Breast cancer biomarkers in clinical testing: analysis of a UK national external quality assessment scheme for immunocytochemistry and in situ hybridisation database containing results from 199 300 patients

Andrew Dodson[1,2]* ⓘ, Suzanne Parry[3], Merdol Ibrahim[3], John MS Bartlett[4], Sarah Pinder[5], Mitch Dowsett[1,2] and Keith Miller[3]

[1]*Ralph Lauren Centre for Breast Cancer Research, The Royal Marsden Hospital, London, UK*
[2]*The Institute of Cancer Research, London, UK*
[3]*UK NEQAS ICC & ISH, University College London Cancer Institute, London, UK*
[4]*Ontario Institute for Cancer Research, Ontario, Canada*
[5]*Cancer Studies, King's College London, London, UK*

*Correspondence to: Andrew Dodson, Ralph Lauren Centre for Breast Cancer Research, Wallace Wing, The Royal Marsden Hospital, Fulham Road, London SW3 6JJ, UK. E-mail: andrew.dodson@icr.ac.uk

Abstract

We describe a collated data set of results from clinical testing of breast cancers carried out between 2009 and 2016 in the United Kingdom and Republic of Ireland. More than 199 000 patient biomarker data sets, together with clinicopathological parameters were collected. Our analyses focused on human epidermal growth factor receptor-2 (HER2), oestrogen receptor (ER) and progesterone receptor (PR), with the aim of the study being to provide robust confirmatory evidence on known associations in these biomarkers and to uncover new data on previously undescribed or unconfirmed associations, thus strengthening the evidence-base in clinical breast cancer testing. Overall, 13.1% of tumours were HER2-positive; 10.6% in ER-positive tumours, and 25.5% in ER-negative tumours. Higher rates of HER2 positivity were significantly associated with patient age <56 years versus age ≥56 years, symptomatic versus screen-detected tumours, testing of involved axillary node versus primary breast cancer, invasive ductal carcinoma (not otherwise specified) versus other histological types, higher histological grade, increasing tumour size, increasing nodal involvement, ER-negative versus ER-positive tumour status, PR-negative versus PR-positive tumour status. Where ER status was known, 82.7% of tumours were ER-positive; 80.9% in women age <56 years, and 83.6% in those age ≥56 years (ER-positive cut-off ≥1.0% positive tumour cells or equivalent). Where PR status was known, 64.9% of tumours were PR-positive; 65.8% in women age <56 years, and 64.4% in women age ≥56 years (PR-positive cut off ≥10.0% or equivalent). These analyses of clinical test results provide contemporary benchmarking data for HER2, ER and PR positive rates.

Introduction

The United Kingdom National External Quality Assessment Scheme for Immunocytochemistry and In Situ Hybridisation (UK NEQAS ICC & ISH) provides services to healthcare laboratories with the aim of helping them to monitor and, where necessary, improve the quality of their clinical testing [1]. It was established in the UK in the early 1980s and has grown into an international service. The scheme distributes materials to participating laboratories, which test for the requested analyte and return stained slides for central assessment. Both individual and collective performance results are fed back at the completion of each assessment run.

In 2007 the scheme extended the scope of its activities by setting up a central spreadsheet-based data

repository into which participants sited in the UK and Republic of Ireland (RoI), could enter data on their human epidermal growth factor receptor-2 (HER2) testing activity in breast cancer. In 2008 the data set was further developed to include two other important breast biomarkers, oestrogen receptor (ER) and progesterone receptor (PR) [2]. The central collection of data was moved to a web-based application in 2009 and continued in this format until early 2016, when the database was closed; more than 199 000 entries were made in to the web-based version. We report here descriptive statistics for the collated data and analyses examining associations for each of the three biomarkers, with the primary aim of producing robust benchmarking data for audit and quality assurance of clinical testing. The data should also provide a relatively close approximation to prevalences in invasive primary breast cancer during the period of collection.

Materials and methods

In 2009 a web-based database system (QuickBase Inc., Cambridge, MA, USA) was made available by UK NEQAS to scheme participants located in the UK and RoI for the entry and collation of HER2, ER, and PR and other associated data. The application was password restricted, allowing centres to access only their own data. The database ran in this format until it was closed in March, 2016.

The majority of data fields in the final version of the web-based application were designed to capture patient demographics, tumour characteristics, and biomarker (HER2, ER, PR) related information. Completion of the 'Originating Trust', 'Date of Test' and 'Unique Record Identifier' fields within each record was mandatory (here 'Trust' designates a hospital or healthcare provider, it is a term used soley in England and Wales, and the more generic term 'centre' has been used in the article). Completion of a categorical 'HER2 Status' field (negative/positive) was also mandatory. Data entry into all other fields was optional.

Additional fields were present allowing entry of technical reagent-specific information and test turnaround times and there were also fields relating to assays for markers such as E-cadherin protein and the proliferation marker Ki67. Comparatively few entries were made in these supplementary areas and they will not be further discussed.

Data were examined for consistency, as records were generated by multiple users into record fields that accepted for the most-part free-text entries. Based on

grouping by the 'Originating Trust' field, data from centres entering <200 records in total were excluded from the analysis. This was done on the basis that data from such 'casual' users were less likely to be entered consistently or be representative of the broad patient population at their centres.

Patient age at date of testing was calculated as the difference between date of birth and date of test. Because menopausal status was not captured, a cut-point of patient age at test ≥56 years has been used to define post-menopausal status in models involving this patient characteristic.

Centres returned ER and PR results as a categorical status (positive/negative/borderline), Allred score, H-Score, or as a proportion of cells staining positive (% positive). Cut-points for defining ER-positive disease were: Allred ≥3, H-Score ≥1, % positive ≥1.0%. For PR-positive disease two cut-point definitions were used, reflecting non-standardised practice: Allred ≥3, H-Score ≥1, % positive ≥1.0% and Allred ≥4, H-Score ≥10, % positive ≥10%.

When combining data to derive a final hormone receptor status for any given tumour, precedence was based on the prevalence of scoring method within the database i.e. categorical status (excepting borderline), Allred score, % positive, and H-Score.

Patient identifiable data were not used in the preparation of this article.

The Excel spreadsheet application (Microsoft Corp., version 15, 2017; Redmond, WA, USA) was used for data storage and handling. Prism software (Graphpad Software Corp., version 7, 2017; San Diego, CA, USA) was used for statistical analyses.

