme, women in the lowest tertiles of both *Cirrocumulus* and *Cirrus* are at 0.52 (95% CI, 0.33 to 0.78) times population risk ($P<0.001$), and this group comprised 14% of controls; see Table 2.

Younger-diagnosed cancer

Table 4 shows that, when fitted alone, there was very strong evidence that the model including *Cirrus* had the best fit, and the fit was improved when further including *Cirrocumulus* ($P<0.001$). The addition of *Cumulus* did not improve the fit ($P=0.8$). The best-fitting model involved *Cirrus* and *Cirrocumulus* with log OPERAs of 0.53 (95% CI, 0.39 to 0.66) and 0.38 (95% CI, 0.24 to 0.52), respectively. The AUC was equivalent to log(OPERA) = 0.82 (95% CI, 0.70 to 0.96).

Women in the highest tertiles of both *Cirrocumulus* and *Cirrus* are at 2.54 (95% CI, 1.93 to 3.33) times population risk ($P<0.001$), and this group comprised 16% of controls. At the other extreme, women in the lowest tertiles of both *Cirrocumulus* and *Cirrus* are at 0.39 (95% CI, 0.24 to 0.62) times population risk ($P<0.001$), and this group comprised 17% of controls; see Table 2.

Figure 1 shows that the receiver operating curves for *Cirrus* and *Cirrocumulus* have different shapes and crossed over. For *Cirrus*, the sensitivity increased rapidly from zero as the specificity decreased from 1, while for *Cirrocumulus*, the specificity increased rapidly from zero as the sensitivity decreased from 1.

Discussion

Our new mammogram-based risk measures based on brightness (*Cirrocumulus*) and texture (*Cirrus*) improved risk prediction beyond the established measure of mammographic density (*Cumulus*). For all three studies, the best fitting model included several measures and performed substantially better than using the established measure alone (all $P<0.001$). Except for interval cancers, the new measures also negated the importance of the established measure on risk prediction.

We also found that, when combined, the new mammogram-based risk measures are at least as accurate in identifying women who will be diagnosed with breast cancer as the recently published polygenic risk score, which has an OPERA of ~1.6 [15] or log(OPERA) = 0.48. For younger-diagnosis breast cancer, the AUC for the combination of our measures was 0.72, equivalent to log(OPERA) = 0.82. Therefore, in terms of differentiating women with or without breast cancer at a young age, our measures were $([0.82-0.48]/0.48)\times100 = 70\%$ more accurate than the polygenic risk score. It is plausible that inclusion of a polygenic risk score with the mammogram-based risk measures will further improve risk prediction [16].

On a population basis, therefore, the combination of these new measures appears to be the strongest of all known breast cancer risk factors. For example, when *Cirrocumulus* and *Cirrus* were combined to predict breast cancer at on average a young age (see Table 1), the OPERA equivalent was 2.28, so the interquartile risk ratio is ~7-fold. In comparison, the interquartile risk ratio is ~4-fold for a multigenerational family history risk score in predicting breast cancer before age 50 years, ~3-fold for the latest polygenic risk score, ~2-fold for conventional mammographic density, ~1.5-fold for *BRCA1* and *BRCA2* mutations, and ~1.2-fold or less for lifestyle-related risk factors [7,8].

From the contrasting shapes of their receiver operating characteristic curves, it can be seen that *Cirrus* has greater sensitivity at high specificity (correctly identifying true negatives), while *Cirrocumulus* has greater specificity at high sensitivity (correctly identifying true positives).

Therefore, *Cirrocumulus* does better at identifying women at higher than average risk, while *Cirrus* does better at identifying women at lower than average risk.

Cirrus gave similar risk gradients in all three settings, suggesting it is tapping into new and fundamental risk-predicting aspects of a mammogram. This was also evident in the original work developing *Cirrus*, which found that a similar risk prediction was achieved for women of Japanese ancestry living in Hawaii and for Australian women [6]. Note that *Cirrus* was designed not to depend on brightness and has only a modest correlation with the brightness measures.

Our new mammogram-based measures are potentially of substantial clinical and population health significance. They not only identify groups of women at substantially increased risk, but they also identify larger groups of women at decreased risk. When categorised by tertiles, *Cirrus* and *Cirrocumulus* divide the population into two extreme groups of approximately the same size (each about 15–20%) containing women who are on average either at twice or more population risk, or at half or less population risk; see Table 2. For interval cancer, about 60% of controls were in the six categories with below population average risk, while for both screen-detected and interval cancer, about 75% of controls were in the six categories with below population average risk.

