Received:
27 April 2017

Revised:
11 June 2017

Accepted:
20 June 2017

<https://doi.org/10.1259/bjr.20170307>

Cite this article as:

Abdolell M, Tsuruda KM, Brown P, Caines JS, Iles SE. Breast density scales: the metric matters. *Br J Radiol* 2017; **90**: 20170307.

SHORT COMMUNICATION

Breast density scales: the metric matters

^{1,2}MOHAMED ABDOLELL, MSc, PStat, ¹KAITLYN M TSURUDA, MSc, ^{1,2}PETER BROWN, MD, ^{2,3}JUDY S CAINES, MD and ^{1,2,3,4}SIÂN E ILES, MD

¹Department of Diagnostic Imaging, Nova Scotia Health Authority, Halifax, NS, Canada

²Department of Diagnostic Radiology, Dalhousie University, Faculty of Medicine, Halifax, NS, Canada

³Department of Diagnostic Radiology, IWK Health Centre, Breast Health Section, Halifax, NS, Canada

⁴Nova Scotia Breast Screening Program, IWK Health Centre, Halifax, NS, Canada

Address correspondence to: Prof Mohamed Abdolell

E-mail: mo@dal.ca

Objective: Measures of percent mammographic density (PMD) are often categorized using various density scales. The purpose of this study was to examine information loss associated with the use of categorical density scales.

Methods: Baseline PMD was assessed at 1% precision for 2,374 females. The data were used to create 21-category, 4-category and 2-category density scales. R-squared and root mean square error were used to evaluate the effect of categorizing PMD. The area under the receiver operator characteristic curves were compared between cancer risk models employing solely categorical PMD scales and solely baseline PMD for a subset of females (424 cases, 848 controls).

Results: R-squared value decreased from 1.00 (1% PMD) to 0.56 (2-category scale), while root mean square error increased from 0.00 (1% PMD) to 10.83 (2-category scale). The area under the receiver operator characteristic curve

decreased from 0.64 for a cancer risk model using 1% PMD to 0.58 for a risk model using a 21-category density scale ($p < 0.0001$), 0.55 for a 4-category Breast Imaging, Reporting and Data System-like scale ($p < 0.0001$) and 0.50 for a 2-category Breast Imaging, Reporting and Data System-like scale (high vs low) ($p < 0.0001$).

Conclusion: Categorizing PMD measures into categorical density scales leads to a significant loss of information. Indeed, a simple high versus low split of PMD using a 50% cut point yields a cancer risk model with no discriminatory power.

Advances in knowledge: Use of categorical mammographic density scales rather than continuous percent mammographic density measures leads to significant loss of information. Breast cancer risk models using categorical mammographic density scales perform more poorly than models using continuous PMD measures.

INTRODUCTION

Percent mammographic breast density (PMD) is a risk factor for breast cancer. Traditionally, visual assessments of PMD have been used in clinical settings and such measurements have been made using four categories, each spanning a 25% interval of the breast density spectrum (which ranges from 0–100%).¹ In an attempt to improve breast cancer risk models, breast density is increasingly used as a predictor of breast cancer. However, the application of breast density in such models is variable, with some models using breast density as a categorical (sometimes dichotomous) variable and others as a continuous variable.^{2–5}

In medical research, it is often thought that categorizing a continuous variable simplifies the statistical model and the interpretation of the ensuing results. While Occam's razor states that more parsimonious models are preferred, oversimplification can in fact present a number of dilemmas. First, the rationale for choice of cut points is typically ad

hoc and unjustified. Additionally, as outlined by Altman and Royston, dichotomizing a continuous variable results in loss of information and reduced power to detect a relationship between the variable and the outcome of interest.⁶ In fact, when considering the relationship between PMD and a second continuous variable, dichotomization of one variable can be equivalent to losing one third of the sample, and dichotomization of both can be equivalent to losing two thirds of the sample.⁷ Furthermore, categorization may conceal non-linear relationships, and patients with similar measurements found on either side of a cut point may be described as being very different when they are not, resulting in false positive findings.^{6,8} Although less information is lost when variables are split into multiple categories, statistical modelling and interpretation becomes more complex.

