Application of the ESMO-Magnitude of Clinical Benefit Scale (V.1.1) to the field of early breast cancer therapies

Shani Paluch-Shimon 1, Nathan I Cherny, Elisabeth G E de Vries, Urania Dafni, Martine J Piccart, Nicola Jane Latino, Fatima Cardoso

ABSTRACT

Background The European Society for Medical Oncology-Magnitude of Clinical Benefit Scale (ESMO-MCBS) is a validated value scale for solid tumour anticancer treatments. Form 1 of the ESMO-MCBS, used to grade therapies with curative intent including adjuvant therapies, has only been evaluated for a limited number of studies. This is the first large-scale field testing in early breast cancer to assess the applicability of the scale to this data set and the reasonableness of derived scores and to identify any shortcomings to be addressed in future modifications of the scale.

Method Representative key studies and meta-analyses of the major modalities of adjuvant systemic therapy of breast cancer were identified for each of the major clinical scenarios (HER2-positive, HER2-negative, endocrine-responsive) and were graded with form 1 of the ESMO-MCBS. These generated scores were reviewed by a panel of experts for reasonableness. Shortcomings and issues related to the application of the scale and interpretation of results were identified and critically evaluated.

Results Sixty-five studies were eligible for evaluation: 59 individual studies and 6 meta-analyses. These studies incorporated 101 therapeutic comparisons, 61 of which were scorable. Review of the generated scores indicated that, with few exceptions, they generally reflected contemporary standards of practice. Six shortcomings were identified related to grading based on disease-free survival (DFS), lack of information regarding acute and long-term toxicity and an inability to grade single-arm de-escalation scales.

Conclusions Form 1 of the ESMO-MCBS is a robust tool for the evaluation of the magnitude of benefit studies in early breast cancer. The scale can be further improved by addressing issues related to grading based on DFS, annotating grades with information regarding acute and long-term toxicity and developing an approach to grade single-arm de-escalation studies.

INTRODUCTION

As the population ages, the incidence and prevalence of cancer are expected to continue to rise both in developed1 and developing countries.2 The estimated total annual economic cost of cancer was US$1.16 trillion in 2010, about 2% of global gross domestic product3 and is continuing to rise exponentially. Breast cancer remains the leading cause of cancer among women4 and the ongoing care of breast cancer patients is estimated to be one of the most significant contributors to growing cancer care expenditure.5

These considerations underscore the need for validated tools to evaluate value of care, where value is recognised as a balance between clinical benefit and cost. With this in mind, both the European Society for Medical Oncology (ESMO) and the American Society of Clinical Oncology (ASCO) established...
Working Groups to address these issues and they have developed and published a platform for evaluating new anticancer therapeutics—the ESMO-Magnitude of Clinical Benefit Scale (ESMO-MCBS)—and the ASCO Framework for assessing value of cancer care.6

The ESMO-MCBS was initially launched and published in 20155 and revised in 2017 with version 1.1.7 The scale aims to provide a validated and rational stratification process for oncology therapies, and its development process has been predicated on ‘accountability for reasonableness’ which incorporated extensive field testing and the peer review of results for ‘reasonableness’.3 Form 1 of the ESMO-MCBS, which is used to grade therapies with curative intent including adjuvant therapies, hitherto, has only been applied in a limited number of studies. Form 1 of the ESMO-MCBS grades therapies with curative intent on a three-point scale A, B and C where scores of A and B represent substantial improvement.

This is the first large-scale field testing of form 1 in early breast cancer to assess the applicability of the ESMO-MCBS in this setting, to determine whether the scoring reflected clinical practice (reasonableness) and to identify shortcomings to be addressed in future versions of the scale. It also provides an overview of the magnitude of benefit for the most common therapies/therapeutic strategies in the field of breast cancer, allowing for a critical reassessment of available options.

METHODOLOGY
ESMO-MCBS V.1.1 form 1, designed to evaluate adjuvant and neoadjuvant studies, was applied to all the selected studies (online supplementary data).

Representative key studies and meta-analyses of the major modalities of adjuvant systemic therapy of breast cancer (chemotherapy or endocrine therapy or anti-HER2 therapy) were identified for each of the major clinical scenarios (HER2-positive, HER2-negative, endocrine-responsive). Studies were identified through PubMed, Food and Drug Administration (FDA) and European Medicines Agency (EMA) registration sites. Pivotal phase 3 studies that have formed the basis for contemporary treatment practice and a randomised phase 2 study that resulted in preliminary drug registration8 were scored.