These observations are highly relevant to considerations of tailored, or personalised, screening based on risk, for which there are now several trials being conducted across the world. These include the Wisdom Study in the United States [17,18], My personalised breast screening (MyPeBS) in France (https://clinicaltrials.gov/ct2/show/NCT03672331), and PROCAS2 in the United Kingdom (https://preventbreastcancer.org.uk/breast-cancer-research/research-projects/early-detection-screening/procas/).

These risk categorisations are in stark contrast to those using BI-RADS alone. Currently, about 40% of screening women in the United States are classified as having dense breasts defined by BI-RADS categories c or d. As a result of a community-led initiative [19], in 35 states it is mandated by law that these women are informed. Research studies in which one or a few radiologists measure

BI-RADS in a controlled manner suggest the increased risk associated with having dense breasts is about 1.6 to 2.2-fold (see IBIS [20] and BOADICEA [21]).

In practice, BI-RADS is measured by multiple radiologists at a given screening service, especially over time, opening the potential for substantial measurement error. For example, from the Supplemental data on 60,000 women screened at a large United States medical center [22], the odds ratio for being classified as having dense breast is only about 1.1, which is far less than found by the research studies ($P<0.001$). This was despite the measurements being recorded by "radiologists who specialized in breast imaging and who had 5–33 years of experience following the American College of Radiology BI-RADS lexicon" [23]. It would appear, therefore, that in practice there could be so much variation across measurers, even experienced specialists in a large city-based service, that clinical BI-RADS measurements might be providing very little information on risk stratification across the population.

There is substantial scope for better addressing the issue of dense breasts by going beyond BI-RADS. A major consequence of having dense breasts is an increased risk of interval cancer. We and others have found that, as well as conventional mammographic density (*Cumulus*), having a family history and other risk factors, such as cumulative exposure to ovarian hormones based on the Pike model [24], combine to predict interval cancer [11]. In our study we have found that *Cirrus* also brings almost as much information as *Cumulus*, and when combined they have an inter-quartile risk ratio for interval cancer of almost 9-fold. Future work will consider how risk of interval cancer, and even of missed cancers, can be further optimised by combining mammogram-based measures with family history, genetic risk scores and other risk factors. This could have a profound impact on the way the issue of dense breasts is addressed in the future.

For our findings to be translated into wider clinical practice, automated use of the mammogram-based and other risk measures needs to be implemented. We are developing a program to measure *Cirrus* automatically from batch files of mammograms and are using deep learning to

develop similar automated measures of *Cumulus* and *Cirrocumulus*. We are developing the empirical evidence, such as in this and other papers [11, 16] to find out how mammogram-based risk measures combine with each other and with other important risk factors to predict risk.

In conclusion, while the established mammographic density measure improved prediction of interval cancer, most likely due to its role in masking tumours, it provided no substantive additional information on risks of screen-detected or younger-diagnosis cancer in addition to our new mammogram-based risk measures. Our findings demonstrate the potential for much improved and more aetiologically relevant breast cancer risk prediction by discovering new ways of extracting information on breast cancer risk from a mammogram. Risk-based personalised breast screening could become part of the precision medicine era [17, 18].

REFERENCES

1. Boyd NF, Guo H, Martin LJ, *et al.* Mammographic density and the risk and detection of breast cancer. N Engl J Med 2007;356:227-36.

2. Nguyen TL, Aung YK, Evans CF, *et al.* Mammographic density defined by higher than conventional brightness threshold better predicts breast cancer risk for full-field digital mammograms. Breast Cancer Res 2015;17:142.

3. Nguyen TL, Aung YK, Evans CF, *et al.* Mammographic density defined by higher than conventional brightness thresholds better predicts breast cancer risk. Int J Epidemiol 2017;46:652-661.

4. Nguyen TL, Choi YH, Aung YK, *et al.* Breast cancer risk associations with digital mammographic density by pixel brightness threshold and mammographic system. Radiology 2018;286:433-442.

5. Nguyen TL, Aung YK, Li S, *et al.* Predicting interval and screen-detected breast cancers from mammographic density defined by different brightness thresholds. Breast Cancer Res 2018;20:152.

6. Schmidt DF, Makalic E, Goudey B, *et al.* Cirrus: an automated mammography-based measure of breast cancer risk based on textural features. JNCI Cancer Spectrum 2018;2(4).pky057.

