

RESEARCH

Open Access



# Evaluation and comparison of different breast cancer prognosis scores based on gene expression data

Avirup Chowdhury<sup>1,2</sup>, Paul D. Pharoah<sup>1</sup> and Oscar M. Rueda<sup>2\*</sup>

## Abstract

**Background** Breast cancer is one of the three most common cancers worldwide and is the most common malignancy in women. Treatment approaches for breast cancer are diverse and varied. Clinicians must balance risks and benefits when deciding treatments, and models have been developed to support this decision-making. Genomic risk scores (GRSs) may offer greater clinical value than standard clinicopathological models, but there is limited evidence as to whether these models perform better than the current clinical standard of care.

**Methods** PREDICT and GRSs were adapted using data from the original papers. Univariable Cox proportional hazards models were produced with breast cancer-specific survival (BCSS) as the outcome. Independent predictors of BCSS were used to build multivariable models with PREDICT. Signatures which provided independent prognostic information in multivariable models were incorporated into the PREDICT algorithm and assessed for calibration, discrimination and reclassification.

**Results** EndoPredict, MammaPrint and Prosigna demonstrated prognostic power independent of PREDICT in multivariable models for ER-positive patients; no score predicted BCSS in ER-negative patients. Incorporating these models into PREDICT had only a modest impact upon calibration (with absolute improvements of 0.2–0.8%), discrimination (with no statistically significant c-index improvements) and reclassification (with 4–10% of patients being reclassified).

**Conclusion** Addition of GRSs to PREDICT had limited impact on model fit or treatment received. This analysis does not support widespread adoption of current GRSs based on our implementations of commercial products.

**Keywords** PREDICT, Breast cancer, Prognosis, Genomic score, Chemotherapy, Calibration, Discrimination, Reclassification

## Introduction

Breast carcinoma is an unregulated growth of cells within the functional units of breast epithelium [1]. It is one of the three most common cancers worldwide (with an estimated 2.26 million new cases in 2020) and the most common malignancy in women [2, 3]. Treatment approaches for breast cancer are diverse and varied: in the UK between 2013 and 2014, 81% of breast cancer patients had surgery as part of their primary treatment regimen, 63% had radiotherapy, 34% had chemotherapy [4], and 62% had endocrine therapy [5].

\*Correspondence:

Oscar M. Rueda

oscar.rueda@mrc-bsu.cam.ac.uk

<sup>1</sup> Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK

<sup>2</sup> MRC Biostatistics Unit, University of Cambridge, East Forvie Building, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Clinicians managing breast cancer must balance the risks of treatment against the potential benefits. Prognostic scores have been developed to assist in these predictions. These include clinical risk scores and genomic risk scores (GRSs). PREDICT is one example of a clinical score. It is based on a multivariable Cox proportional hazards model incorporating patient age, tumour size, tumour grade, tumour protein expression (ER, HER2 and KI67), positive nodes and mode of diagnosis [6, 7]. It provides estimates of absolute treatment benefit for hormone therapy, chemotherapy, adjuvant trastuzumab and bisphosphonate therapy. PREDICT is recommended by the National Institute for Health and Care Excellence (NICE) as a tool for supporting clinical decisions on adjuvant treatment benefit [8] and has been endorsed by the American Joint Committee on Cancer [9]. The underlying model is flexible, enabling additional biomarkers to be incorporated.

GRSs (also called genomic prognostic scores) based on RNA gene expression data were developed in response to the concern that clinicopathological features are imperfect estimators of disease risk and chemosensitivity [10]. They have the theoretical advantages of optimal use of continuous variables and added robustness (e.g. by gathering information on ER activity through a cluster of genes) [11]. Many GRSs have been developed, but few are commercially available and fewer still are endorsed by clinical bodies. At present, only Oncotype DX [12], EPclin [13] and Prosigna [14] have been approved for use in clinical practice in the UK under specific circumstances [15]. Another signature, MammaPrint [16], was deemed clinically effective but not cost-effective.