To identify the pivotal phase 3 studies, a PubMed search was performed with the following search criteria: “breast cancer”[Title] AND breast[Title] AND cancer[Title] AND adjuvant[Title] OR neo-adjuvant AND “2002”[Date—Publication] : “2019”[Date—Publication] AND English[Language] AND “randomized controlled trial” OR “phase 3” OR “randomized phase 2” NOT retrospective[Title/Abstract] NOT historical[Title/Abstract] NOT “systematic review”[Title] NOT advanced[Title] NOT metastatic[Title] NOT irradiation[Title] NOT safety[Title] NOT insights[Title] NOT observations[Title] NOT “quality of life”[Title] NOT biosimilar[Title] NOT analysis[Title] NOT analyses[Title] NOT radiation[Title]. There were 597 studies identified from the search. Relevant studies that were comparative phase 3 randomised controlled studies were identified and subsequently cross-referenced with the FDA and EMA registration sites and ESMO9 and National Comprehensive Cancer Network (NCCN)10 guidelines to identify pivotal and practice changing studies. Key meta-analyses referenced by ESMO9 and NCCN10 guidelines were identified.

Studies were eligible for scoring if they were randomised comparative studies comparing new therapies to standard of care or meta-analyses of those studies. Studies were scored if they met the scoring criteria defined by the ESMO-MCBS guideline according to the criteria in form 1. Where missing data impeded scoring, the corresponding author was contacted with a request for data or clarification. If no response was received, the study was either marked as not scorable (this occurred for only one study11 and one meta-analysis12) or excluded (if there was inadequate data reported). All scoring was reviewed for accuracy by members of the Magnitude of Clinical Benefit Working Group and the generated scores were reviewed by the ESMO Breast Cancer Faculty for reasonableness.

Scoring was performed in accordance with the rules for application of the ESMO-MCBS.5,7 Studies initially evaluated based on disease-free survival (DFS) criteria alone or pathological complete remission (PCR) rate were re-evaluated when mature overall survival (OS) data are available and a final score was determined based on these OS results. The only exception was for studies that were un-blinded after compelling early DFS results with subsequent access to the superior arm, whereby OS results were contaminated by the crossover and therefore were not evaluable.

Studies that could not be scored were classified into one of three groups: (1) studies that did not achieve statistical significance, designated ‘no evaluable benefit’ (NEB), (2) non-inferiority studies in which non-inferiority was not verified, designated ‘negative non-inferiority’ (NNI), (3) studies that could not be scored because required data were not included in the publication, designated ‘scoring not applicable’ (SNA) and (4) not-scorable subgroup data

RESULTS
Sixty-five studies were eligible for evaluation: 59 individual studies and 6 meta-analyses (5 of which were individual patient-level data meta-analyses), which yielded data relevant to 101 therapeutic comparisons, 61 of which demonstrated significant benefit or non-inferiority and could be scored.

Adjuvant chemotherapy
Polychemotherapy versus no chemotherapy
Both cyclophosphamide methotrexate and 5-fluorouracil (CMF) and anthracycline-based therapy were found to be superior to no chemotherapy (in a predominantly node-positive population), both scoring an A compared with no treatment in the meta-analysis, with a 15-year gain
Post-neoadjuvant chemotherapy trials comparing TC \times 6 to combinations including AC regimens.22

In all the dose-dense (dd) regimen trials, the high-risk
node-positive population demonstrated a benefit to \text{AC} \times 4, demonstrating a 6% gain in OS at 7-
year follow-up (grade A).21 However, joint analysis of three
trials comparing TC \times 6 to combinations including AC
colours consisted predominantly of patients with node-
positive disease.

Table 1 First-generation adjuvant chemotherapy

| Tested agent | Trial name | Setting | Primary outcome | Median follow-up (years) | DFS control (%) | DFS gain (%) | OS control | OS HR* | pCR | ESMO-MCBS V.1.1 | Reference |
|--------------|------------|---------|----------------|-------------------------|----------------|-------------|------------|--------|-----|----------------|-----------|
| Polychemotherapy vs none (meta-analysis) | EBCTCG | Anthracycline vs nil 82% Node+ | OS | 15 | 47.4 | 8.0 | 0.73 (0.68–0.79) | Breast cancer mortality 35.8% | 6.5 | 0.79 (0.72–0.85) | A | 13 |
| Polychemotherapy vs none (meta-analysis) | EBCTCG | CMF vs nil 34% Node+ | OS | 15 | 39.8 | 10.2 | 0.70 (0.63–0.77) | Breast cancer mortality 27.8% | 6.2 | 0.76 (0.68–0.84) | A | 13 |
| CMF \times 6 vs CAF \times 6 | INT 0102 | High-risk Node- | DFS | 10 | 75.0 | 2.0 | 1.05 (0.94–1.27) | DFS 82.0% | 3.0 | 1.19 (0.99–1.43) | NEB | 14 |
| CMF \times 6 vs CEF \times 6 | MA5 | Node+ | RFS | 10 | 45.0 | 7.0 | 1.31 (1.06–1.61) | RFS 58.0% | 4.0 | 1.18 (0.94–1.49) | NEB | 15 |
| CMF vs AC \times 4 (meta-analysis) | EBCTCG | 61% Node+ | OS | 10 | 42.1 | 1.1 | 0.99 (0.90–1.08) | Breast cancer mortality 32.5% | 0.9 | 0.98 (0.89–1.08) | NEB | 13 |
| CMF vs 4 AC \times 4 (meta-analysis) | EBCTCG | 61% Node+ | OS | 10 | 33.8 | 2.6 | 0.89 (0.82–0.96) | Overall mortality 34.6% | 1.2 | 0.97 (0.89–1.07) | NEB | 13 |
| CMF vs CAF/FEC (meta-analysis) | EBCTCG | 53% Node+ | OS | 10 | 33.8 | 2.6 | 0.89 (0.82–0.96) | Breast cancer mortality 24.1% | 4.1 | 0.80 (0.72–0.88) | B | 13 |