7. Hopper JL, Nguyen TL, Schmidt DF, et al. Going beyond conventional mammographic density to discover novel mammogram-based predictors of breast cancer risk. J Clin Med 2020; 9:(3).pii: E627.

8. Hopper JL. Odds per adjusted standard deviation: comparing strengths of associations for risk factors measured on different scales and across diseases and populations. Am J Epidemiol 2015;182:863-867.

9. Baglietto L, Krishnan K, Stone J, *et al.* Associations of mammographic dense and nondense areas and body mass index with risk of breast cancer. Am J Epidemiol 2014;179:475-83.

10. Krishnan K, Baglietto L, Apicella C, *et al.* Mammographic density and risk of breast cancer by mode of detection and tumor size: a case-control study. Breast Cancer Res 2016;18:63.

11. Nguyen TL, Li S, Dite GS, *et al.* Interval breast cancer risk associations with breast density, family history, and breast tissue ageing. Int J Cancer 2019;145:1768-1773.

12. Byng JW, Yaffe MJ, Jong RA, *et al.* Analysis of mammographic density and breast cancer risk from digitized mammograms. Radiographics 1998;18:1587-98.

13. StataCorp. Stata statistical software: Release 15 College Station, TX: StataCorp LLC, 2017.

14. Fisher RA. Statistical methods and scientific induction. J Roy Statist Soc B 1995;17:69-78.

15. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet 2019;104:21-34.

16. Vachon CM, Scott CG, Tamimi RM, *et al.* Joint association of mammographic density adjusted for age and body mass index and polygenic risk score with breast cancer risk. Breast Cancer Res 2019;21:68.

17. Esserman LJ, and the WISDOM Study and Athena Investigators. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. NPJ Breast Cancer 2017;3;34.

18. Shieh Y, Eklund M, Madlensky L, et al. Breast cancer screening in the precision medicine era: risk-based screening in a population-based trial. J Natl Cancer Inst. 2017;109(5): doi:10.1093/jnci/djw290.

19. Cappello NM, Richetelli D, Lee CI. The impact of breast density reporting laws on women's awareness of density-associated risks and conversations regarding supplemental screening with providers. J Am Coll Radiol 2019;**16**:139-46.

20. Warwick J, Birke H, Stone J, et al. Mammographic breast density refines Tyrer-Cuzick estimates of breast cancer risk in high-risk women: findings from the placebo arm of the International Breast Cancer Intervention Study I. Breast Cancer Research 2014;16:451-6.

21. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genetics in Medicine 2019;21:1708-1718.

22. Yala A, Lehman C, Schuster T, et al. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology 2019;291:60-66.

23. Lehman CD, Yala A, Schuster T, et al. Mammographic breast density assessment using deep learning: clincal implementation. Radiology 2019;290:52-58.

24. Pike MC, Krailo MD, Henderson BE, Casagrande JT, Hoel DG. 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. Nature 1983;303:767-70.

Table 1. For interval cancer, OPERA (95% CI) estimates of odds ratio per adjusted standard deviation from univariable and multivariable analyses of *Cumulus* (as a percentage), *Cirrocumulus* and *Cirrus*, all normalised, adjusted for age and body mass index, and standardised.

Measure	**Interval cancer**						
	Univariable analyses			**Multivariable analyses**			
Cumulus	2.32 (1.85 to 2.91)	–	–	2.23 (1.61 to 3.10)	1.81 (1.41 to 2.31)	–	1.71 (1.19 to 2.44)
Cirrocumulus	–	1.84 (1.50 to 2.24)	–	1.05 (0.78 to 1.41)	–	1.52 (1.23 to 1.88)	1.07 (0.79 to 1.46)
Cirrus	–	–	2.13 (1.74 to 2.61)	–	1.72 (1.38 to 2.14)	1.91 (1.55 to 2.36)	1.72 (1.38 to 2.14)
-2 LL[1]	66.66	41.06	67.36	66.75	91.88	83.20	92.07
AUC[2]	0.70 (0.65 to 0.74)	0.65 (0.60 to 0.70)	0.71 (0.66 to 0.75)	0.69 (0.64 to 0.73)	0.73 (0.68 to 0.77)	0.72 (0.69 to 0.77)	0.72 (0.68 to 0.76)
OPERA[3]				2.00 (1.68 to 2.42)	2.35 (1.97 to 2.84)	2.31 (1.93 to 2.80)	2.28 (1.90 to 2.76)

[1] LL = log likelihood relevant to that for the null model; [2] AUC = are under the receiver operating characteristic curve; [3] OPERA = change in odds per unit standard deviation for the controls adjusted for age and body mass index

Table 2. For *Cirrus* and *Cirrocumulus*, risk relative to the population average, with 95% CI in parentheses, and numbers of cases and controls as a ratio, by tertile-by-tertile.