Several key metrics are used to assess model fit, including calibration, discrimination and reclassification. Calibration is defined as the agreement between observed outcomes and predictions, often presented as an absolute difference in values. Discrimination is a model's ability to differentiate between those with and without the outcome [17], typically expressed using *c*-indices [18]. Reclassification refers to the movement of individuals between risk categories with the introduction of a new prediction model (or extension of a model through the addition of new variables) [17, 19]. Even if the calibration or discrimination of a model does not change, a change in risk categories may result in an individual receiving different treatment according to clinical guidelines [19].

There is a paucity of evidence comparing GRSs to current validated clinical scores. Previous studies comparing GRSs to clinicopathological scores use models less comprehensive than PREDICT and tend to reduce continuous variables into categories [20–28].

This study aims to assess whether GRSs provide any additional clinical benefit beyond PREDICT, the current

standard of care, using a head-to-head comparison in an external cohort. It also analyses the impact of model fit when GRSs are incorporated into the PREDICT algorithm. GRSs included for comparison were those referred to by NICE in their most recent guideline [15]: EPclin, Oncotype DX, Prosigna and MammaPrint.

## Material and methods

Linked clinical and gene expression data were obtained from the METABRIC study [29, 30], described in detail elsewhere. The hazard ratio (HR) functions from PREDICT version 2.1 were used in this analysis [7]. Surrogate KI67 status was calculated using gene expression data for *MKI67*, the gene which codes the KI67 protein, using the *mclust* package [31]. Proportions of KI67 status grouped by cancer grade, stage, number of lymph nodes positive and hormonal status were similar to those previously reported [32].

Four GRSs were adapted in line with the specifications of their respective papers—EndoPredict, Oncotype DX (ODx), Prosigna and MammaPrint. Code was adapted from the *genefu* R package [33] to make it suitable for use on *z*-score normalised expression data. 10-year breast cancer-specific survival (BCSS) was the outcome of interest, defined as the percentage of patients who did not die from breast cancer over ten years.

## Building Cox proportional hazards models

Cox proportional hazards models were built using the *survival* package [34]. The primary outcome of interest was breast cancer-specific death. Separate models were built for ER-positive and ER-negative patients, since the baseline hazard is different in these two groups [7, 12]. In all models, the PREDICT prognostic index was constrained to have a coefficient of one and it was included as an offset. This avoids overfitting and serves as an independent validation of the PREDICT model. No constraints were placed upon GRSs as this information was unavailable, and so this does not serve as an independent validation of these scores as used clinically. Unlike in the original PREDICT model, follow-up was not censored at 15 years.

Univariable models were built for each GRS in turn. Multivariable models were built using PREDICT plus a GRS to assess whether the prognostic information provided by GRSs was independent of PREDICT. Since PREDICT is already a validated multivariable model which incorporates key clinical factors known to be associated with breast cancer prognosis, no additional terms were included in the model to prevent overfitting; for this reason, EndoPredict was used in place of EPclin, and the ROR-C score chosen for Prosigna, in multivariable models. Models were built using single

GRSs, since multiple scores are unlikely to be used simultaneously in a clinical setting due to prohibitive cost.

#### Adding GRS terms into the PREDICT algorithm

We also assessed the impact of including GRSs on the calibration, discrimination and reclassification of the PREDICT algorithm. GRS terms were incorporated as additional terms into PREDICT after rescaling such that the average HR across the GRS distribution was one. This ensures that the baseline hazard used in PREDICT is appropriate.

To account for differences in follow-up time, the expected 10-year survival probability of each patient was calculated using PREDICT. Calibration was reported as the absolute difference in 10-year BCSS between the predicted results (the mean survival of all patients as reported by each algorithm) against the observed results (calculated using the *survival* package [34]). Discrimination was reported for each algorithm in turn by producing a univariable Cox proportional hazards model, and statistical significance tested using the *survcomp* [35] package. Goodness of fit was reported using log-likelihoods and tested using one-way ANOVA tests. Log-likelihoods are equivalent to Akaike information criterion in this case since all models have the same number of variables.

In order to account for the effect of using the same observations to estimate the hazard ratios and to measure model performance, we computed the optimism [36] using a bootstrap procedure adapted from the *rms* R package [36]. We resampled 100 times from the original dataset, fitted the model for each of these samples and compared the performance estimated in the bootstrap sample with a testing sample that contained the observations not sampled in each iteration. The difference between them is the optimism and gives an indication of the amount of overfitting in the model.

