



## Measurement challenge: protocol for international case–control comparison of mammographic measures that predict breast cancer risk

Dench, Evenda; Bond-smith, Daniela; Darcey, Ellie; Lee, Grant; Aung, Ye K; Chan, Ariane; Cuzick, Jack; Ding, Ze Y; Evans, Chris F; Harvey, Jennifer; Highnam, Ralph; Hsieh, Meng-kang; Kontos, Despina; Li, Shuai; Mariapun, Shivaani; Nickson, Carolyn; Nguyen, Tuong L; Pertuz, Said; Procopio, Pietro; Rajaram, Nadia; Repich, Kathy; Tan, Maxine; Teo, Soohwang; Trinh, Nhut Ho; Ursin, Giske; Wang, Chao; Dos-santos-silva, Isabel; McCormack, Valerie; Nielsen, Mads; Shepherd, John; Hopper, John L; Stone, Jennifer

*Published in:*  
BMJ Open

*DOI:*  
[10.1136/bmjopen-2019-031041](https://doi.org/10.1136/bmjopen-2019-031041)

*Publication date:*  
2019

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/)

*Citation for published version (APA):*  
Dench, E., Bond-smith, D., Darcey, E., Lee, G., Aung, Y. K., Chan, A., ... Stone, J. (2019). Measurement challenge: protocol for international case–control comparison of mammographic measures that predict breast cancer risk. *BMJ Open*, 9(12), 1-6. [e031041]. <https://doi.org/10.1136/bmjopen-2019-031041>

# BMJ Open Measurement challenge: protocol for international case-control comparison of mammographic measures that predict breast cancer risk

Evenda Dench <sup>1</sup>, Daniela Bond-Smith,<sup>2</sup> Ellie Darcey,<sup>1</sup> Grant Lee,<sup>3</sup> Ye K Aung,<sup>3</sup> Ariane Chan,<sup>4</sup> Jack Cuzick <sup>5</sup>, Ze Y Ding,<sup>6</sup> Chris F Evans,<sup>3</sup> Jennifer Harvey,<sup>7</sup> Ralph Highnam,<sup>4</sup> Meng-Kang Hsieh,<sup>8</sup> Despina Kontos,<sup>8</sup> Shuai Li,<sup>3</sup> Shivaani Mariapun,<sup>9,10</sup> Carolyn Nickson,<sup>3,11</sup> Tuong L Nguyen <sup>3</sup>, Said Pertuz,<sup>12,13</sup> Pietro Procopio,<sup>3,11</sup> Nadia Rajaram,<sup>9,10</sup> Kathy Repich,<sup>7</sup> Maxine Tan,<sup>6,14</sup> Soo-Hwang Teo,<sup>9</sup> Nhut Ho Trinh,<sup>3</sup> Giske Ursin,<sup>15</sup> Chao Wang <sup>16</sup>, Isabel dos-Santos-Silva,<sup>17</sup> Valerie McCormack,<sup>18</sup> Mads Nielsen,<sup>19</sup> John Shepherd,<sup>20</sup> John L Hopper,<sup>3</sup> Jennifer Stone<sup>1</sup>

**To cite:** Dench E, Bond-Smith D, Darcey E, *et al.* Measurement challenge: protocol for international case-control comparison of mammographic measures that predict breast cancer risk. *BMJ Open* 2019;**9**:e031041. doi:10.1136/bmjopen-2019-031041

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-031041>).

Received 18 April 2019

Revised 29 November 2019

Accepted 04 December 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Jennifer Stone;  
[jennifer.stone@uwa.edu.au](mailto:jennifer.stone@uwa.edu.au)

## ABSTRACT

**Introduction** For women of the same age and body mass index, increased mammographic density is one of the strongest predictors of breast cancer risk. There are multiple methods of measuring mammographic density and other features in a mammogram that could potentially be used in a screening setting to identify and target women at high risk of developing breast cancer. However, it is unclear which measurement method provides the strongest predictor of breast cancer risk.