Results

Database scope and patient characteristics

When it closed in March 2016 the database contained 199 300 individual patient records, representing approximately 55% of all invasive breast cancer cases tested during the 7+ year period when the database was in existence (using 50 000 per year as an approximation for the total number of cases eligible for testing in the UK and RoI). After exclusion of records from centres that entered <200 records, 187 368 (94.0%) were available for analysis; of these 182 413 (97.4%) were collected between 2009 and 2015. The total number of records identifiable as originating from UK-based centres was 144 733 (77.2%), those from RoI centres totalled 15 369 (8.2%) and for 27 266 (14.6%) it was not possible to identify from which of the two regions they originated. Data and analyses relating to

the UK-specific and whole patient population are presented here, results relating to the RoI patient population are given in the Supplementary RoI file.

The number of entries into the separate subject fields varied greatly. However, due to the overall size of the database, fields where coverage was far from complete often included information on substantial patient numbers.

After exclusion of centres entering <200 records the median number of records entered per centre was 862 (range: 206–7134). Figure 1A shows further data for the number of tests entered by year and Supplementary material, Table S1 shows a breakdown of numbers of entries by centre.

Distributions for patient age were very similar in the UK and the whole population (for both median age = 61 years; inter-quartile range [IQR] = 51–71 years); in each population 64% of women were age ≥56 years. Figure 1B illustrates the age data. Further patient demographics together with figures for tumour characteristics are shown in Table 1 and a detailed breakdown of tumour size distribution is given in Supplementary material, Figure S1.

HER2-positive rates

Based on the categorical HER2 status field data entries 13.3% of UK patients and 13.1% of the total patient population were HER2-positive (Supplementary material, Table S2). When the data were examined by year of testing a noticeable but non-significant trend was seen for a decline in positivity-rates by year for both UK and overall data ($p > 0.05$ for both). This is illustrated in Figure 2A, and Supplementary material, Table S3 gives further details.

Comparative data for HER2-positive rates grouped by patient demographics and tumour characteristics are given in Table 2; HER2-positive rates were calculated using centre-assigned HER2 status for these analyses. In the majority of cases, the data are similar for both the UK and the whole population, and for brevity only the findings for the whole population are discussed here (Table 2 gives further details). HER2-positive rates were significantly higher in tumours from patients:

- who were age <56 years versus age ≥56 years; 16.1 and 11.4%, respectively ($p < 0.001$). The effects on HER2-positive rates of patient age and hormonal receptor status are discussed in more detail below;
- who presented with symptomatic disease compared to those whose tumours were screen-detected; 13.3 and 9.0%, respectively ($p < 0.001$).
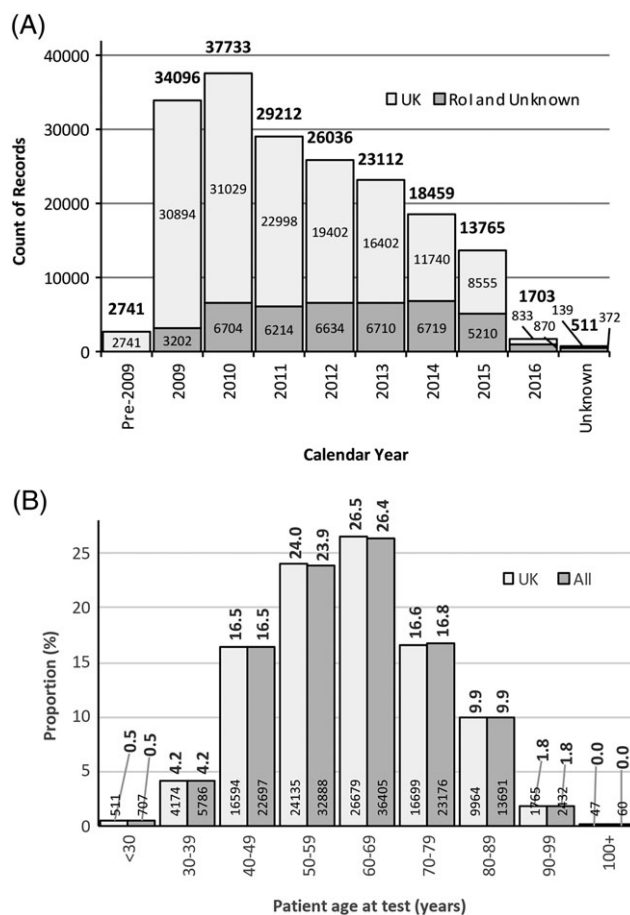

Figure 1. (A) Total number of records submitted in each calendar year. Those shown as 'Unknown' represent records entered with no date or with an unfeasible date. Numbers above columns are overall totals, those within columns indicate total number of patients submitted by UK centres, records submitted by RoI centres have been grouped with those where the location of the centre was unknown. (B) Distributions of patient age at date of HER2 testing. Median age in the UK and the total patient populations was 61 years (IQR: 51–71 years). Numbers above columns are proportions (%), those within column base are counts of patients.

The HER2-positive rates in diagnostic cores (13.0%) and surgical excisions (13.1%) were not significantly different but axillary/nodal samples showed a significantly higher rate of positivity compared to primary breast cancer specimens (15.7%; $p < 0.001$).

The rate in bone metastases was significantly lower than in cutaneous and visceral disease samples (9.2, 15.0, 15.3%, respectively; both $p < 0.05$); but the rates in the latter two were not significantly different from each other. It is possible that this lower positivity rate seen in bone metastases compared to soft tissues may