The effect of second-generation chemotherapy upon BCSS was used to assess reclassification. Locally, the Cambridge Breast Unit uses PREDICT to stratify patients into three groups according to the predicted benefit of adjuvant chemotherapy: absolute increases in BCSS of <3%, 3–5% and >5% [37]. The first group is usually not offered chemotherapy, and chemotherapy is recommended in the third group; for the middle group, a discussion of the pros and cons of treatment is conducted. These thresholds were used to categorise patients; reclassification was assessed using reclassification tables.

All analyses were conducted in RStudio (version 4.1.0, RStudio, Inc., MA, USA); analysis code and data are provided as Additional file 1.

## Results

### Study population characteristics

Matched clinical outcome and gene expression data were available for 1980 patients in the METABRIC cohort (Table 1). Median follow-up in the study population was 9.56 years (range 0–29.2 years). There were 646 breast cancer-specific deaths during the study period.

### Cox proportional hazards models

Additional file 2: Table S1 summarises key metrics from univariable Cox proportional hazards models. In the ER-positive cohort, all scores except MammaPrint had statistically significant HRs. No GRS had a significant HR in ER-negative patients.

The discrimination of PREDICT (c-index 0.687) was better than GRSs in ER-positive cases; MammaPrint was the best GRS (0.652). GRS discrimination was poor for ER-negative patients with PREDICT performing substantially better (0.667).

In multivariable models, EndoPredict and MammaPrint statistically significantly improved the fit of PREDICT in ER-positive patients (Table 2). While adding Prosigna significantly improved model fit in ER-negative patients, the overall hazard ratio remained non-significant. There was no significant change in discrimination with the addition of GRSs in either ER-positive or ER-negative patients.

### Modified PREDICT algorithm

The PREDICT algorithm was modified to incorporate GRS coefficients from the multivariable models. All modified algorithms underestimated 10-year absolute survival in the METABRIC cohort (Table 3). Survival in the METABRIC cohort at 10 years was 74.0% for ER-positive patients and 58.5% in ER-negative patients. Estimates in ER-positive patients underestimated survival by between 12.2% and 13%, with the closest estimate from PREDICT + MammaPrint. Estimates in ER-negative patients underestimated survival by between 2.3 and 8.8%, with the closest estimate from PREDICT + Prosigna. Subgroup analyses are reported in Additional file 2: Figures S1–S6.

Including GRSs in PREDICT resulted in statistically significant improvements in model fit. Point estimates of discrimination were improved in ER-positive patients with the inclusion of any GRS. However, none of these changes were statistically significant.

The majority of patients remained in the same clinical group when using the original and modified forms of the PREDICT model. A total of 1878 patients were included in these analyses, with the remaining 102 excluded due to missing event data. There were 74 (4%) reclassifications when Oncotype DX was included in PREDICT. This

**Table 1** Study population characteristics

Characteristic	Patients (n = 1980)
Median age at diagnosis, years (range)	61.8 (21.9–96.3)
Inferred menopausal state (%)	
Pre-menopausal	424 (21)
Menopausal	1556 (79)
Oestrogen receptor status (%)	
Positive	1506 (76)
Negative	474 (24)
Progesterone receptor status (%)	
Positive	1040 (53)
Negative	940 (47)
HER2 receptor status (%)	
Positive	247 (12)
Negative	1733 (88)
Stage (%)	
In situ	12 (1)
I	501 (25)
II	825 (42)
III	118 (6)
IV	10 (1)
Not specified	514 (26)
Grade (%)	
1	169 (9)
2	772 (39)
3	955 (48)
Not specified	84 (4)
Median tumour size, mm (range)	23 (1–403)
Lymph node involvement (%)	
Negative	1043 (53)
Positive	937 (47)
Surgery (%)	
Breast-conserving	785 (40)
Mastectomy	1170 (59)
Not specified	25 (1)
Chemotherapy (%)	
None	1568 (79)
Second generation	395 (20)
Third generation	17 (1)
Radiotherapy (%)	
Not received	810 (41)
Received	1170 (59)
Hormone therapy (%)	
Not received	764 (39)
Received	1216 (61)

was lower than those for EndoPredict (132; 7%), MammaPrint (154; 8%) and Prosigna (183; 10%).