While the use of categorical PMD scales is appealing owing to the simplicity of interpretation, the use of such

a measure of PMD assumes equal risk within a given category.⁹ For example, a female with 76% PMD is assumed to have the same risk as a female with 95% PMD if the Breast Imaging, Reporting and Data System fourth edition density classification scale is used.¹ Such categorizations introduce a high degree of variability within subject groups and reduce the power of the model to detect a relationship between breast density and cancer risk, which leads to inaccurate risk estimates. Moreover, when empirically derived quantiles of PMD are used on the basis of a particular study sample, the external validity is difficult to assess and results can be difficult to compare across studies.⁹

The objective of this study was to examine the extent of information loss associated with decreased precision of PMD assessments and the commensurate loss of predictive power in breast cancer risk models in which screen-detected breast cancer is the outcome of interest.

METHODS AND MATERIALS

Two-view, bilateral “for presentation” full-field digital screening mammograms were obtained from 1,272 females attending a population-based breast screening program between 2009 and 2011. These females were part of a case-control study that has been previously described.¹⁰ Briefly, this sample consists of 424 females with pathologically confirmed unilateral screen-detected breast cancer and 848 age-matched controls.

An automated PMD measurement software (DM-Density 2.0, Densitas Inc., Halifax, NS) was used to assess PMD for each mammography image at 1% precision. The mean PMD value of the four images per study (LCC, RCC, LMLO, RMLO) was used

for analysis. Less precise categorical PMD scales were subsequently derived by rounding upward to the nearest 5%, 25% and 50%. A random variable with values ranging from 0 to 100% was also created to evaluate the information-bearing value of a variable that contained no signal (only noise).

The coefficient of determination (R-squared) and root mean square error (RMSE) from univariate linear regression models were used to evaluate the impact of decreasing precision in PMD measures (predictor variable) to predict the baseline assessment of PMD measured in 1% density increments (response variable). R-squared is a goodness-of-fit measure of the percentage of response variable variation that is explained by the univariate linear regression model. An R-squared value of 1 indicates that all of the response variable variation is explained by the model, and a value of 0 indicates that none of the response variable variation is explained by the model. RMSE is a goodness-of-fit measure of how close the predicted values from the linear model are to the observed values; higher observed RMSE values indicate poorer model fit with $RMSE_{max}$ being the theoretical upper limit. Therefore, a value of $1 - \frac{RMSE}{RMSE_{max}} = 1$ indicates a model where the response variable and the predictor variable are identical and the model fits the data perfectly ($RMSE = 0$), whereas a value of $1 - \frac{RMSE}{RMSE_{max}} = 0$ indicates a model that does not fit the data at all ($RMSE = RMSE_{max}$). R-squared, RMSE and $1 - \frac{RMSE}{RMSE_{max}}$ were used to assess the goodness of fit of the linear regression models predicting PMD measured at a precision of 1% as a function of PMD measured increasingly imprecisely (at 5%, 25% and 50%), and finally as a function of random noise, which provided the upper boundary value for $RMSE_{max}$.

Table 1. R-squared and root mean squared error (RMSE) from linear regression models setting the response variable as percent mammographic density (PMD) at 1% precision and the predictor variables as increasingly imprecise categorical PMD scales, culminating with a continuous random PMD variable, and area under the curve (AUC) of the receiver operator characteristic curve for univariate logistic regression risk models setting the response variable as presence of screen-detected breast cancer and the predictor variable as the PMD measures with varying precisions. Larger R-squared and $1 - \frac{RMSE}{RMSE_{max}}$ values (both bounded by 0 and 1) indicate better goodness of fit. Larger AUC values (bounded by 1 and 0.5) indicate greater signal/information carried by the predictor variable.