Table 1. Patient and tumour characteristics

	UK	All
Patient records, n (%)	144 733 (77.2)	187 368 (100.0)
Patient age at testing (years), median (IQR)	61 (51–71)	61 (51–71)
Age (menopausal status surrogate), n (%)		
<56 years	36 171 (36.0)	49 561 (36.0)
≥56 years	64 397 (64.0)	88 281 (64.0)
Total records*	100 568 (69.5*)	137 842 (73.6*)
Referral pathway, n (%)		
Screen-detected	2348 (38.2)	4620 (35.4)
Symptomatic	3792 (61.8)	8414 (64.6)
Total records*	6140 (4.2*)	13 034 (7.0*)
Primary specimen sample type, n (%)		
Axilla/nodes	5618 (5.7)	6428 (4.7)
Core biopsy	68 333 (69.1)	96 269 (71.0)
Excision	24 876 (25.2)	32 931 (24.3)
Total records*	98 827 (68.3*)	135 628 (72.4*)
Metastatic presentation site, n (%)		
Bone	221 (9.0)	445 (11.4)
Cutaneous	1072 (43.5)	1483 (37.9)
Visceral	1173 (47.6)	1984 (50.7)
Total records*	2466 (1.7*)	3912 (2.1*)
Histological type, n (%)		
Carcinoma *in situ*	2346 (3.6)	3753 (4.5)
IDC/NOS	51 780 (80.1)	66 432 (79.4)
ILC	7922 (12.3)	10 416 (12.4)
Mixed IDC/ILC	1236 (1.9)	1461 (1.7)
Special type	1337 (2.1)	1652 (2.0)
Total records*	64 621 (44.6*)	83 714 (44.7*)
Histological grade, n (%)		
Grade 1	10 792 (16.3)	13 032 (16.1)
Grade 2	35 147 (53.0)	43 519 (53.7)
Grade 3	20 351 (30.7)	24 424 (30.2)
Total records*	66 290 (45.8*)	80 975 (43.2*)
Histological tumour size, n (%)		
Diameter < 20 mm	7647 (53.2)	10 075 (51.5)
Diameter 20 to <50 mm	5487 (38.2)	7596 (38.9)
Diameter ≥ 50 mm	1234 (8.6)	1876 (9.6)
Total records*	14 368 (9.9*)	19 547 (10.4*)
Nodal status, n (%)		
N0	6913 (62.4)	9138 (62.4)
N1	2691 (24.3)	3613 (24.7)
N2	825 (7.4)	1054 (7.2)
N3	649 (5.9)	843 (5.8)
Total records*	11 078 (7.7*)	14 648 (7.8*)

Summary data for count of patient records and for patient age at testing, together with data on other patient and tumour characteristics. Data are presented for the UK and the whole patient population
n = number in group, % = proportion within group.
*'Total records' the proportion is that of whole population, and thus indicates coverage.

be at least in part attributeable to technical effects of decalcification rather than a true biological difference.

The following tumour characteristics were indicative of significantly higher HER2-positive rates (all at $p < 0.001$):

- invasive carcinoma of ductal/not otherwise specified (IDC/NOS) type compared to carcinoma *in situ* (15.3 versus 8.0%);

- IDC/NOS compared to invasive lobular carcinoma (ILC) (15.3 versus 5.5%);
- IDC/NOS compared to mixed IDC/ILC (15.3 versus 7.9%);
- IDC/NOS compared to invasive disease of special type (15.3 versus 5.3%);
- increasing histological grade (Grade 1: 2.5%, Grade 2: 10.8%, Grade 3: 23.8%);
- increasing categorical tumour size (<20 mm: 11.8, 20 to <50 mm: 13.9%, ≥50 mm:14.6%);
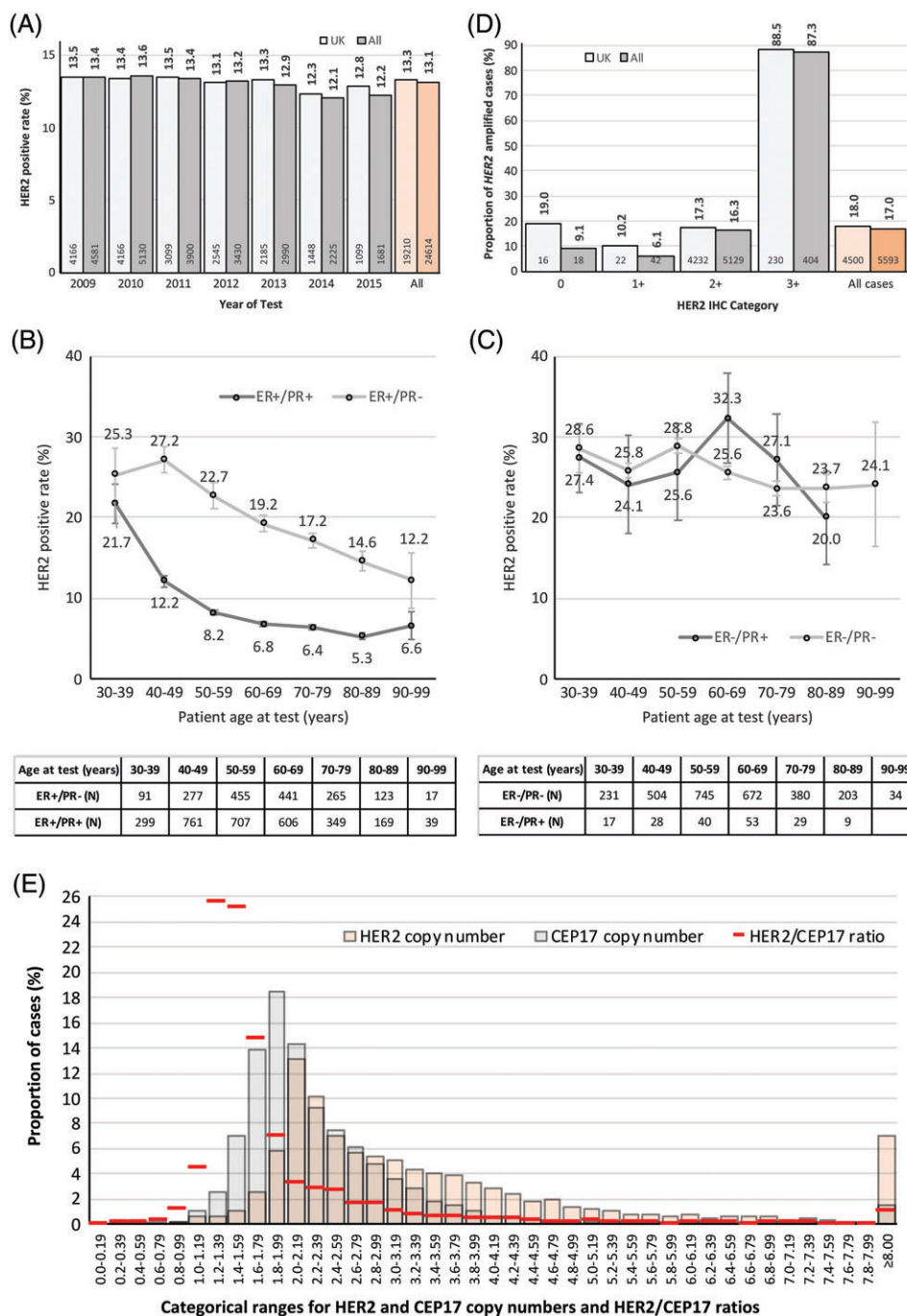
Figure 2. Legend on next page.