**Methods and analysis** The measurement challenge has been established as an international resource to offer a common set of anonymised mammogram images for measurement and analysis. To date, full field digital mammogram images and core data from 1650 cases and 1929 controls from five countries have been collated. The measurement challenge is an ongoing collaboration and we are continuing to expand the resource to include additional image sets across different populations (from contributors) and to compare additional measurement methods (by challengers). The intended use of the measurement challenge resource is for refinement and validation of new and existing mammographic measurement methods. The measurement challenge resource provides a standardised dataset of mammographic images and core data that enables investigators to directly compare methods of measuring mammographic density or other mammographic features in case/control sets of both raw and processed images, for the purposes of the comparing their predictions of breast cancer risk.

**Ethics and dissemination** Challengers and contributors are required to enter a Research Collaboration Agreement with the University of Melbourne prior to participation in the measurement challenge. The Challenge database of collated data and images are stored in a secure data repository at the University of Melbourne. Ethics approval for the measurement challenge is held at University of Melbourne (HREC ID 0931343.3).

## Strengths and limitations of this study

- The measurement challenge is a unique annotated resource that enables timely and prompt comparison of different mammographic measurement methods applied to full field digital mammogram images, and thereby robust assessment of which measurement method provides the 'best' predictor of breast cancer risk within different screening settings.
- The measurement challenge currently includes anonymised mammographic images and core data from 1650 cases and 1929 controls from five different sample populations.
- The measurement challenge is an ongoing collaboration and we are continuing to expand the resource to include additional image sets across different populations (from contributors) and to apply additional measurement methods (by challengers).
- The measurement challenge provides a much-needed standardised dataset with which to refine and assess new or existing mammographic measurement methods but is currently underpowered to examine the association between mammographic measures and breast cancer risk in non-European/Caucasian populations.

## INTRODUCTION

Mammograms are the current standard in population-based screening for breast cancer in women aged 40 or above. For women of the same age and body mass index (BMI), increased mammographic density is one of the strongest predictors of breast cancer risk.<sup>1-3</sup> There are multiple methods of measuring mammographic density and other features of a mammogram that could potentially be used in a screening setting



to identify and target women at high risk of developing breast cancer. However, it is unclear which measurement method provides the strongest predictor of breast cancer risk.

The measurement challenge is an international resource that offers a common set of raw and/or processed full field digital mammograms for measurement and analysis. By providing an annotated dataset of mammogram images and core data, this resource will enable investigators to directly compare methods of measuring mammographic density or other mammographic features as predictors of breast cancer risk.

Initiated in 2016, the measurement challenge has collated full field digital mammograms images and core data from 1650 cases and 1929 controls from five countries (Australia, Malaysia, Norway, UK and the USA) to date. Images have been randomly divided to form a training set and a validation set for challenge measurement. Further contributions are welcomed.

There are currently eight challenger research groups who have applied measurement methods to the resource dataset. These include fully automated approaches (ie, AutoDensity, BIOMEDIQ, DeepRisk, Laboratory for Individualized Breast Radiodensity Assessment (LIBRA), MMTEXT, OpenBreast V.1.0 and VolparaDensity) and semiautomated methods (ie, Cumulus, including AltoCumulus and CirroCumulus).

The measurement challenge is an ongoing collaboration and we are continuing to expand the resource to include additional image sets across different populations and to apply additional measurement methods. More image sets are needed to improve statistical power and provide sufficient data to address incompatibilities between measurement methods and image type and/or manufacturer.