Predictor variables	Goodness-of-fit statistics for univariate linear regression models			Accuracy of univariate logistic regression risk models
	R-squared	RMSE	$1 - \frac{RMSE}{RMSE_{max}}$	AUC
PMD measured using scales of varying precision				
1% precision (continuous) ^a	1.000	0.000	1.000	0.6413
5% precision (21 categories)	0.992	1.419	0.913	0.5791
25% precision (4 categories)	0.864	6.009	0.631	0.5500
50% precision (2 categories)	0.557	10.832	0.335	0.5000
Randomly generated values at 1% precision (continuous) ^b	0.0003	16.278	0.000	0.5000

^aTheoretical maximum goodness of fit: R-squared = 1.0, $1 - \frac{RMSE}{RMSE_{max}} = 1.0$ and 100% signal (AUC = 1.0) indicate complete information capture from the predictor variable.

^bTheoretical minimum goodness of fit: R-squared = 0.0, $1 - \frac{RMSE}{RMSE_{max}} = 0.0$ and 100% signal (AUC = 0.5) indicate total information loss from the predictor variable.

Figure 1. Categorizing continuous mammographic density measures leads to increasingly more loss of information as categories become increasingly broad and relates to reduced accuracy of cancer risk models; dichotomizing continuous density measures can lead to loss of a significant amount of information such that risk model performance may be reduced to no better than random guessing (area under curve = 0.5). PMD, percent mammographic density; ROC, receiver operator characteristic.

Univariate logistic regression was used to model the risk of screen-detected breast cancer predicted solely by PMD at varying levels of precision. The sample for these analyses consisted of the subset of females for whom cancer outcome information was known; one female was excluded from this analysis owing to missing data, resulting in 424 cases and 847 controls for these analyses. DeLong's test was used to compare the area under the curve (AUC) for the receiver operator characteristic curve between breast cancer risk models using PMD measures with varying levels of precision, where the reference standard was the model using PMD measured at 1% precision.¹¹

All analyses were performed using SAS® software version 9.3 for Windows (SAS Institute Inc., Cary, NC)

This project was approved by the Capital District Health Authority Research Ethics Board. Informed consent was not obtained from participants as this study involved the secondary use of de-identified data.

RESULTS

The average age of all females included in this study was 59 years (range 40–75 years) and the mean PMD measured at 1% precision was 31% (range 4 to 77%).

R-squared decreased from 1.00 for PMD measured at 1% precision to 0.56 for PMD categorized into high and low. R-squared was 0.00028 when a set of random values at 1% precision ranging from 0% to 100% were used to model the baseline PMD measured at 1% precision. Conversely, the RMSE increased from 0.00 when PMD was measured at 1% precision to 10.83 for PMD categorized as high and low. The maximum RMSE was 16.28 for the model generated using the variable with no signal (Table 1).

When used to evaluate cancer risk using univariate logistic regression models, the AUC curve for the baseline PMD measured at 1% precision (AUC = 0.64) was significantly higher than the AUC for PMD measured at 5% precision (AUC = 0.58, $p < 0.0001$). It was also significantly higher than the AUC for PMD assessed using a 4-category scale at 25% precision (AUC = 0.55, $p < 0.0001$), and for PMD assessed using a 2-category high and low scale at 50% precision (AUC = 0.50, $p < 0.0001$) (Figure 1).

There was a commensurate decrease in AUC with the decrease in precision of measurement of PMD (Table 1). A model developed using the random noise predictor variable had the same AUC (0.50) as the model using PMD assessed either as high or low.

DISCUSSION

Use of less precise density measures achieved by categorizing PMD into density scales leads to significant loss of information and more poorly performing cancer risk models.