Chart blanks—relevant variables not available in manuscript.
*HR values in parentheses refer to 95% CI.
A, doxorubicin; C, cyclophosphamide; DFS, disease-free survival; E, epirubicin; ESMO-MCBS, The European Society for Medical Oncology—Magnitude of Clinical Benefit Scale; F, fluorouracil; M, methotrexate; NEB, no evaluable benefit; Node+, node-positive; Node-, node-negative; OS, overall survival; RFS, relapse-free survival.
Table 2  Adjuvant chemotherapy with the addition of taxane

| Study | Trial name | Setting | Primary outcome | Median follow-up | DFS control (%) | DFS gain (%) | DFS HR* | OS control (%) | OS gain (%) | OS HR* | ESMO-MCBS V.1.1 | Reference |
|-------|------------|---------|-----------------|------------------|----------------|-------------|---------|----------------|-------------|---------|----------------|-----------|
| TAC vs FAC | GEICAM 9805 | High-risk Node− | DFS | 77 months | 81.8 | 6.0 | 0.68 (0.49–0.93) | 93.5% | 1.7 | 0.76 (0.45–1.26) | NEB | 16 |
| FEC-P vs FEC | GEICAM 9906 | Node+ | DFS | 66 months | 72.1 | 6.4 | 0.74 (0.60–0.92) | 87.1% | 2.8 | 0.78 (0.57–1.06) | NEB | 17 |
| AC×4 vs AC-P | NSABP-B28 | Node+ | DFS and OS | 64.6 months | 72.0 | 4.0 | 0.83 (0.72–0.95) | 85.0% | 0 | 0.93 (0.78–1.12) | NEB | 18 |
| AC-P vs CEF×6 | MA-21 | High-risk Node− and + | RFS | 30 months | 85.0† | 5.1 | 1.49 (1.12–1.99)‡ | A§ | 19 |
| AC-P vs dose dense EC×6 >P×4 q21 | MA-21 | High-risk Node− and + | RFS | 30 months | 85.0† | 4.5 | 1.68 (1.25–2.27)‡ | A§ | 19 |
| Paclitaxel q21 d vs q7 day | E1199 | High-risk LN−/LN+ | DFS | 12 years | 65.5‡ | 5.2 | 0.84 (0.73–0.96) | 75.3%‡ | 2.4 | 1.02 (0.88–1.18) | NEB | 79 82 |
| Paclitaxel q21 d vs docetaxel q21 day | E1199 | High-risk LN−/LN+ | DFS | 12 years | 65.5‡ | 6.4 | 0.79 (0.68–0.90) | 75.3%‡ | 3.2 | 0.86 (0.73–1.00) | NEB | 79 82 |
| TAC vs AC-T | BCIRG 005 | Node+ | DFS | 65 months | 79.0 | 0 | 1.0 (0.86–1.16) | 88.0% | 1.0 | 0.91 (0.75–1.11) | NEB | 81 |
| TAC vs dd AC-P | NSABP-B38 | Node+ | DFS | 64 months | 80.1 | 2.1 | NS | 89.6% | −0.5 | NS | NEB | 82 |
| TAC vs dd AC-PG | NSABP-B38 | Node+ | DFS | 64 months | 80.1 | 0.5 | NS | 89.6% | 0.8 | NS | NEB | 82 |
| TC vs AC | US Oncology 9735 | High-risk Node− and 1–3+nodes | DFS | 7 years | 75.0 | 6.0 | 0.74 (0.56–0.98) | 84.0% | 6.0 | 0.69 (0.50–0.97) | A | 20 21 |
| TC×6 vs Tax AC | ABC Trials— joint analysis | iDFS non-inferiority | 3.3 years | 88.2† | 2.5 | 1.2 (0.97–1.49) | NNI | 22 |
| TC×6 vs Tax AC Chemo+Taxane (meta-analysis) | EBCTCG | Chemo+Taxe+ (Node+100%) | iDFS non-inferiority | 8 years | 34.8 | 4.6 | 0.84 (0.78–0.91) | Breast cancer mortality 23.9% | 2.8 | 0.86 (0.79–0.93) | C | 13 |
| Chemo+Tax (meta-analysis) | EBCTCG | Tax/chemo vs different non-Tax regimen (Node+82%) | OS | 8 years | 22.0 | 2.9 | 0.86 (0.82–0.91) | Breast cancer mortality 11.5% | 1.4 | 0.88 (0.81–0.95) | C | 13 |