- increasing nodal involvement (N0: 11.0%, N1: 13.3%, N2: 16.3%, N3: 21.8%);
- ER-negative versus ER-positive tumours (25.5 versus 10.6%). Discussed in greater detail below;
- PR-negative versus PR-positive tumours (23.4 versus 8.7%). Discussed in greater detail below.

When tumours were categorised according to ER status, a marked difference in the frequency of HER2-positive cases was observed in ER-positive versus ER-negative breast cancers with respect to nodal status and grade. Our data confirm a strong relationship between increasing frequency of HER2-positive

cases and grade in ER-positive but not in ER-negative disease as has been previously reported [3]. However, the reciprocal relationship between nodal status, ER status and HER2 status was not observed (Supplementary material, Tables 4A and 4B).

Table 3 shows the numbers and proportions of cases classified into each HER2 category sub-according to ER and PR status. Comparing ER-positive and ER-negative tumours:

- a substantially larger proportion of ER-negative tumours were HER2 3+ (25.4 versus 7.9%);
- a larger proportion of ER-positive tumours were HER2 2+ (13.7 versus 9.6%);
- within the 2+ category, a larger proportion of ER-negative tumours were *HER2* gene amplified (25.0 versus 15.9%);
- in ER-positive tumours PR status affected the proportion of tumours classified as either HER2 0 or 1+ (PR-negative: 54.5%, PR-positive: 67.3%) and the proportion classified as HER2 3+ (PR-negative: 16.6%, PR-positive: 5.7%);
- in ER-negative tumours PR status did not substantially affect HER2 categorisation, but it should be remembered that the number of ER-negative/PR-positive cases was very small ($n = 693$, 1.1% of total population) and confidence in this result is less than for others reported here.

Figure 2B,C illustrates the relationships between HER2-positive rates and patient age in 10-year groups; the rate was broadly similar in patients presenting at a young age (30–39 years) with either ER-positive or ER-negative tumours, regardless of PR status (range: 28.6–21.7%). In ER-positive/PR-positive tumours, the HER2-positivity rate declined rapidly with increasing patient age, up to age 60 years after which it plateaued at approximately 6.5%. In contrast, in ER-positive/PR-negative tumours the rate declined much more slowly, and substantial separation of the curves is seen in patients aged 40 years and above; PR-negative tumours in patients aged between 90 and 99 years had a HER2-positive rate that remained approximately twice that seen in PR-positive tumours of patients matched for age (12.2 versus 6.6%). These differences in rates were significant ($p < 0.05$). In contrast, in ER-negative patients there was no significant change in the rates of HER2-positivity with increasing age, and PR status had no observable effect.

HER2 amplification status

Figure 2D shows the proportions of amplified cases for each HER2 immunohistochemistry (IHC) category (Supplementary material, Tables S5B,C and S6A–J give detailed results). It is important to remember when interpreting this data that it reflects, for the most part, *HER2* gene amplification rates for those cases that were reflexed to *in situ* hybridisation (ISH) testing following IHC, and as such is heavily weighted towards HER2 2+ cases, with comparatively few cases in other IHC categories. Using data from centre-assigned HER2 IHC category (0, 1+, 2+, 3+), combined with centre-assigned *HER2* ISH status (non-amplified/amplified), 16.3% of all HER2 2+ tumours

FIGURE 2 (A) HER2-positive rates for records grouped by year of test. Figures above columns indicate positivity-rates, those at bases of columns show the total number of HER2-positive patients in that group. The data shown are for centre-assigned HER2 status, and exclude tests conducted pre-2009 and in 2016 where insufficient numbers of tests were present to allow reliable analysis. In the indicated date-range the mean HER2-positive rate for UK patients was 13.2% [IQR: 12.8–13.5%; standard deviation (SD) = 0.44%], and for all patients it was 13.2% (IQR: 12.2–13.4%; SD = 0.60%). The mean HER2-positive rate for all UK patients (light-orange column) was 13.3%, and for all patients (orange column) it was 13.1%. (B) Mean HER2-positive rates for patients with ages between 30 and 99 years in 10-year groupings. This figure is for ER-positive patients stratified by PR status and should be compared with figure (C), which shows similar data for ER-negative patients. This is an analysis of the whole population to maintain adequate patient numbers in some subgroups. The table indicates number of HER2-positive patients in group. Error bars indicate standard error of the mean. (C) Mean HER2-positive rates for patients with ages between 30 and 99 years in 10-year groupings (90–99 group omitted for ER-negative/PR-positive patients due insufficient numbers in group). This figure, which is for ER-negative patients, should be compared with figure (B), which shows ER-positive patients. This is an analysis of the whole population to maintain adequate patient numbers in some subgroups. The table indicates number of HER2-positive patients in group. Error bars indicate standard error of the mean, ER-negative/PR-positive status is uncommonly encountered and hence the patient population is comparatively small, leading to large standard errors. (D) *HER2* gene amplification rates by HER2 IHC category. The distribution of cases is very heavily weighted towards HER2 2+, which comprise slightly less than 96% of cases (98% for UK data) in the analysis. The figures above columns indicate proportions of amplified cases, those within the bases of columns show number of amplified cases. A small number of centres also used non-standard '1+/2+' and '2+/3+' categories; these are not recognised in any published HER2 assessment guidelines and have been excluded (UK cases, $n = 1837$); All cases, $n = 2000$. See also Supplementary material, Table 5B,C. (E) Distribution of *HER2* gene and CEP17 copy numbers, and *HER2*/CEP17 ratio for cases where these data were reported ($n = 12\,049$). Median number of *HER2* copies per cell was 2.73 (IQR: 1.55–11.75); for CEP17 it was 1.90 (IQR: 1.24–3.80). Median *HER2*/CEP17 ratio was 1.33 (IQR: 0.96–4.36).