		Cirrus								
Cirrocumulus		Interval			Screen-detected			Younger-Diagnosis		
		low		high	low		high	low		high
low		0.27 (0.10 to 0.59)	0.59 (0.26 to 1.22)	1.17 (0.58 to 2.24)	0.50 (0.32 to 0.75)	0.58 (0.37 to 0.91)	1.42 (0.97 to 2.06)	0.25 (0.13 to 0.44)	0.35 (0.17 to 0.66)	0.53 (0.28 to 0.95)
		7/78	10/50	15/38	30/170	27/131	49/98	14/149	12/91	15/75
		0.20 (0.05 to 0.53)	0.83 (0.43 to 1.52)	1.45 (0.81 to 2.52)	0.82 (0.55 to 1.18)	0.76 (0.50 to 1.14)	1.22 (0.85 to 1.72)	0.73 (0.46 to 1.13)	0.85 (0.56 to 1.27)	1.89 (1.33 to 2.69)
		4/62	16/57	23/47	42/146	34/127	54/126	29/106	37/116	66/93
high		0.46 (0.11 to 1.34)	1.11 (0.62 to 1.90)	2.45 (1.67 to 3.60)	0.58 (0.32 to 1.00)	0.97 (0.67 to 1.38)	1.96 (1.50 to 2.55)	0.84 (0.47 to 1.46)	0.69 (0.43 to 1.08)	2.45 (1.86 to 3.21)
		4/26	22/59	67/81	17/83	48/141	121/175	19/60	28/108	134/146

Table 3. For screen-detected cancer, OPERA (95% CI) estimates of odds ratio per adjusted standard deviation from univariable and multivariable analyses of *Cumulus* (as a percentage), *Cirrocumulus* and *Cirrus*, all normalised, adjusted for age and body mass index, and standardised.

Measure	**Screen-detected cancer**						
	Univariable analyses			**Multivariable analyses**			
Cumulus	1.29 (1.15 to 1.45)	–	–	1.12 (0.94 to 1.33)	1.03 (0.90 to 1.18)	–	0.81 (0.66 to 0.99)
Cirrocumulus	–	1.32 (1.18 to 1.48)	–	1.22 (1.02 to 1.44)	–	1.16 (1.03 to 1.31)	1.34 (1.12 to 1.61)
Cirrus	–	–	1.55 (1.39 to 1.74)	–	1.53 (1.35 to 1.74)	1.49 (1.32 to 1.67)	1.59 (1.39 to 1.81)
-2 LL[1]	20.11	23.41	63.69	24.97	63.89	69.81	74.18
AUC[2]	0.57 (0.54 to 0.60)	0.57 (0.54 to 0.61)	0.62 (0.59 to 0.65)	0.57 (0.54 to 0.61)	0.62 (0.59 to 0.65)	0.63 (0.60 to 0.66)	0.63 (0.60 to 0.66)
OPERA[3]				1.31 (1.16 to 1.47)	1.54 (1.37 to 1.74)	1.58 (1.41 to 1.79)	1.59 (1.41 to 1.79)

[1] LL = log likelihood relevant to that for the null model; [2] AUC = are under the receiver operating characteristic curve; [3] OPERA = change in odds per unit standard deviation for the controls adjusted for age and body mass index