The potential benefit of pooling and standardising mammographic images from different populations includes providing a large set of images with which to refine and assess new or existing mammographic measurement methods. We anticipate that the measurement challenge will become a unique annotated resource for timely and prompt comparison of different mammographic measurement methods with the aim of identifying mammographic measures that the best predict breast cancer risk in different screening settings. Given the increased measurement and reporting of mammographic density internationally, identifying the measurement method/s that best discriminates women who will or will not be diagnosed with breast cancer in the future could substantially improve existing risk prediction models and ultimately help improve the efficiency and effectiveness of screening programs by triaging women according to risk. The cost and ease of implementing routine mammographic measurement in local clinical settings would depend on the requirements and idiosyncrasies of each screening programmes.

## METHODS AND ANALYSIS

As part of the first phase of the measurement challenge, full field digital mammogram images and core data from 1650 cases and 1929 controls have been collated. The resource has been collated from five contributing studies in Australia, Malaysia, Norway, the UK and the USA (henceforth contributors). Both contributors of case-control image sets and investigators interested in applying their measurement methods (henceforth challengers) are required to enter a Research Collaboration Agreement with The University of Melbourne prior to participation in the measurement challenge.

Data are obtained from existing epidemiological case-control studies or nested case-control studies. For the former, diagnostic or prediagnostic mammograms are obtained for cases with comparable mammograms obtained for controls. For nested case-control studies, mammograms taken at the time of entry into the cohort are used and incident cancers diagnosed within 1 year after the baseline screen negative mammogram are excluded to avoid masking bias or false negatives. Control selection differed slightly for each study (references provided in [table 1](#)); at a minimum, controls were frequency matched for age post hoc.

[Table 1](#) describes the mammographic image data available by contributor, to date. Both raw and processed images are available for 1003 cases and 1283 controls; processed images are available for 1650 cases and 1929 controls. Core participant data consist of year at mammogram, year of birth, year of interview, BMI at mammogram and ethnicity. For breast cancer cases, breast cancer laterality and year of diagnosis is also available. Additional image data include laterality, view and mammography machine manufacturer.

Anonymised mammographic images and core data are available to challengers interested in applying their measurement methods in two sets. Images from each source are randomly divided (50:50) in two (stratified by case-control status) and assigned to a 'test set' or a 'validation set'. The test set is intended for use in refining new measurement methods or for training purposes, and the corresponding core data includes breast cancer status. It is made available to all challengers for a specified period of time. The validation set is provided for measurement validation and the corresponding core data does not include any information regarding breast cancer status. Both sets of images are securely transferred to challengers for the measurement using their method (multiple methods are permitted).

Thus far, eight challengers from Australia, Denmark, Finland, Malaysia, New Zealand, the UK and the USA have accessed the measurement challenge core data. Measurement methods include AutoDensity,<sup>4</sup> BIOMEDIQ,<sup>5</sup> Cumulus (Sunnybrook Women's Hospital, Toronto Canada), including AltoCumulus and CirroCumulus,<sup>6</sup> DeepRisk, LIBRA,<sup>7</sup> MMTEXT,<sup>8</sup> OpenBreast V.1.0<sup>9</sup> and VolparaDensity.<sup>10</sup> AutoDensity automatically preprocesses images by identifying the breast area using

**Table 1** Characteristics of available measurement challenge images and data

Country (where images were sourced)	Study design	No case/controls	Ethnic majority	Mean age (SD)	Mean BMI (SD)	Image type	Mammogram view	Machine manufacturer
Australia <sup>12</sup>	Nested Case/control	464/463	European	62.9 (7.1)	27.3 (5.4)	P	CC, MLO	Fuji, Hologic, Siemens
Malaysia <sup>13</sup>	Case/control	202/202	Chinese	50.9 (11.5)	23.4 (3.8)	R, P	CC, MLO	Hologic
UK <sup>14</sup>	Case/control	403/683	European	62.1 (9.8)	26.2 (5.4)	R, P	CC	GE
USA <sup>15</sup>	Case/control	398/398	European	55.1 (10.9)	27.8 (6.3)	R, P	CC, MLO	GE, Hologic
Scandinavia <sup>11</sup>	Nested Case/control	183/183	European	59.7 (6.2)	24.7 (3.4)	P	CC, MLO	GE
Combined (all data)		1650/1929	European	59.3 (10.2)	26.2 (5.3)	R, P	CC, MLO	GE, Hologic