Table 2. HER2-positive rates for tumours from UK patients and those from the whole patient population, using the recording centre's categorical assignment to define HER2 status

	UK	P value	All	P value
All cases, n (%)	19 210 (13.3)	na	24 614 (13.1)	na
Age at test				
<56 years	5932 (16.4)		8002 (16.1)	
≥56 years	7445 (11.6)	<0.0001	10 091 (11.4)	<0.0001
Referral pathway				
Screen detected	205 (8.7)		418 (9.0)	
Symptomatic	520 (13.7)	<0.0001	1119 (13.3)	<0.0001
Tissue type (sample type)				
Axilla/nodes	838 (14.9)	<0.05*	1006 (15.7)	<0.0001*
Primary tumour – core biopsy	9017 (13.2)	<0.05[†]	12 516 (13.0)	<0.0001[†]
Primary tumour – excision	3386 (13.6)	ns[‡]	4330 (13.1)	ns[‡]
Metastatic disease by site				
Bone	20 (9.0)	<0.05[§]	41 (9.2)	<0.05[§]
Cutaneous	162 (15.1)	ns[¶]	222 (15.0)	<0.05[¶]
Visceral	171 (14.6)	ns**	304 (15.3)	ns**
Histological type				
Carcinoma *in situ*	133 (5.7)		302 (8.0)	
IDC/NOS	7887 (15.2)		10 134 (15.3)	
ILC	446 (5.6)		573 (5.5)	
Mixed IDC/ILC	105 (8.5)		115 (7.9)	
Special type	68 (5.1)	See legend notes[††]	87 (5.3)	See legend notes[††]
Tumour grade				
Grade 1	268 (2.5)		332 (2.5)	
Grade 2	3723 (10.6)		4716 (10.8)	
Grade 3	4915 (24.2)	<0.0001[‡‡]	5821 (23.8)	<0.0001[‡‡]
Tumour size category (diameter)				
<20 mm	922 (12.1)		1188 (11.8)	
20 to <50 mm	831 (15.1)		1053 (13.9)	
≥50 mm	203 (16.5)	<0.0001[‡‡]	274 (14.6)	<0.0001[‡‡]
Nodal status				
N0	787 (11.4)		1006 (11.0)	
N1	388 (14.4)		482 (13.3)	
N2	141 (17.1)		172 (16.3)	
N3	163 (25.1)	<0.0001[‡‡]	184 (21.8)	<0.0001[‡‡]
ER status				
Negative	3182 (25.3)		4183 (25.5)	
Positive	6320 (10.5)	<0.0001	8317 (10.6)	<0.0001
PR status				
Negative	4106 (22.7)		5566 (23.4)	
Positive	2916 (8.5)	<0.0001	3836 (8.7)	<0.0001
Combined ER and PR status				
ER-negative/PR-negative	2492 (25.4)		3388 (26.1)	
ER-negative/PR-positive	161 (25.2)	ns	197 (25.9)	ns
ER-positive/PR-negative	1604 (19.5)		2155 (20.1)	
ER-positive/PR-positive	2750 (8.1)	<0.0001	3633 (8.4)	<0.0001

ER-negative: Centre assigned, or Allred <3, or % positive <1%, or H-score <1.
ER-positive: Centre assigned, or Allred ≥3, or % positive ≥1%, or H-score ≥1.
PR-negative: Centre assigned, or Allred <4, or % positive <10%, or H-score <10.
PR-positive: Centre assigned, or Allred ≥4, or % positive ≥10%, or H-score ≥10.
Refer to Table 1 for data on the total number of cases in each category.
na = not applicable, ns = not significant at $p < 0.05$.
*Axilla/nodes versus primary tumour – core biopsy.
[†]Axilla/nodes versus primary tumour – excision.
[‡]Primary tumour – core biopsy versus primary tumour – excision.
[§]Bone versus cutaneous.
[¶]Bone versus visceral.
**Cutaneous versus visceral.
[††]Rate for IDC/NOS was significantly different from all other groups ($p < 0.001$), which were not significantly different from each other at $p < 0.05$.
[‡‡]Test for trend.

Table 3. Distribution of cases by HER2 category, and by *HER2* gene amplification status for the 2+ category

	ER-neg/PR-neg	ER-neg/PR-pos	Total ER-neg	ER-pos/PR-neg	ER-pos/PR-pos	Total ER-pos	Totals
HER2 IHC status, *n* (%)							
0	4269 (36.7)	235 (33.9)	4504 (36.5)	2725 (26.3)	12 569 (30.2)	15 294 (29.4)	19 798 (30.8)
1+	2183 (18.7)	155 (22.4)	2338 (18.9)	2929 (28.2)	15 434 (37.1)	18 363 (35.3)	20 701 (32.2)
2+	1119 (9.6)	63 (9.1)	1182 (9.6)	1502 (14.5)	5621 (13.5)	7123 (13.7)	8305 (12.9)
3+	2955 (25.4)	177 (25.5)	3132 (25.4)	1719 (16.6)	2368 (5.7)	4087 (7.9)	7219 (11.2)
Totals*	11 645 (18.1*)	693 (1.1*)	12 338 (19.2*)	10 377 (16.1*)	41 613 (64.7*)	51 990 (80.8*)	64 328 (100.0*)
2+ amplification status, *n* (%)							
2+ (non-amplified)	837 (74.8)	49 (77.8)	886 (75.0)	1218 (81.1)	4770 (84.9)	5988 (84.1)	5988 (72.1)
2+ (amplified)	282 (25.2)	14 (22.2)	296 (25.0)	284 (18.9)	851 (15.1)	1135 (15.9)	1135 (15.7)

This analysis has been further categorised by ER and PR status (defined by final categorical result)
neg = negative, pos = positive.
*Percentage figures shown for the totals are proportions of the total cases in the whole analysis.

were ISH-amplified (UK: 17.3%). For all cases tested by ISH, regardless of IHC category, 17.0% were ISH-amplified (UK: 18.0%).

In cases reported as HER2 IHC 0, 0/1+, or 1+ the frequencies of those found to be ISH-amplified were 12.9% in UK cases (*n* = 39/303) and 7.0% in all cases (*n* = 62/891), indicative of disprortionately high false-negative rates when compared to those cited in UK guidelines (<2.0%) and other contemporaneous series (1.1 and 4.0%) [4–6]. Further analysis indicated that, within the UK group, 59.0% (*n* = 23/39) of the cases could be identified as originating from just two centres. By removing these the UK frequency became 5.7% (*n* = 16/280), which is still high, but perhaps less alarmingly so. For the whole population 59.7% (*n* = 37/62) of cases could be ascribed to three centres. Discounting these gave a realigned false-negative rate of 2.9% (*n* = 25/854), which is closely similar to published rates.