The most important clinical category of ER-positive patients to consider is intermediate benefit, since the benefit of adjuvant chemotherapy is unclear in this

group. Reclassification varied by GRS (Table 4). Similar numbers of patients were reclassified into and out of the intermediate benefit category with Oncotype DX (40 vs. 34) and EndoPredict (71 vs. 61). More patients were reclassified out of intermediate benefit than into it using MammaPrint (66 vs. 88), while more were classified as intermediate when using Prosigna (102 vs. 80).

No patients were reclassified from high to low benefit or vice versa when Oncotype DX, EndoPredict or MammaPrint were used; with Prosigna, 1 patient was reclassified from low to high benefit.

## Discussion

Overall, EndoPredict, MammaPrint and Prosigna demonstrated prognostic power independent of PREDICT in multivariable models in ER-positive patients; however, discrimination was not significantly improved. Incorporating GRSs into the PREDICT algorithm did not improve calibration, with underestimation of 10-year BCSS in the METABRIC cohort in ER-positive and ER-negative cohorts. Measures of discrimination were not significantly changed. GRS inclusion caused 4–10% of patients to be reclassified into different clinical categories.

This analysis addresses some key gaps in the current evidence base. In their most recent guideline on the topic, NICE [15] states that “there are no data available to compare the tumour profiling tests with PREDICT, or to define the clinical risk groups using PREDICT”. Previous studies tended to compare GRS against one another or against suboptimal clinicopathological parameters. This study used the current clinical standard of care, making it easier to assess whether clinical predictions are improved.

By creating modified versions of the PREDICT algorithm which accepted GRSs as an additional term, this analysis leveraged absolute risk predictions from PREDICT to allow inferences about the impact of GRSs on model fit. Although this is not the same as calculating the calibration of the scores themselves, it nonetheless demonstrates what the impact of combining PREDICT and GRSs would be.

PREDICT had poor calibration in this analysis, underestimating 10-year BCSS by 13%. Independent validation of PREDICT by Gray et al. [38] on the Scottish Cancer Registry showed much higher calibration, with 5% overestimation of 5-year mortality and 2% underestimation of 10-year mortality. The differences in findings between this analysis and previous work may be due to differences in the outcome of interest (overall survival versus BCSS) or the cohorts themselves (as outcomes in the general population may be different to those in a highly selected cohort like METABRIC, from the UK and Canada). It

**Table 2** Multivariable Cox proportional hazards models for each genomic prognostic signature adjusted for PREDICT in (a) ER-positive and (b) ER-negative patients

Model	Hazard ratio (95% CI)	Log-likelihood	Model fit <i>p</i> value	c-index (95%CI)	c-index optimism
(a)					
PREDICT	2.72 <sup>†</sup> (NA)	− 2860.6	−	0.687 (0.661–0.713)	−
Oncotype DX	1 (0.99–1)	− 2859.1	0.0835	0.693 (0.667–0.718)	− 4.17 × 10 <sup>−4</sup>
EndoPredict	1.13 (1.02–1.24)	− 2858	0.024	0.692 (0.666–0.717)	0.003
MammaPrint	0.52 (0.17– 0.87)	− 2853.9	2.32 × 10 <sup>−4</sup>	0.699 (0.674–0.724)	6.72 × 10 <sup>−4</sup>
Prosigna	1.58 (1.11– 2.05)	− 2858.7	0.0546	0.694 (0.668–0.719)	− 8.68 × 10 <sup>−4</sup>
(b)					
PREDICT	2.72 <sup>†</sup> (NA)	− 1072	−	0.667 (0.630–0.704)	−
Oncotype DX	0.998 (0.986– 1.01)	− 1072	0.77	0.667 (0.63–0.704)	0.011
EndoPredict	0.884 (0.657– 1.11)	− 1071.4	0.287	0.667 (0.630–0.705)	0.007
MammaPrint	0.854 (0.196– 1.51)	− 1071.9	0.638	0.669 (0.632–0.706)	0.005
Prosigna	0.252 (− 0.875–1.38)	− 1069.3	0.0166	0.661 (0.622–0.699)	7.84 × 10 <sup>−4</sup>