CC, craniocaudal; GE, general electric; MLO, mediolateral-oblique; P, processed image; R, raw image.

thresholding algorithms (drawing on header information where available, or otherwise through iterative optimisation), removing background objects such as image tags and image acquisition artefacts, and then reducing noise and improving image contrast. AutoDensity then segments the breast area into dense and non-dense tissue by identifying an optimal intensity threshold, outputting image files of the breast area and dense tissue data tables including the image/woman identifier, breast area, dense area and percent density. BIOMEDIQ is a deep-learning-based mammographic texture-based risk assessment. The deep-learning framework has a 5-layer convolutional neural network that maps mammographic patches to a cancer risk score. The Cumulus measures define mammographic density at successively higher pixel brightness thresholds. DeepRisk uses deep-learning-based methods to estimate breast cancer risk on mammograms. Deep learning differs from other previous methods in that it automatically determines and extracts useful features using deep-learning-based methods without relying on manual extraction of features by experts. LIBRA is an adaptive multicluster fuzzy c-means segmentation algorithm. For the MMTEXT algorithm, images are first downsized,<sup>11</sup> leading to a new image for which each pixel had a much larger physical area than the original. A gray-level co-occurrence matrix (GLCM) is formed based on pixels within the breast area in each image, and a GLCM summary statistic ('sum average') is calculated for prediction. OpenBreast V.1.0 is an open software for breast cancer risk assessment based on fully automated parenchymal analysis of mammographic images. Parenchymal analysis is performed by applying computerised texture descriptors on each mammogram, followed by stepwise feature selection allowing first order interactions between features. The Volpara algorithm used an internal fat reference point within the mammogram to convert the intensities at each pixel to a thickness of fibroglandular tissue.

All challengers are required to provide their measurements from the validation set for statistical analysis at the University of Western Australia. Statistical criteria for assessing and comparing performance of each measurement method are described in the Overview of analysis protocol below and will be completed by a statistician with no vested interest in the outcome.

### Overview of analysis protocol

To prepare the data for analysis, we exclude images of the affected breast for all cases. We also ensure that age distributions are similar across cases and controls. Age frequency matching is performed randomly within rolling 5-year intervals and separately by each image type and source.

For most women, four mammograms are available (ie, both sides and views). If multiple measurements have been submitted for the same person, we average these separately for each image type. The VolparaDensity challenger has provided a specific custom average which is



used. Details regarding its calculation can be provided on request.

Not all challengers can measure each image due to image type and/or quality; measured proportions will be reported by technique. A complete dataset that contains all images that every challenger can measure is used for the principal analysis. In addition, we also run all calculations individually for each challenger using all images that this specific challenger submitted.

Logistic regression is used to estimate the association between the mammographic measure and breast cancer risk for each method, stratifying by image type (raw or processed). Adjustment for covariates includes age, BMI, machine manufacturer and source.

Box-Cox transformations are used to better approximate normality,

$$f(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(x) & \text{if } \lambda = 0 \end{cases}$$

where  $x$  is the mammographic measurement and  $\lambda$  is chosen to maximise the profile likelihood of a linear model of  $x$ . The value of  $\lambda$  obtained and used for transformation will be reported for each measurement method.

We calculate standard ORs per unit increase in mammographic measure and as ORs per adjusted SD (OPERAs).<sup>1</sup> To calculate OPERAs, the control population mammographic measures are regressed on the adjustment covariates. The standardised residuals from this regression are an adjusted and independent mammographic measure that effectively controls for key population differences, strengthening our ability to compare ORs across measurement approaches. OPERAs are used to calculate any further metrics for assessment.