It is also notable that higher than expected frequencies of non-amplified cases were seen amongst the HER2 IHC 3+ tumours tested, with 11.5% of UK cases (*n* = 30/260), and 12.7% of all cases (*n* = 59/463) being non-amplified by ISH, contrasting significantly with published 'false-positive' rates of, for example, 3.9% in a single-centre series [6], and <6.0% in UK guidelines [5]. In a similar way to the false-negative findings, analysis indicated that the majority of cases arose within a small number of centres. Three centres produced 60.0% (*n* = 18/30) of the UK cases. By removing these the UK frequency became 5.0% (*n* = 12/242). Similarly, for the whole population 62.7% (*n* = 37/59) of cases could be ascribed to five centres. When these were removed the false-positive rate became 5.2% (*n* = 22/426). Interestingly, two centres appeared in both the false-negative and false-positive outlier lists, adding weight to the identification of their data as being outwith that of the testing community as a whole.

By considering the set of cases having both *HER2* gene and chromosome 17 centromeric probe (CEP17) copy number information, direct comparison of *HER2* gene amplification status as defined by *HER2* copy number alone and by *HER2*/CEP17 ratio could be made, based on published guidelines [5,7]. To avoid having small numbers of cases in some categories data are shown for the whole patient population only.

The findings were similar for all analysable cases (*n* = 12 049) and in the data set restricted to HER2 2+ cases (*n* = 10 523) (Table 4 and Supplementary material, Tables S7A–F):

- the proportion of amplified cases was significantly larger in ER-negative than in ER-positive tumours, regardless of how amplification status was defined (all *p* < 0.001);
- significantly fewer amplified cases were identified when copy number alone was used to define amplification status compared to the ratio alone, which in turn identified significantly fewer amplified cases than did the combination algorithm (all *p* < 0.001).

Figure 2E shows the distribution of *HER2* gene and CEP17 copy numbers and *HER2*/CEP17 ratio (figure legend gives detailed results).

ER and PR status

ER status was recorded for tumours from 94 887 patients, and in this population 82.7% (*n* = 78 465) were ER-positive; 82.6% of tumours from patients tested in UK centres were ER-positive (*n* = 59 909/72 510) (Supplementary material, Table S8).

Considering the effect of patient age on the prevalence of ER-positive tumours in the whole patient cohort; 66.6% of tumours in younger patients (age <35 years) were ER-positive, the proportion increased to 72.8% for patients 35–39 years of age and

Table 4. *HER2* gene amplification status as defined by *HER2* copy number alone, *HER2*/CEP17 ratio alone, and by algorithms for combined ratio/copy number given in published guidelines (see main paper text for references)

Category	Amplified by copy number (≥6.0)		Amplified by ratio (≥2.0)		Amplified by ratio and copy number combined algorithm	
	Count of cases (*n*)	Proportion of cases (%)	Count of cases (*n*)	Proportion of cases (%)	Count of cases (*n*)	Proportion of cases (%)
All HER2 IHC categories						
ER-positive	738	10.2	1235	17.0	1414	19.5
ER-negative	162	17.9	246	27.1	274	30.2
All	1328	11.0	2148	17.8	2415	20.0
HER2 IHC 2+ cases only						
ER-positive	594	9.4	1054	16.6	1204	19.0
ER-negative	123	16.1	200	26.1	225	29.4
All	1066	10.1	1842	17.5	2050	19.5

A total of 12 049 cases were available for analysis when HER2 IHC category was not taken into consideration of which, 10 523 (87.3%) were HER2 2+ by IHC

continued to increase with increasing age up to age-range 45–49, where it was 82.6%. After this age the rate remained about constant until age 75–79 years, after which it began to increase again, such that it was 86.7% in tumours from patients age ≥90 years. Results for the UK-specific patient population were closely similar to those quoted for the overall patient set (Figure 3A and Supplementary material, Table S8).

In tumours assessed for ER by Allred score it was possible to analyse the degree of positivity in addition to absolute ER status. The well documented bimodal distribution of Allred score is seen, with the vast majority of patients' tumours being strongly positive (score = 7 or 8) or completely negative (score = 0). In all patient age groups the majority of carcinomas were categorised as scoring 8 by the Allred score system but the proportion scoring 8 increased with increasing age (Figure 3B and Supplementary material, Table S10).

Using centre-assigned PR status, 65.1% were classified as positive (UK: 66.0%). No information was available on the cut-point employed, but modelling in the whole population using the percentage score data showed that a cut-point of ≥1% produced a PR-positive rate of 73.5%, and a cut-point of ≥10% gave a rate of 65.8%. Similar results were produced when Allred data were analysed, with a cut-point of ≥3 yielding a 70.8% PR-positive rate, and ≥4 a 66.9% rate. These data suggest that the use of a ≥10% or equivalent cut-point was more common (Supplementary material, Tables S11 and S12).

The effect of patient age on PR-positivity rate is illustrated in Figure 3C. In younger patients (age <35 years), 49.9% of tumours in the overall population were PR-positive and, in a similar way to ER, the positive proportion increased with age to a maximum level of 69.7% in the age-range 45–49 years. But, in contrast to the ER data, after this the positive rate declined again

until age-range 55–59 years where it was 63.1%, with the trend being significant ($p < 0.001$). From age 60 onwards it remained more or less constant. Results for the UK-specific patient populations were closely similar to those quoted for the whole patient set (Supplementary material, Table S13).

Combining data for tumours in the whole patient population where both ER and PR status were available, 63.9% were ER-positive/PR-positive, 15.9% ER-positive/PR-negative, 19.2% ER-negative/PR-negative, and 1.1% ER-negative/PR-positive (Supplementary material, Table S14).

Discussion

To the authors' knowledge, this publication presents the analysis of the largest clinical testing breast cancer biomarker data set reported to date. As such it represents an invaluable, contemporary resource for clinicians, clinical trialists, policy makers and scientists engaged in the area, providing robust benchmarking figures for the distributions of HER2, ER, and PR and their associations with patient and pathological variables.