95% CI, 95% confidence interval

<sup>†</sup> Hazard ratio has been constrained to this value**Table 3** Calibration of original and modified PREDICT models in the METABRIC cohort for (a) ER-positive and (b) ER-negative patients

Model	10-Year survival %	Difference	Difference optimism
(a)			
<b>METABRIC cohort</b>	<b>74.0</b>	Reference	Reference
PREDICT	61.1	− 13	−
PREDICT + Oncotype DX	61.3	− 12.8	− 0.694
PREDICT + Endo-Predict	61.6	− 12.5	− 0.711
PREDICT + MammaPrint	61.9	− 12.2	− 0.640
PREDICT + Prosigna	61.2	− 12.8	− 0.706
(b)			
<b>METABRIC cohort</b>	<b>58.5</b>	Reference	Reference
PREDICT	49.7	− 8.8	−
PREDICT + Oncotype DX	50.5	− 8	− 0.307
PREDICT + Endo-Predict	52.9	− 5.6	− 0.403
PREDICT + MammaPrint	48.3	− 10.2	− 0.150
PREDICT + Prosigna	56.2	− 2.3	− 0.336

In bold, reference survival (METABRIC cohort)

may also be due to calibration drift, where risk estimates change over time due to changes in population characteristics or disease incidence [39]. This may also explain the lower underestimates of calibration in ER-negative

disease (Additional file 2: Figure S1), for whom treatment advances have been more limited.

There is a need to develop prognostic molecular scores which take into account the standard clinicopathological variables already used in a clinical setting and then explore how much additional benefit is gained from the inclusion of genomic signatures. In this way, the developed scores will reflect real-world clinical practice and be more relevant to clinical decision-making.

#### Limitations and future work

A key limitation of this work is that surrogate GRSs were used rather than their commercial counterparts due to cost considerations. Although all GRSs in this study were derived from published papers in line with the original authors' instructions, there is nonetheless a risk that these differ from commercial scores. Future work should use commercial scores where possible.

Due to the time period in which METABRIC patients were recruited, certain treatments (e.g. bisphosphonates) were less commonly used. There is also a risk of time-varying confounding due to available treatment regimens changing over time (and affecting patient survival). This is likely given that breast cancer survival rates have improved dramatically over the past decades, a finding attributed in major part to improved treatment [1, 2].

The use of pre-processed METABRIC data presented some challenges for conversion of GRSs. Changing the type of data normalisation post hoc is challenging and risks introducing further biases [40]. These issues continue to exist even if raw data are analysed; the only way to eliminate them is to use the genomic test with the

**Table 4** Reclassification tables comparing clinical categories of chemotherapy benefit from standalone PREDICT and PREDICT with (a) Oncotype DX, (b) EndoPredict, (c) MammaPrint and (d) Prosigna in ER-positive patients

		PREDICT + Oncotype DX		
		Low	Intermediate	High
(a)				
PREDICT				
Low		893	19	0
Intermediate		27	434	7
High		0	21	477
		PREDICT + EndoPredict		
		Low	Intermediate	High
(b)				
PREDICT				
Low		890	22	0
Intermediate		46	407	15
High		0	49	449
		PREDICT + MammaPrint		
		Low	Intermediate	High
(c)				
PREDICT				
Low		873	39	0
Intermediate		54	380	34
High		0	27	471
		PREDICT + Prosigna		
		Low	Intermediate	High
(d)				
PREDICT				
Low		877	34	1
Intermediate		60	388	20
High		0	68	430

official platform. This was not feasible due to the high cost of requesting such testing.

Only BCSS was considered in this analysis. Clinical outcome measures used in cancer are diverse and include measures of survival, disease recurrence and disease-free survival. Although these outcomes are correlated to some extent, future studies would benefit from considering multiple clinical outcomes to ensure that these findings are consistent and sustained across clinically relevant subgroups.

This study likely underestimates PREDICT's performance, since several variables required for optimal prediction (screening status, KI67 status and bisphosphonate use) were unavailable. Although alternative methods were used to infer KI67 status, these are unlikely to be as accurate as established histopathological techniques.