Chart blanks—relevant variables not available in manuscript.
*HR values in parentheses refer to 95% CI.
†AC-T6 arm.
‡Three weekly paclitaxel arm.
§No OS data published.
††TC×6 arm.
A, doxorubicin; C, cyclophosphamide; Chemo, chemotherapy; dd, dose dense; DFS, disease-free survival; ESMO-MCBS, The European Society for Medical Oncology-Magnitude of Clinical Benefit Scale; F, fluorouracil; G, gemcitabine; iDFS, invasive disease-free survival; NEB, no evaluable benefit; NNI, negative non-inferiority; Node+, node-positive; Node−, node-negative; OS, overall survival; P, paclitaxel; RFS, relapse-free survival; T, docetaxel; Tax, taxane.
Table 3  Chemotherapy approaches: dose density, neoadjuvant approaches, post-neoadjuvant

| Study                                | Trial name     | Setting | Primary outcome | Median follow-up | DFS control | DFS gain | DFS HR* | OS control | OS gain (%) | OS HR* | pCR | Reference |
|--------------------------------------|----------------|---------|-----------------|------------------|-------------|----------|---------|------------|-------------|--------|-----|-----------|
| dd studies                           |                |         |                 |                  |             |          |         |            |             |        |     |           |
| AC-P q21 vs q14                       | INT C9741/ CALGB 9741 | Node+  | DFS             | 36 months        | 75.00%      | 7.00%    | 0.74 (0.59-0.93) | 90.00%      | 2.00   | 0.69 (0.50-0.93) | C   | 23 |
| EG-P q21 vs q14                       | AGO            | Node+  | EFS             | 62 months        | 62.00%      | 8.00%    | 0.72 (0.59-0.87) | 77.00%      | 5.00   | 0.76 (0.59-0.97) | B   | 24 |
| (F)EC-P q21 vs q14                    | GIM            | Node+  | DFS             | 7 years          | 76.00%      | 5.00%    | 0.77 (0.65-0.92) | 89.00%      | 5.00   | 0.65 (0.51-0.84) | B   | 25 |
| Meta-analysis—dd vs regular schedule | 2 weekly vs 3 weekly (same regimen) | DFS and OS | 10 years | 28.30% | 4.30% | 0.83 (0.76-0.91) | Breast cancer mortality 19.6% | 1.80   | 0.86 (0.77-0.95) | C   | 26 |
| Meta-analysis—dd vs regular schedule | Pooled-analysis—all dd and sequential | DFS and OS | 10 years | 32.00% | 3.60% | 0.65 (0.81-0.89) | Breast cancer mortality 19.6% | 2.70   | 0.87 (0.82-0.92) | C   | 26 |
| Meta-analysis—dd vs regular schedule | Pooled-analysis—all dd and sequential | DFS and OS | 10 years |         |         |         |         | All-cause mortality 25.5% | 3.00   | 0.87 (0.82-0.91) | B   | 26 |
| dd vs regular schedule                |                |         |                 |                  |             |          |         |            |             |        |     |           |
| Stratified for HR status              | DDFS and OS    | Variable 2-10 years | All | 0.84 (0.77-0.91) | 0.85 (0.79-0.93) | B | 27 |
| Stratified for HR status              | DDFS and OS    | Variable 2-10 years | HR-ve | 0.80 (0.69-0.92) | 0.74 (0.69-0.89) | B | 27 |
| Stratified for HR status              | DDFS and OS    | Variable 2-10 years | HR+ve | 0.93 (0.82-1.05) | 0.81 (0.73-0.91) | NEB | 27 |
| Post-neoadjuvant chemotherapy         |                |         |                 |                  |             |          |         |            |             |        |     |           |
| Capecitabine vs placebo               | CREATE—X       |         | Residual disease after neoadjuvant therapy | DFS | 3.6 years | All 67.6% | 6.50% | 0.70 (0.53-0.92) | 83.60% | 5.60 | 0.59 (0.39-0.90)* | A | 28 |
| Capecitabine vs placebo               | CREATE—X       |         | Residual disease after neoadjuvant therapy | DFS | 3.6 years | Triple negative 56.1% | 13.70% | 0.58 (0.39-0.87) | 70.30% | 8.50 | 0.52 (0.30-0.90) | A | 28 |
| Capecitabine vs placebo               | CREATE—X       |         | Residual disease after neoadjuvant therapy | DFS | 3.6 years | HR+/HER2- neg 73.4% | 3.00% | 0.81 (0.55-1.17) | 90.00% | 3.40 | 0.73 (0.38-1.14) | NEB | 28 |
| Neoadjuvant carboplatin               |                |         |                 |                  |             |          |         |            |             |        |     |           |
| Neoadjuvant Peg-A+P+Bev vs Peg-A+P+Bev+Carbo | GeparSixto | Triple negative | pCR | 35 months | 76.10% | 9.70% | 0.56 (0.33-0.96) | 41% vs 56.8% (ns) | 41% vs 56.8% (ns) | A | 29 |
| P vs P + carboplatin                  | BRIGHTNESS     | Triple negative | pCR |             |         |         |         |            |             |        |     |           |
| P + carboplatin vs P + carboplatin + veliparib | BRIGHTNESS | Triple negative | pCR |             |         |         |         |            |             |        |     |           |
| Neoadjuvant AC+P ± carboplatin        | CALBG 40603    | Triple negative | pCR | 3 years | 71.00% | 5.00% | 0.84 (0.58-1.22) | 85.00% | -4.00 | 1.15 (0.74-1.79) | 41% vs 54% (ns) | NEB | 31 |
| Neoadjuvant AC-P ± bevacizumab        | CALBG 40603    | Triple negative | pCR | 3 years | 72.00% | 3.00% | 0.80 (0.55-1.17) | 81.00% | 4.00 | 0.76 (0.49-1.19) | 44% vs 52% (ns) | NEB | 31 |