Breast cancer risk model performance benefits from the use of continuous measurements of PMD. More coarse or broadly defined breast density scales, such as a 21-point density scale, the Breast Imaging, Reporting and Data System fourth edition density scale or any dichotomization of PMD into high versus low density categories, are less informative for inclusion in image-based breast cancer risk models. The use of a categorical density scale can obscure the more analogue relationship of density to risk by arbitrarily forcing divisions that are not necessarily clinically relevant, such that each category can encompass multiple levels of risk. The more precise continuous mammographic density measures more accurately reflect the escalating dose–response relationship between density and risk, providing greater precision in personalized cancer risk estimates.

While radiologists have demonstrated an ability to assess density using a 1% scale (*i.e.* using a Visual Analogue Scale) and 5% scale (*i.e.* visual assessment on a 21-point scale), inter- and intra-rater variability remain a limitation of these methods.^{12,13} Automated breast density algorithms provide standardized, reproducible PMD assessments on a continuous or near-continuous scale and our results indicate that this would provide more information than categorical density scales in breast cancer risk models. The results of our study show that more precise measures of PMD

enable a more accurate assessment of screen-detected breast cancer risk.

Precise PMD measurements should be used when evaluating a female's personalized risk of having breast cancer in the context of

population-wide breast screening programs that are implementing risk-based stratification within the screen-eligible population.

CONFLICT OF INTEREST

Mohamed Abdolell is CEO and founder of Densitas Inc.

REFERENCES

1. D'Orsi CJ, Bassett LW, Berg WA. BI-RADS: mammography. D'Orsi CJ, Mendelson EB, Ikeda DM. *Breast imaging reporting and data system: ACR BI-RADS—breast imaging atlas*. 4th ed. Reston, VA: American College of Radiology; 2003.
2. Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *J JNCI Nat Cancer Ins* 2006; **98**: 1215–26. doi: <https://doi.org/10.1093/jnci/djj332>
3. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med* 2008; **148**: 337–47. doi: <https://doi.org/10.7326/0003-4819-148-5-200803040-00004>
4. Kerlikowske K, Gard CC, Sprague BL, Tice JA, Miglioretti DL, Kerlikowske K. Breast Cancer Surveillance Consortium. One versus two breast density measures to predict 5- and 10-Year breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2015; **24**: 889–97. doi: <https://doi.org/10.1158/1055-9965.EPI-15-0035>
5. Brentnall AR, Harkness EF, Astley SM, Donnelly LS, Stavrinou P, Sampson S, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res* 2015; **17**: 147. doi: <https://doi.org/10.1186/s13058-015-0653-5>
6. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006; **332**: 1080. doi: <https://doi.org/10.1136/bmj.332.7549.1080>
7. Cohen J. The cost of dichotomization. *Appl Psychol Meas* 1983; **7**: 249–53. doi: <https://doi.org/10.1177/014662168300700301>
8. Austin PC, Brunner LJ. Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. *Stat Med* 2004; **23**: 1159–78. doi: <https://doi.org/10.1002/sim.1687>
9. Bennette C, Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med Res Methodol* 2012; **12**: 21. doi: <https://doi.org/10.1186/1471-2288-12-21>
10. Abdolell M, Tsuruda KM, Lightfoot CB, Payne JJ, Caines JS, Iles SE. Utility of relative and absolute measures of mammographic density vs clinical risk factors in evaluating breast cancer risk at time of screening mammography. *Br J Radiol* 2016; **89**: 20150522: 20150522: . doi: <https://doi.org/10.1259/bjr.20150522>
11. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**: 837–45. doi: <https://doi.org/10.2307/2531595>
12. Gao J, Warren R, Warren-Forward H, Forbes JF. Reproducibility of visual assessment on mammographic density. *Breast Cancer Res Treat* 2008; **108**: 121–7. doi: <https://doi.org/10.1007/s10549-007-9581-0>
13. Sergeant JC, Walshaw L, Wilson M, Seed S, Barr N, Beetles U, et al. Same task, same observers, different values: the problem with visual assessment of breast density. In: Abbey CK, Mello-Thoms CR. ed. SPIE; 2013. pp. 86730T–86730T–8.