It should result in a realignment of the HER2 positivity rate from 25%, which is often still quoted in the literature, and results from early reported analyses of populations heavily weighted towards the metastatic setting, to a figure of 13%, that is much more representative of modern-day testing regimes within the primary breast cancer setting. Support for this realignment can be found in the recent literature. For example, the paper by Lin *et al* examined HER2-positive rates within more than 95 000 patients tested in the US state of California between 2006 and 2011, finding a base-line rate of 15.9%. They also saw a similar trend to the one we report for higher HER2-positive
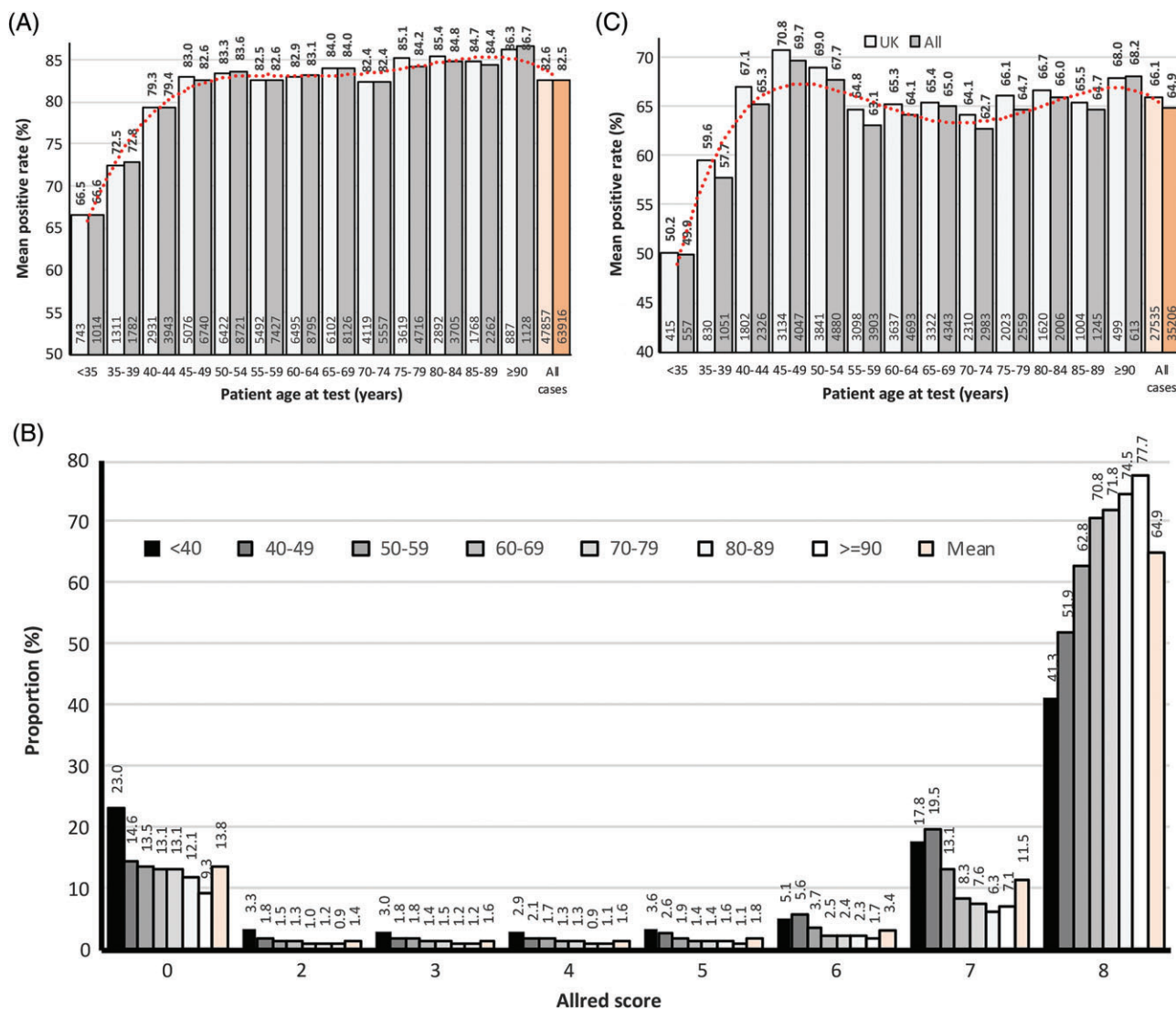
Figure 3. (A) Distribution of ER-positive rates by patient age at test. The final categorical status has been used to define ER positivity. Data are shown for UK and for all patients. A polynomial trend-line (red dotted-line) is displayed for the all patient data set. Figures above the columns are proportions of tumours, those inside the base of the columns are counts of ER-positive tumours. (B) Allred scores for ER expression by patient age. This is for all patients where ER status was categorised by Allred score (n = 56 282). Data are presented as patients <40 years and ≥90 years, with those from 40 to 89 being grouped in 10-year intervals. Figures above the bars on the graph indicate proportion (%) within that group. (C) Distribution of PR-positive rates by patient age at test. The final categorical status has been used to define PR positivity. Data are shown for UK and for all patients. A polynomial trend-line (red dotted-line) is displayed for the all patient data set. Figures above the columns are proportions of patients, those inside the base of the columns are numbers of PR-positive patients.

rates in younger compared to more elderly patients (26.9% in age <40 years, 17.9% in age 65+ years), and a higher prevalence for HER2-positive cases in the ER-positive versus the ER-negative tumour type (31.2% in ER-positive, 13.3% in ER-negative) [8]. The data presented also supports previously published UK studies such as that by Purdie et al, which reported a HER2-positive rate of 13.9% in a contemporary single-centre series of patients. The authors of

that paper also saw a similar negative association between HER2 status and PR status in ER-positive patients to that reported here [9].

Similarly, historically quoted ER-positive rates for primary invasive breast cancer of 70% should be replaced by the higher rate of approximately 83% observed here. In support of this, the previously cited work of Purdie et al in a UK population, and two recently published studies examining the incidence of ER-positive tumours

in patients in Denmark and in Ireland report rates of 80.2, 79.1 and 82% respectively [9–11]. The higher rate seen in data from contemporary populations may, at least in part, result from the use of more sensitive primary antibodies and detection systems and suggest a need to re-examine the cut-points employed. Given our finding that the prevalence of ER-positive tumours increases with patient age beyond 75 years, and the predictions from other groups that breast cancer incidence will rise especially in women age 70–84 years [12], it can be expected that the overall ER-positive rate may increase further in the coming years.