Continued
Table 3 Continued

| Study | Trial name | Setting | Primary outcome | Median follow-up (DFS gain) | DFS HR* | OS control (%/OS gain) | OS HR* | pCR | ESMO-MCBS V1.1 | Reference |
|-------|------------|---------|----------------|-----------------------------|---------|------------------------|---------|-----|----------------|-----------|
| Neoadjuvant other agents | | | | | | | | | | |
| D+AC vs DG+AC | NSABP-B40 | HER2-negative | pCR | 4.7 years | 72.80% | 1.10% | 0.90 (0.67–1.19) | 80.90% | 4.80 | 0.73 (0.51–1.04) | 32.7% vs 31.8% (ns) | NEB | 11 |
| D+AC vs DX+AC | NSABP-B40 | HER2-negative | pCR | 4.7 years | 72.80% | −0.20% | 1.01 (0.77–1.33) | 80.90% | 0.60 | 0.96 (0.68–1.32) | 32.7% vs 29.7% (ns) | NEB | 11 |
| Above regimens ± Bev | NSABP-B40 | HER2-negative | pCR | 4.7 years | 72.80% | 4.00% | 0.80 (0.63–1.01) | 80.90% | 0.65 | 0.65 (0.49–0.88) | 28.2 vs 34.5% (ns) | B/C | 11 |
| EC+D vs EC+D+Bev | GeparQuinto | Neoadjuvant—all subtypes | pCR | 3.8 years | 81.50% | −2.00% | 1.03 (0.84–1.25) | 88.70% | 2.00 | 0.97 (0.75–1.26) | 14.9 vs 18.4% (ns) | NEB | 83 |
| P-EC vs Nab-P-EC | GeparSepto | Neoadjuvant—all subtypes | pCR | 29% vs 38.4% (ss) | | | | | | | |

Chart blanks—relevant variables not available in manuscript. *HR values in parentheses refer to 95% CI.

A, doxorubicin; Bev, bevacizumab; C, cyclophosphamide; Carbo, carboplatin; D, docetaxel; D+AC, docetaxel + cisplatin; DFS, disease-free survival; E, epirubicin; EFS, event-free survival; ESMO-MCBS, The European Society for Medical Oncology—Magnitude of Clinical Benefit Scale; F, fluorouracil; G, gemcitabine; HER2-neg, HER2-negative; HR+, hormone-positive; iDFS, invasive disease-free survival; Nab-P, Nab-paclitaxel; nc, not statistically significant; NEB, no evaluable benefit; OS, overall survival; P, paclitaxel; pCR, pathological complete response; Peg-A, pegylated doxorubicin; q21, every 21 days; ss, statistically significant; V, veliparib; X, capecitabine.

In the BRIGHTNESS study of 15.8% compared with the non-carboplatin arm.30 The CALGB 40603 did not demonstrate an outcome benefit from the addition of neoadjuvant carboplatin or bevacizumab despite improvements in pCR and was categorised as NEB.31 In the NSABP B40 study, there was no benefit of the addition of gemcitabine or capecitabine to standard neoadjuvant chemotherapy regimens.11 32 This study reported an OS benefit from the addition of neoadjuvant bevacizumab with a HR of 0.65 (95% CI 0.49–0.88); however, since the absolute survival benefit was not published, this was not evaluatable (SNA).11

In the GeparSepto study, neoadjuvant nab-paclitaxel demonstrated a limited improvement in pCR rate compared with paclitaxel, however the gain was below the ESMO-MCBS threshold for scoring the ≥30% relative and >15% absolute pCR gain.33