Similarly, this analysis likely overestimates GRS prognostic power, since GRS coefficients were unconstrained, while PREDICT was constrained to one. This has the effect of allowing GRS coefficients to be re-estimated in the current dataset, effectively creating an overfitting problem whereby the predictive power of included variables is overestimated. The impact of this may be quite large: when univariable PREDICT models were built using the unconstrained variable, model fit was dramatically improved compared to the constrained model (log-likelihood -2823.6 versus -2860.2).

Although this study found that there were largely non-significant changes in model performance as a result of incorporation of genomic prognostic scores, these changes need to be modelled economically. In particular, the small number of reclassifications may be important at a health system level if they result in

improvements in patient care (which reduce long-term costs of readmission, for example).

## Conclusion

This study evaluates the impact of adding GRSs to current standards of care for breast cancer predictive modelling using key model performance metrics (calibration, discrimination and reclassification). Three GRSs (EndoPredict, MammaPrint and Prosigna) demonstrated power to predict BCSS in breast cancer independent of PREDICT. However, incorporating these models into PREDICT had only a modest impact upon calibration (underestimating 10-year BCSS by around 12%), discrimination (with *c*-indices non-significantly different to the original PREDICT algorithm) and reclassification (with 4–10% of patients reclassified into different clinical categories). Performance was much better in ER-positive than ER-negative patients. Although these small improvements in model fit might be clinically useful, economic analyses are needed to assess whether this justifies the increased cost.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13058-023-01612-9>.

**Additional file 1.** Code and data used in the study.

**Additional file 2.** Supplementary Tables and Figures.

## Acknowledgements

Not applicable.

## Author contributions

PDP and OMR conceptualised and designed the study; AC undertook the formal analysis and wrote the manuscript under the supervision of PDP and OMR. All authors read and approved the final manuscript.

## Funding

OMR was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and the UKRI (UK; MC\_UU\_00002/16).

## Availability of data and materials

All data generated or analysed during this study are included in this published article (and its supplementary information files).

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

PDP receives a share of fees received by Cambridge Enterprises for licensing of PREDICT. AC and OMR have no conflicts of interest to declare.

Received: 25 June 2022 Accepted: 27 January 2023

Published online: 08 February 2023

## References

- Winters S, Martin C, Murphy D, Shokar NK. Breast cancer epidemiology, prevention, and screening. *Prog Mol Biol Transl Sci*. 2017;151:1–32.
- Harbeck N, Gnant M. Breast cancer. *Lancet*. 2017;389:1134–50.
- Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: an overview. *Int J Cancer*. 2021;149:778–89.
- Service NCR& A, UK CR. Chemotherapy, radiotherapy and tumour resections in England: 2013–2014 workbook. 2017.
- Bastiaannet E, Charman J, Johannesen TB, Schrodi S, Siesling S, van Eycken L, et al. A European, observational study of endocrine therapy administration in patients with an initial diagnosis of hormone receptor-positive advanced breast cancer. *Clin Breast Cancer*. 2018;18:e613–9.
- Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearns O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res*. 2010;12:R1.
- Candido dos Reis FJ, Wishart GC, Dicks EM, Greenberg D, Rashbass J, Schmidt MK, et al. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. *Breast Cancer Res*. 2017;19:58.
- NICE. Early and locally advanced breast cancer: diagnosis and management (NG101). NICE; 2020.
- American Joint Committee on Cancer. AJCC cancer staging manual. Berlin: Springer; 2017.
- Waks AG, Winer EP. Breast cancer treatment: a review. *JAMA*. 2019;321:288–300.
- Gyorffy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L. Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res*. 2015;17:1–7.
- Chia SKL. Clinical application and utility of genomic assays in early-stage breast cancer: key lessons learned to date. *Curr Oncol*. 2018;25:S125–30.
- Filipits M, Rudas M, Jakesz R, Dubsy P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res*. 2011;17:6012–20.
- Parker JS, Mullins M, Cheung MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27:1160–7.
- NICE. Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer [DG34]. NICE; 2018.
- Van't Veer LJ, Dai H, Van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530–6.
- Steyerberg EW, Vickers AJ, Cook NR, Gerdts T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128–38.
- Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247:2543–6.
- Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115:928–35.
- Beumer IJ, Persoon M, Witteveen A, Dreezen C, Chin SF, Sammut SJ, et al. Prognostic value of MammaPrint in invasive lobular breast cancer. *Biomark Insights*. 2016;11:139–46.
- Buus R, Sestak I, Kronenwett R, Denkert C, Dubsy P, Krappmann K, et al. Comparison of EndoPredict and EPclin With Oncotype DX recurrence score for prediction of risk of distant recurrence after endocrine therapy. *J Natl Cancer Inst*. 2016;108:djw149.
- Drukker CA, Elias SG, Nijenhuis MV, Wesseling J, Bartelink H, Elkhuisen P, et al. Gene expression profiling to predict the risk of locoregional recurrence in breast cancer: a pooled analysis. *Breast Cancer Res Treat*. 2014;148:599–613.
- Martin M, Brase JC, Calvo L, Krappmann K, Ruiz-Borrego M, Fisch K, et al. Clinical validation of the EndoPredict test in node-positive, chemotherapy-treated ER+/HER2- breast cancer patients: results from the GEICAM 9906 trial. *Breast Cancer Res*. 2014;16:R38.
- Yao K, Goldschmidt R, Turk M, Wesseling J, Stork-Sloots L, de Snoo F, et al. Molecular subtyping improves diagnostic stratification of patients with primary breast cancer into prognostically defined risk groups. *Breast Cancer Res Treat*. 2015;154:81–8.
- Zhang Y, Schnabel CA, Schroeder BE, Jerevall P-L, Jankowitz RC, Fornander T, et al. Breast cancer index identifies early-stage estrogen