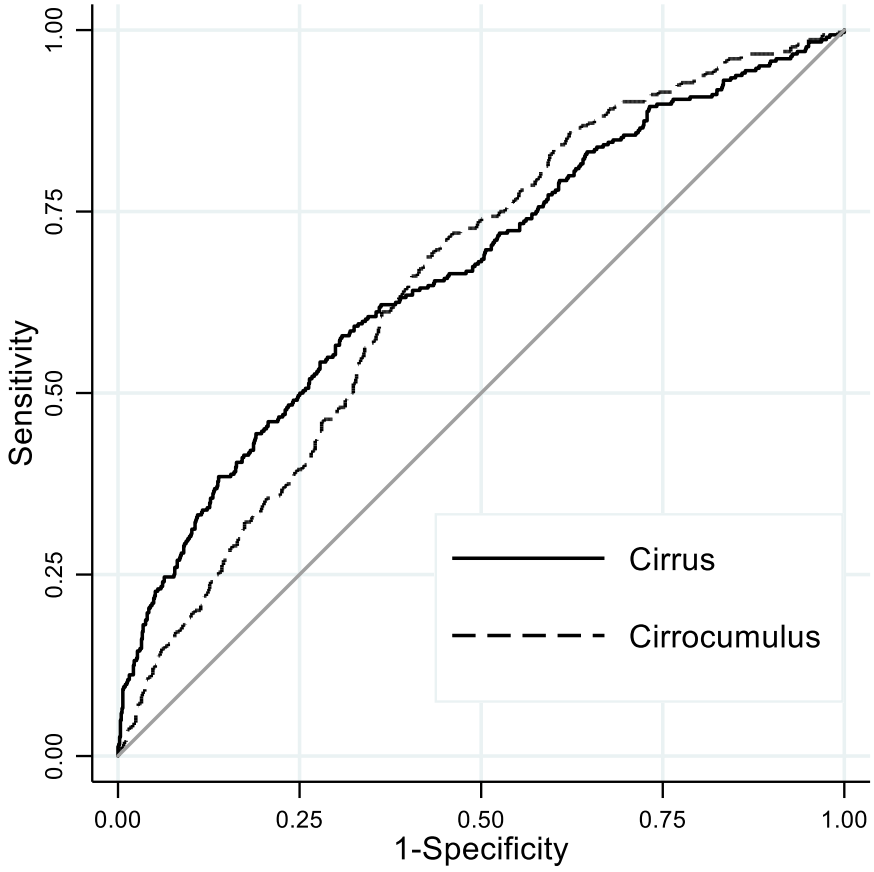
Table 4. For younger-diagnosis cancer, OPERA (95% CI) estimates of odds ratio per adjusted standard deviation from univariable and multivariable analyses of *Cumulus* (as a percentage), *Cirrocumulus* and *Cirrus*, all normalised, adjusted for age and body mass index, and standardised.

Measure	**Younger-diagnosis cancer**						
	Univariable analyses			**Multivariable analyses**			
Cumulus	1.51 (1.32 to 1.71)	–	–	1.15 (0.98 to 1.35)	1.14 (0.99 to 1.32)	–	0.90 (0.75 to 1.07)
Cirrocumulus	–	1.74 (1.52 to 1.98)	–	1.60 (1.37 to 1.88)	–	1.46 (1.27 to 1.68)	1.54 (1.31 to 1.82)
Cirrus	–	–	1.89 (1.66 to 2.15)	–	1.79 (1.55 to 2.06)	1.70 (1.48 to 1.94)	1.75 (1.52 to 2.02)
-2 LL[1]	52.38	84.60	119.59	87.52	122.86	149.39	150.87
AUC[2]	0.63 (0.60 to 0.66)	0.67 (0.64 to 0.70)	0.70 (0.67 to 0.73)	0.67 (0.64 to 0.70)	0.70 (0.67 to 0.73)	0.72 (0.69 to 0.75)	0.72 (0.69 to 0.75)
OPERA[3]				1.87 (1.66 to 2.12)	2.10 (1.85 to 2.40)	2.28 (2.01 to 2.61)	2.29 (2.02 to 2.62)

[1] LL = log likelihood relevant to that for the null model; [2] AUC = are under the receiver operating characteristic curve; [3] OPERA = change in odds per unit standard deviation for the controls adjusted for age and body mass index

Legend to Figures

Figure 1: Receiver operating characteristic (ROC) curves for *Cirrocumulus* and *Cirrus* for case-control study of younger-diagnosis cancer.