Anti-HER2 therapies

All the 12-month adjacent trastuzumab studies demonstrated substantial benefit (grade A or B).34–36 Two years of trastuzumab was not superior to 12 months.34 While several studies failed to demonstrate non-inferiority of shorter duration of trastuzumab therapy,37–39 the PERSEPHONE study demonstrated non-inferiority for 6 months versus 12 months of trastuzumab and scored a B based on non-inferiority and reduced cost (table 4).40

Dual blockade

Four of the five studies testing double blockade with trastuzumab plus pertuzumab to treat pCR demonstrated large improvement in DFS compared with trastuzumab alone.41 The node-positive subgroup was not scorable since this was 1 of 12 evaluated subgroups in an exploratory analysis and was therefore not eligible for grading (of note, the ESMO-MCBS allows only for scoring of subgroups if there were up to three planned subgroup analyses in the study design).41

In the APHINITY study, evaluating the addition of pertuzumab to trastuzumab in the ITT population, which had a similar pCR gain but no OS benefit.42 43

Second-generation anti-HER2 therapies

In the ExteNET study, the addition of neratinib for completion of 1 year of trastuzumab scored a grade A.44
| Study | Trial name | Setting | Primary outcome | Median follow-up | DFS control | DFS gain | DFS HR* | OS control | OS gain | OS HR* | ESMO-MCBS V1.1 | Reference |
|-------|------------|---------|----------------|------------------|-------------|----------|---------|------------|---------|---------|--------------|-----------|
| Chemotherapy+trastuzumab | HERA | Adjuvant or neoadjuvant HER2+ tumours | DFS | 2 years | DFS 77.4% | 8.4% | 0.54 (0.43–0.67) | – | Early crossover at interim analysis | A | 34 |
| AC-P vs AC-PH or TCaH | BCIRG006 | Adjuvant HER2+ tumours | DFS | 65 months | 75.0% | AC-PH-9% | 0.63 | 87.0% | AC-PH-5% | 0.75 | A | 36 |
| AC-P vs AC-PH or TCaH | BCIRG006 | Adjuvant HER2+ tumours | DFS | 65 months | 75.0% | TCaH-6% | 0.75 | TCaH-4% | 0.77 | B | (Slamon, 2011 #1663; Slamon, 2011 #1663) |
| AC-P vs AC-PH | NSABP B31-NCCGT | Adjuvant HER2+ tumours | DFS | 8.4 years | 10 years | DFS 62.2 | 11.5% | 10 years OS | 75.2% | 0.63 (0.54–0.73) | A | (Perez, 2014 #1669) |
| Adjuvant chemotherapy+trastuzumab | Meta-analysis | HER2+, <2 cm stratified for HR and nodal status | DFS and OS HR+all | 8 years | 75.7% | 7.0% | 0.70 (0.59–0.83) | 88.4% | 3.8% | 0.68 (0.52–0.89) | B | (O’Sullivan, 2015 #1811) |
| Adjuvant chemotherapy+trastuzumab | Meta-analysis | HER2+, <2 cm stratified for HR and nodal status | DFS and OS HR+<1 node | 8 years | 81.6% | 3.8% | 0.64 (0.47–0.83) | 92.6% | 2.1% | 0.68 (0.42–1.10) | NEB | (O’Sullivan, 2015 #1811) |
| Adjuvant chemotherapy+trastuzumab | Meta-analysis | HER2+, <2 cm stratified for HR and nodal status | DFS and OS HR- all | 8 years | 66.4% | 9.4% | 0.66 (0.49–0.88) | 78.8% | 8.8% | 0.59 (0.47–0.74) | A | (O’Sullivan, 2015 #1811) |
| Adjuvant chemotherapy+trastuzumab | Meta-analysis | HER2+, <2 cm stratified for HR and nodal status | DFS and OS HR- <1 node | 8 years | 73.7% | 5.9% | 0.77 (0.59–1.00) | 87.8% | 4.0% | 0.69 (0.66–1.04) | NEB | (O’Sullivan, 2015 #1811) |

Chart blanks—relevant variables not available in manuscript.
*HR values in parentheses refer to 95% CI.
A, doxorubicin; C, cyclophosphamide; Ca, carboplatin; chemo, chemotherapy; DFS, disease-free survival; ESMO-MCBS, The European Society for Medical Oncology-Magnitude of Clinical Benefit Scale; H, trastuzumab; HER2+, HER2 overexpression; HR+, hormone receptor-positive; NEB, no evaluable benefit; OS, overall survival; P, paclitaxel; T, docetaxel.
Table 5  HER2 double blockade and second-generation anti-HER2 therapies