The following diagnostic statistics are considered for the assessment:

**Akaike's information criterion (AIC):** AIC is a diagnostic statistic derived from information theory. It is calculated as minus twice the maximised log-likelihood plus twice the number of parameters. AIC is a relative statistic capturing information loss, here relative to a reference model without the mammographic measure. The smaller the AIC, the stronger the model.

**Brier score:** The Brier score can be used as a calibration measure reflecting a loss function. It is calculated as the mean of the squared residuals. Again, the model with the lowest Brier score is preferred.

**Area under the receiver operating curve (AUC):** AUC is a measure of discrimination between cases and controls. ROC curves visualise the trade-off between sensitivity and specificity visualised on a plane erected by sensitivity and 1-specificity. A higher AUC means a better classification. They are calculated using the trapezoidal rule. We will also report a CI for the AUC bootstrapped 1000 iterations.

**Partial AUC (pAUC):** We will also report pAUC focusing on the section of the Receiver Operating Characteristic (ROC) curve between 0.9 and 1 (high pAUC)

and between 0 and 0.1 (low pAUC). CIs will be calculated and reported as for the aggregate AUC.

**Adjusted for covariate AUC (AAUC):** The AAUC can be adjusted for covariates. For analysing the mammographic measure–breast cancer association, age and BMI are critical confounders. Adjusting the AUC takes into account that covariates may affect the control distribution of the mammographic measure such that the underlying probability thresholds for discriminating between cases and controls are modelled to vary with age and BMI.

**Net Reclassification Index (NRI):** The (continuous) NRI reflects the proportion of cases correctly assigned a higher breast cancer risk and controls correctly assigned a lower breast cancer risk when moving from a null model to one including a mammographic measure as a predictor. It is calculated as the sum of  $NRI_{cases}$  and  $NRI_{controls}$ , where  $NRI_{cases}$  is the proportion of cases that experience an increase in breast cancer risk minus the proportion with a decrease, and  $NRI_{controls}$  is the proportion of controls with a decrease minus the proportion with an increase in breast cancer risk.

**Integrated discrimination improvement (IDI):** The IDI is calculated similarly to the NRI, with  $IDI_{cases}/IDI_{controls}$  representing the difference between the mean of the predicted cancer risk from a model containing a mammographic measure as a predictor and the mean risk probability from the reference model without the mammographic measure for cases/controls, respectively. This means that the  $IDI_{cases}$  capture the difference in the average sensitivity of the two models and similarly, the  $IDI_{controls}$  represent the difference in the average 1-specificity.

**Comparisons:** We will also test for equality of AUCs, using the AUCs. P values are reported for all possible individual comparisons of the methods by mammogram type as well as a joint comparison of all methods by mammogram type (testing for equality of all AUCs). Hence, small p values suggest that the challengers' methods produce significantly different results.

**Other considerations:** Large sample sizes are needed to facilitate stratification by type of image (eg, processed vs raw; machine manufacturer) and ethnicity. The resource includes a mix of diagnostic and prediagnostic mammograms and only the contralateral mammogram is used for breast cancer cases.

## PATIENT AND PUBLIC INVOLVEMENT

No patients are involved in this study. Patient data and images provided by contributors are supplied in an anonymised format, so that the data cannot be linked to personal information. Challengers agree to treat the Challenge Database and its contents as Confidential Information of the University (or challenge contributor as the case may be); and not attempt to identify or identify the individuals whose information appears in the images and data.

## ETHICS AND DISSEMINATION

The measurement challenge is an ongoing collaboration and we are continuing to expand the resource to include additional image sets across different populations (from contributors) and to apply additional measurement methods (by challengers). Contributors are invited to submit large sets of case–control mammographic images (raw and processed full field digital mammograms) from a range of populations, plus corresponding core data timed to the mammogram. Each contributor is responsible for the ethics approval and participant consent for their respective sample populations; written informed consent is expected unless a waiver for informed consent is obtained. Challengers are invited to apply their measurement methods to the validation set of case–controls images and provide their measurements for central analysis. All contributors and challengers are required to enter a Research Collaboration Agreement with the University of Melbourne prior to participation in the measurement challenge.