We believe that the frequency distrubtions and patterns of reporting presented will be of value to the clinical pathologist. For example, the finding that within the HER2 2+ category, a larger proportion of ER-negative than ER-positive tumours are subsequently shown to be *HER2* gene amplified (25.0 versus 15.9%) is novel. Similarly, the observation that within the group of ER-positive tumours, the proportion classified as either negative or positive by IHC varies by PR status is potentially valuable for comparison with ones own local data.

An important *proviso* that should be noted is that no reliable method for predicting HER2 status using the data presented here is suggested by the authors. Guidelines mandate the testing of all primary breast cancers for HER2 irrespective of statistical associations with clinical, pathological, or any other features.

This study's major strengths are:

- the size of the data set, comprising results on testing of almost 200 000 patients across a number of years, making it largely immune to temporal variability in reagents and platforms;
- that it draws data from almost all UK clinical testing laboratories and a significant proportion of those in the RoI;
- that it amalgamates results obtained by all testing modalities in clinical routine use.

Its major weakness is that data were entered largely as free-text, without error checking at source, and as such are subject to data entry errors. As noted previously, an attempt to minimise potential errors arising from this has been made by excluding data from centres entering <200 individual entries on the basis that such users are less experienced in data entry and also less likely to use the data for their own internal audit and quality control processes (and thus correct any errors discovered). Where it has been possible to compare the results to other sources this has been done and indications from these 'reality checks' are that the data is robust and concordant.

Space considerations have meant that only the major analyses for the three main breast cancer biomarkers

have been presented, but substantial additional data have been made available in the accompanying supplementary tables and figures, enabling researchers to derive further sub-analyses of their own.

Conclusions

In summary of the main findings.
HER2 positivity rate:

- 13.1% for the whole analysed population (n = 24 614/187 368)
- 10.6% in ER-positive tumours (n = 8317/78 465)
- 25.5% in ER-negative tumours (n = 4183/16 422)

Increasing HER2 positivity rates were significantly associated with:

- patient age <56 years versus age ≥56 years
- symptomatic versus screen-detected tumours
- involved axillary node sample versus primary breast cancer sample
- IDC/NOS versus other histological types
- higher histological grade
- increasing tumour size
- increasing nodal involvement
- PR-negative versus PR-positive tumour status

ER positivity rate:

- 82.7% for the whole analysable population (n = 78 465/94 887)
- 80.8% in tumours from women age <56 years (n = 23 586/29 173)
- 83.6% in tumours from women age ≥56 years (n = 40 330/48 256)

PR positivity rate:

- 64.9% for the whole analysable population (n = 44 070/67 878)
- 65.8% in tumours from women age <56 years (n = 13 578/20 632)
- 64.4% in tumours from women age ≥56 years (n = 21 628/33 600)

Acknowledgements

Author contributions statement

KM, MI, and SP initiated and administered the running of the database. SP oversaw data collation. AD and MD conceived the analyses. AD conducted data analysis. MD, JMSB, and SP critically reviewed the article. JMSB suggested additional analyses. All authors were involved in writing the paper and had final approval of the submitted and published versions.

REFERENCES

1. UK-NEQAS ICC & ISH. [Accessed 01 May 2018]. Available from: www.ukneqasiccish.org
2. Ibrahim M, Herriot K, Parry S, *et al*. Audit of breast biomarkers markers, HER2.ER & PR in the UK: an update by the UK National External Quality Assessment Scheme for Immunocytochemistry and In Situ Hybridisation (UK NEQAS ICC & ISH). *Virchows Archiv* 2011; **459**: S74.
3. Bartlett JM, Ellis IO, Dowsett M, *et al*. Human epidermal growth factor receptor 2 status correlates with lymph node involvement in patients with estrogen receptor (ER) negative, but with grade in those with ER-positive early-stage breast cancer suitable for cytotoxic chemotherapy. *J Clin Oncol* 2007; **25**: 4423–4430.
4. Kaufman PA, Bloom KJ, Burris H, *et al*. Assessing the discordance rate between local and central HER2 testing in women with locally determined HER2-negative breast cancer. *Cancer* 2014; **120**: 2657–2664.
5. Rakha EA, Pinder SE, Bartlett JM, *et al*. Updated UK recommendations for HER2 assessment in breast cancer. *J Clin Pathol* 2015; **68**: 93–99.
6. Purdie CA, Jordan LB, McCullough JB, *et al*. HER2 assessment on core biopsy specimens using monoclonal antibody CB11 accurately determines HER2 status in breast carcinoma. *Histopathology* 2010; **56**: 702–707.
7. Wolff AC, Hammond ME, Hicks DG, *et al*. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol* 2013; **31**: 3997–4013.
8. Lin CY, Carneal EE, Lichtensztajn DY, *et al*. Regional variability in percentage of breast cancers reported as positive for HER2 in California: implications of patient demographics on laboratory benchmarks. *Am J Clin Pathol* 2017; **148**: 199–207.
9. Purdie CA, Quinlan P, Jordan LB, *et al*. Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. *Br J Cancer* 2014; **110**: 565–572.
10. Mullooly M, Murphy J, Gierach GL, *et al*. Divergent oestrogen receptor-specific breast cancer trends in Ireland (2004-2013): amassing data from independent western populations provide etiologic clues. *Eur J Cancer* 2017; **86**: 326–333.
11. Anderson WF, Rosenberg PS, Petito L, *et al*. Divergent estrogen receptor-positive and -negative breast cancer trends and etiologic heterogeneity in Denmark. *Int J Cancer* 2013; **133**: 2201–2206.
12. Rosenberg PS, Barker KA, Anderson WF. Estrogen receptor status and the future burden of invasive and in situ breast cancers in the United States. *J Natl Cancer Inst* 2015; **107**: djv159.

SUPPLEMENTARY MATERIAL ONLINE

Figure S1. Tumour size at histological assessment
Table S1. Records entered
Table S2. HER2 positivity rates
Table S3. HER2 positivity rates by year
Table S4. HER2 positivity rates by ER status
Table S5. *HER2* gene amplification rates distributed by HER2 IHC category
Table S6. HER2 status associations
Table S7. *HER2* gene amplification by copy number, *HER2*/CEP17 ratio and combined
Table S8. ER status
Table S9. ER positive rates by patient age
Table S10. ER status by age, categories by Allred score
Table S11. PR status using two different cut-points
Table S12. PR status by different scoring methods
Table S13. PR status, using finalised ER status
Table S14. Combined ER and PR status case distribution

Supplementary material file containing the data from the Republic of Ireland