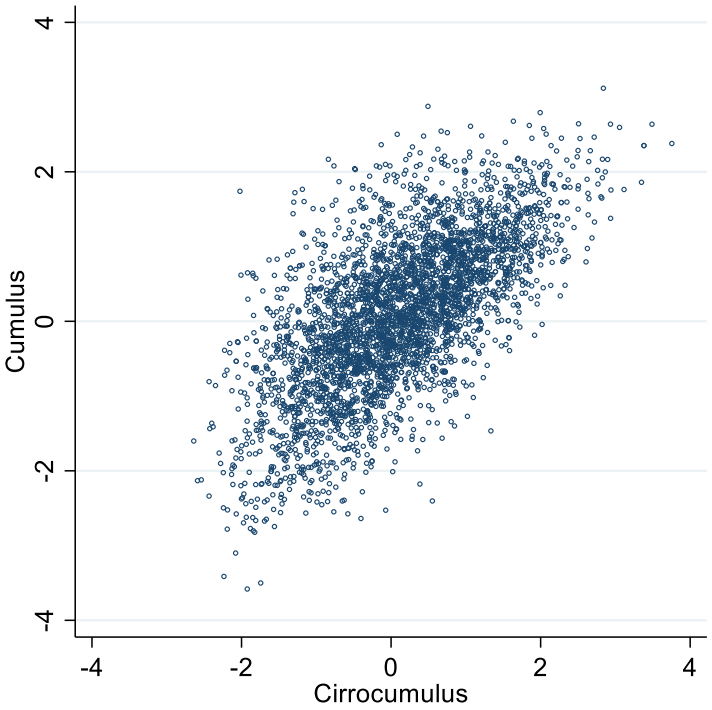


Supplementary Table 1. For *Cirrus* and *Cumulus*, risk relative to the population average, with 95% CI in parentheses, and numbers of cases and controls as a ratio, by tertile-by-tertile for interval cancer.

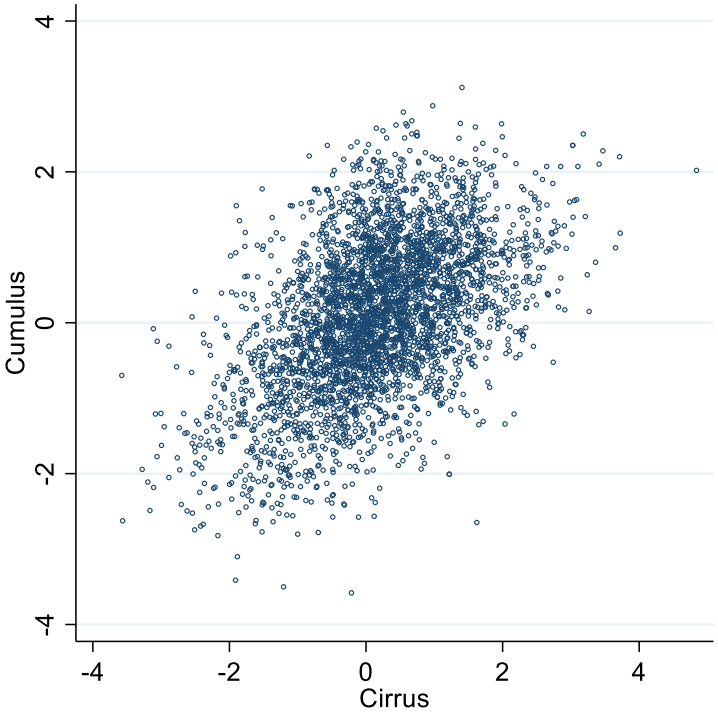
		Cirrus									
Cumulus		Interval				Screen-detected				Younger-Diagnosis	
		low		high		low		high		low		high
low		0.19 (0.07 to 0.44)	0.54 (0.23 to 1.15)	0.86 (0.31 to 2.12)		0.60 (0.42 to 0.84)	0.76 (0.49 to 1.16)	1.30 (0.79 to 2.10)		0.44 (0.29 to 0.66)	0.74 (0.45 to 1.20)	1.27 (0.70 to 2.25)
		6/93	9/49	7/24		46/219	32/119	28/61		31/187	24/86	20/42
		0.21 (0.06 to 0.59)	0.92 (0.48 to 1.66)	1.40 (0.82 to 2.36)		0.59 (0.37 to 0.92)	0.78 (0.53 to 1.12)	1.42 (0.99 to 2.01)		0.57 (0.33 to 0.93)	0.62 (0.38 to 0.97)	1.62 (1.13 to 2.29)
		4/56	17/55	26/55		27/129	43/156	57/114		21/99	26/112	63/104
high		0.87 (0.25 to 2.51)	1.05 (0.60 to 1.80)	2.45 (1.68 to 3.56)		0.89 (0.47 to 1.61)	0.78 (0.51 to 1.17)	1.76 (1.37 to 2.25)		0.92 (0.40 to 1.97)	0.62 (0.38 to 0.96)	2.10 (1.60 to 2.73)
		5/17	22/62	72/87		16/51	34/124	139/224		10/29	27/117	132/168

Supplementary Figure 1

1(a) *Cumulus* versus *Cirrocumulus*



1(b) *Cumulus* versus *Cirrus*



1(c) *Cirrocumulus* versus *Cirrus*