| Study | Trial name | Setting | Primary outcome | Median follow-up | DFS control (%) | DFS gain (%) | DFS HR* | OS control (%) | OS gain (%) | OS HR* | pCR | QoL | Toxicity | ESMO-MCBS V.1.1 | Reference |
|-------|------------|---------|----------------|-----------------|-----------------|--------------|---------|----------------|--------------|---------|-----|-----|----------|----------------|-----------|
| Neratinib vs placebo | ExtentNET | Stage 2–3 HER2+ after 12 months trastuzumab, stratified for HR status | iDFS all | 5.2 years | 87.7 | 2.5 | 0.73 (0.57–0.92) | After first month similar for both arms | 40% grade 3 diarrhoea | A | 44 |
| Neratinib vs placebo | ExtentNET | Stage 2–3 HER2+ after 12 months trastuzumab, stratified for HR status | iDFS HR+ | 5.2 years | 86.8 | 4.4 | 0.60 (0.43–0.83) | After first month similar for both arms | 40% grade 3 diarrhoea | A | 44 |
| Neratinib vs placebo | ExtentNET | Stage 2–3 HER2+ after 12 months trastuzumab, stratified for HR status | iDFS HR-ve | 5.2 years | 88.8 | 0.1 | 0.95 (0.66–1.35) | After first month similar for both arms | 40% grade 3 diarrhoea | NEB | 44 |
| AC-PH vs AC-PHPz | APHINITY | HER2+, stratified for nodal status | iDFS all | 45.4 months | 93.2 | 0.9 | 0.81 (0.66–1.00) | | | | | | | | B | 41 |
| AC-PH vs AC-PHPz | APHINITY | HER2+, stratified for nodal status | iDFS Node+ | 44.5 months | 90.2 | 1.8 | 0.77 (0.62–0.96) | | | | | | | | Not scorable† | 41 |
| TDM1 vs H | KATHERINE | HER2+ residual disease after neoadjuvant therapy | iDFS | 3 years | 77.0 | 11.3 | 0.50 (0.39–0.64) | | | | | A | 45 |
| TDM1 vs H | NeoSphere | HER2+ (phase 2) | pCR | | | | | | | | | | 29.0% vs 45.8% | | C | 8 |
| H vs LH | NeoALTO | Neoadjuvant | pCR | 3.77 years | 76.0 | 8.0 | 0.78 (0.47–1.28) | 85.0 | 6.0 | 0.62 (0.3–1.25) | 24.7% vs 29.5% vs 51.3% | | | | ESMO-MCBS V.1.1 | Reference | 42 |
| H vs LH | ALTTO | Adjuvant | DFS | 4.5 years | 86.0 | 2.0 | 0.84 (0.70–1.02) | 94.0 | 1.0 | 0.80 (0.62–1.03) | | | | | | |
| H vs LH | PERSEPHONE | HER2+ | DFS (non-inferiority) | 4 years | 89.8 | 89.4 | 1.07 (0.93–1.24) | | | | | | | | | | | |

*HR values in parentheses refer to 95% CI.
†More than three prespecified subgroups violate scoring rules.
A, adriamycin; C, cyclophosphamide; chemo, chemotherapy; DFS, disease-free survival; ESMO-MCBS, The European Society for Medical Oncology Magnitude of Clinical Benefit Scale; H, trastuzumab; HER2+, HER2 overexpressed; HR+, hormone-positive; HR-, hormone-negative; DFS, disease-free survival; L, lapatinib; NEB, no evaluable benefit; OS, overall survival; P, paclitaxel; pCR, pathological complete response; Pz, pertuzumab; RFS, relapse-free survival; T, docetaxel.
**Adjuvant endocrine therapy**

**Tamoxifen**

The addition of 5 years of tamoxifen compared with placebo was graded an A based on increased long-term OS by 6% and 9% at the individual trial level and in the meta-analysis level, respectively (table 6).46 47

**Aromatase Inhibitors**

The aromatase inhibitor studies to score an A were the Intergroup Exemestane (IES) study and the Italian Tamoxifen Anastrozole (ITA) study. The ITA study score was credited based on DFS results alone in the absence of mature OS data.48 Among the five studies with mature OS data, the data in two did not meet significance thresholds49–52 and the OS gain merited scores of B53–55 or C in the other three.56–58 Comparison aromatase inhibitor alone for 5 years with a switch regimen including tamoxifen and an aromatase inhibitor (2.5 years each) were credited on the basis of non-inferiority in OS and reduced toxicity compared with aromatase inhibitor alone (table 6).47

Meta-analysis data resulted in a C score for the use of an aromatase inhibitor alone in the adjuvant setting, and a C when used a part of a switch after tamoxifen.60

In the premenopausal population, the addition of an aromatase inhibitor (with ovarian function suppression) scored a C when compared with tamoxifen with ovarian function suppression, in the combined SOFT-TEXT study,61–63 but it did not score in the ABCSG-12 study.64