- receptor-positive breast cancer patients at risk for early- and late-distant recurrence. *Clin Cancer Res.* 2013;19:4196–205.
26. Zhao X, Rodland EA, Sorlie T, Vollan HKM, Russnes HG, Kristensen VN, et al. Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status. *BMC Cancer.* 2014;14:211.
  27. Filipits M, Nielsen TO, Rudas M, Greil R, Stoger H, Jakesz R, et al. The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. *Clin Cancer Res.* 2014;20:1298–305.
  28. Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z, et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Ann Oncol.* 2014;25:339–45.
  29. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012;486:346–52.
  30. Rueda OM, Sammut S-J, Seoane JA, Chin S-F, Caswell-Jin JL, Callari M, et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature.* 2019;567:399–404.
  31. Scrucca L, Fop M, Murphy TB, Raftery AE. Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J.* 2016;8:289–317.
  32. Abubakar M, Orr N, Daley F, Coulson P, Ali HR, Blows F, et al. Prognostic value of automated Ki67 scoring in breast cancer: a centralised evaluation of 8088 patients from 10 study groups. *Breast Cancer Res.* 2016;18:1–13.
  33. Gendoo DMA, Ratanasirigulchai N, Schröder MS, Paré L, Parker JS, Prat A, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics.* 2016;32:1097–9.
  34. Therneau TM. A package for survival analysis in R [Internet]. *Compr. R Arch. Netw. Comprehensive R Archive Network (CRAN)*; 2021. Cited 28 June 2021. <https://cran.r-project.org/package=survival>.
  35. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics.* 2011;27:3206–8.
  36. Harrell FEJ. *Regression modeling strategies.* Berlin: Springer; 2001.
  37. Loh S-W, Rodriguez-Miguel M, Pharoah P, Wishart G. A comparison of chemotherapy recommendations using Predict and Adjuvant models. *Eur J Surg Oncol.* 2011;37:S21–2.
  38. Gray E, Marti J, Brewster DH, Wyatt JC, Hall PS. Independent validation of the PREDICT breast cancer prognosis prediction tool in 45,789 patients using Scottish Cancer Registry data. *Br J Cancer.* 2018;119:808–14.
  39. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med.* 2019;17:1–7.
  40. Qin S, Kim J, Arafat D, Gibson G. Effect of normalization on statistical and biological interpretation of gene expression profiles. *Front Genet.* 2013;4:160.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