The Challenge database of collated data and images will be stored in a secure data repository at the University of Melbourne.

The measurement challenge is being organised by an international committee including: JLH (University of Melbourne), JeS (University of Western Australia), JoS (UCSF), IdSS (LSHTM), MN (University of Copenhagen), VM (IARC). To collaborate or request access to data and images from the measurement challenge, contact the corresponding author at [jennifer.stone@uwa.edu.au](mailto:jennifer.stone@uwa.edu.au).

## ADDITIONAL CONSIDERATIONS

The measurement challenge is a unique annotated resource that enables timely and prompt comparison of different mammographic measurement methods applied to full field digital mammogram images. However, full field digital mammograms differ by image type (processed vs unprocessed) and the processing differs by machine manufacturer, potentially introducing systematic bias to risk estimates. Therefore, more image sets are needed to improve statistical power and provide sufficient data to address incompatibilities between measurement methods and image type and/or manufacturer. Also, screening programmes differ by ethnicity of participants, equipment used (eg, manufacturer) and screening protocols (eg, only a small proportion store unprocessed images). Therefore, we are comparing different mammographic measures within different screening settings to inform potential users about which measurement method provides the ‘best’ predictor of breast cancer risk. Identifying women at increased risk due to high breast density could help target those who could benefit from supplemental screening using ultrasound, MRI or other modalities. Clinical trials are currently underway to determine evidence-based screening recommendations for women with dense breasts.

## Author affiliations

- <sup>1</sup>Centre for Genetic Origins of Health and Disease, University of Western Australia, Crawley, Western Australia, Australia
- <sup>2</sup>School of Population and Global Health, University of Western Australia, Crawley, Western Australia, Australia
- <sup>3</sup>Centre for Epidemiology & Biostatistics, The University of Melbourne School of Population and Global Health, Melbourne, Victoria, Australia
- <sup>4</sup>Science and Technology, Volpara Health Technologies, Wellington, New Zealand
- <sup>5</sup>Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
- <sup>6</sup>Electrical and Computer Systems Engineering, School of Engineering, Monash University - Malaysia Campus, Bandar Sunway, Selangor, Malaysia
- <sup>7</sup>Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, Virginia, USA
- <sup>8</sup>Department of Radiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA
- <sup>9</sup>Cancer Research Malaysia, Subang Jaya, Selangor, Malaysia
- <sup>10</sup>Department of Applied Mathematics, Faculty of Engineering, University of Nottingham - Malaysia Campus, Semenyih, Selangor, Malaysia
- <sup>11</sup>Cancer Research Division, Cancer Council New South Wales, Sydney, New South Wales, Australia
- <sup>12</sup>Laboratory of Signal Processing, Tampere University of Technology, Tampere, Pirkanmaa, Finland
- <sup>13</sup>Connectivity and Signal Processing group, Universidad Industrial de Santander, Bucaramanga, Colombia
- <sup>14</sup>School of Electrical and Computer Engineering, University of Oklahoma Norman Campus, Norman, Oklahoma, USA
- <sup>15</sup>Cancer Registry of Norway, Oslo, Norway
- <sup>16</sup>Faculty of Health, Social Care and Education, Kingston University and St George's, University of London, Kingston-Upon-Thames, UK
- <sup>17</sup>Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
- <sup>18</sup>Section of Environment and Radiation, International Agency for Research on Cancer, IARC, Lyon, France
- <sup>19</sup>University of Copenhagen, Copenhagen, Denmark
- <sup>20</sup>University of Hawai'i Cancer Center, Honolulu, Hawaii, USA

**Twitter** Chao Wang @excel\_wang and Jennifer Stone @DrJenniferStone

**Contributors** JeS, JLH, JoS, IdSS, MN and VM are the Organizing Committee of the measurement challenge and made substantial contributions to the conception or design of the work. JeH, SM, NR, KR, S-HT, GU, IdSS and JLH are phase 1 contributors of images and data who made substantial contributions to the acquisition of data for the work. GL facilitates the capture, renaming, security and distribution of images in the challenge, and is involved in design and conduct of the study. DB-S and EID performed the statistical analyses and made a substantial contribution to analysis of the work. YKA, AC, JC, ZYD, CFE, RH, M-KH, DK, SL, CN, TLN, SP, PP, MT, NHT, CW, MN and JLH are phase 1 challengers who have made a substantial contribution to interpretation of data for the work. EvD and JeS prepared the manuscript and all authors revised it critically for important intellectual content. All authors provided final approval of the version to be published and agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** Ethics approval for the Challenge is held at University of Melbourne (HREC ID 0931343.3).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.



## ORCID iDs

Evenda Dench <http://orcid.org/0000-0002-4513-6628>

Jack Cuzick <https://orcid.org/0000-0001-7420-7512>

Tuong L Nguyen <http://orcid.org/0000-0002-6597-8363>

Chao Wang <https://orcid.org/0000-0002-7873-0594>

## REFERENCES

- 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;182:863–7.
- McCormack VA, dos Santos Silva I, Silva dosS I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006;15:1159–69.
- Bond-Smith D, Stone J. Methodological challenges and updated findings from a meta-analysis of the association between mammographic density and breast cancer. *Cancer Epidemiol Biomarkers Prev* 2019;28:22–31.
- Nickson C, Arzhaeva Y, Aitken Z, et al. AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes. *Breast Cancer Res* 2013;15.
- Wanders JOP, van Gils CH, Karssemeijer N, et al. The combined effect of mammographic texture and density on breast cancer risk: a cohort study. *Breast Cancer Res* 2018;20.
- 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;46:652–61.
- Keller BM, Chen J, Daye D, et al. Preliminary evaluation of the publicly available laboratory for breast Radiodensity assessment (LIBRA) software tool: comparison of fully automated area and volumetric density measures in a case-control study with digital mammography. *Breast Cancer Res* 2015;17:117.
- Wang C, Brentnall AR, Cuzick J, et al. A novel and fully automated mammographic texture analysis for risk prediction: results from two case-control studies. *Breast Cancer Res* 2017;19.
- Pertuz S, Torres GF, Tamimi R, IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). *Open framework for Mammography-based breast cancer risk assessment*. Chicago, IL, 2019.
- Highnam R, Brady SM, Yaffe MJ, et al. Robust breast composition measurement - volpara™. In: Martí J, Oliver A, Freixenet J, et al, eds. *Digital mammography*. Berlin, Heidelberg: Springer, 2010.
- Ellingjord-Dale M, Vos L, Hjerkind KV, et al. Alcohol, physical activity, smoking, and breast cancer subtypes in a large, nested case-control study from the Norwegian breast cancer screening program. *Cancer Epidemiol Biomarkers Prev* 2017;26:1736–44.
- Nickson C, Procopio P, Velentz LS, et al. Prospective validation of the NCI breast cancer risk assessment tool (Gail model) on 40,000 Australian women. *Breast Cancer Res* 2018;20.
- Tan M-M, Ho W-K, Yoon S-Y, et al. A case-control study of breast cancer risk factors in 7,663 women in Malaysia. *PLoS One* 2018;13:e0203469.
- Eng A, Gallant Z, Shepherd J, et al. Digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods. *Breast Cancer Res* 2014;16.
- Brentnall AR, Cohn WF, Knaus WA, et al. A case-control study to add volumetric or clinical mammographic density into the Tyrer-Cuzick breast cancer risk model. *J Breast Imaging* 2019;1:99–106.