**Extended endocrine therapy**

In the MA-17 study of 5 years letrozole or placebo after 5 years tamoxifen, the node-positive subgroup scored A based on DFS criteria.65 66 Other studies of extended aromatase inhibitor failed to demonstrate improvement in OS.67–69 The ATLAS (Adjuvant Tamoxifen: Longer Against Shorter) study of 5 years versus 10 years of adjuvant tamoxifen demonstrated a 2.8% reduction in breast cancer mortality (grade C) (table 8).70

**Ovarian function suppression in premenopausal women**

Three studies were evaluated. Two mature studies did not demonstrate significant OS gain.61 64 71 72 In the SOFT study, a 1.8% OS advantage was observed in the tamoxifen with ovarian function suppression (OPS) arm, scoring a C, and in the subgroup of patients who had received prior chemotherapy the observed gain in OS was 4.3% (grade B) (table 9).63

**Adjuvant bone-modifying agents**

None of the six individual studies demonstrated a survival advantage. A meta-analysis identified a reduction in breast cancer mortality of 1.8% (grade C), largely derived from the benefit observed in postmenopausal subgroup where the benefit was 3.3% (grade B) (table 10).73

**Expert peer review of the generated results**

The scores generated in this field testing were reviewed by the ESMO Breast Cancer Faculty for reasonableness.
| Study | Trial name | Setting | Primary outcome | Median follow-up | DFS control (%) | DFS gain | DFS HR* | OS control | OS gain (%) | OS HR* | ESMO-MCBS V.1.1 | Reference |
|-------|------------|---------|----------------|-----------------|----------------|----------|---------|------------|-------------|--------|-----------------|-----------|
| 5 years vs Tam 5 years | ATAC | Postmenopausal HR+ | DFS | 120 months | 76.0 | 4.30% | 0.86 (0.78–0.95) | 77.5% | 1.0 | 0.95 (0.84–1.06) | NEB | 59-51 |
| L vs Tam (5 years) | BIG 1–98 | Postmenopausal HR+ | DFS | 97.2 months | 72.0 | 4.4% | 0.82 (0.74–0.92) | 81.4% | 4.0 | 0.79 (0.69–0.90) | B | 59-55 |
| L vs Tam→L vs L→Tam | BIG 1–98 | Postmenopausal HR+ | DFS (Tam→L) | 71 months | 87.9 | –1.7 | 1.05 (0.84–1.32) | 93.4% | –1.0 | 1.13 (0.83–1.53) | B | 56-59 |
| L vs Tam→L vs L→Tam | BIG 1–98 | Postmenopausal HR+ | DFS (L→Tam) | 71 months | 87.9 | –0.3% | 0.96 (0.76–1.21) | 93.4% | –0.3 | 0.90 (0.65–1.24) | B | 56-59 |
| Tam 2–3→E 2–3 vs Tam 5 years | IES | Postmenopausal HR+ and unknown | DFS | 55.7 months | – | All 3.3% | 0.76 (0.66–0.88) | 88.0% | 1.3 | 0.85 (0.71–1.02) | B | 56-57 |
| Tam 2–3→E 2–3 vs Tam 5 years | IES | Postmenopausal HR+ | DFS | 55.7 months | – | HR+ 3.5% | 0.75 (0.65–0.87) | 87.9% | 1.8 | 0.83 (0.69–0.99) | A | 56-57 |
| Tam 2→A 3 years vs Tam 5 year | ARNO-95 | Tam 2→A 3 years vs Tam 5 years | DFS | 30.1 months | 89.3 | 4.2% | 0.66 (0.44–1.00) | p=0.049 | 94.3% | 2.6 | 0.53 (0.28–0.99) | C | 58 |
| E vs Tam→E | TEAM | E 5 years vs Tam 2–3 years →E 5 years | DFS | 5.1 years | 85.0** | 1.0% | 0.97 (0.88–1.08) | 91.0% | 0 | 1.00 (0.89–1.14) | B | 52 |
| Tam vs Tam→A | ITA | Postmenopausal HR+ Node+ | DFS | 36 months | 85.8 | 8.8% | 0.35 (0.18–0.68) | | | | A | 49 |
| 5 years Tam vs 5 years AI (meta-analysis) | EBCTCG | Postmenopausal HR+ | DFS and OS | 10 years | 22.7 | 3.6% | 0.80 (0.73–0.88) | Breast cancer mortality 14.2% | 2.1 | 0.85 (0.75–0.96) | C | 59 |
| 5 years Tam vs 5 years AI (meta-analysis) | EBCTCG | Postmenopausal HR+ | DFS and OS | 10 years | 22.7 | 3.6% | 0.80 (0.73–0.88) | Overall mortality 24% | 2.7 | 0.89 (0.8–0.97) | C | 59 |
| 5 years Tam vs Tam→AI (meta-analysis) | EBCTCG | Postmenopausal HR+ | DFS and OS | 10 years | 19.0 | 2.0% | 0.82 (0.75–0.91) | Breast cancer mortality 10% | 1.5 | 0.84 (0.72–0.96) | C | 59 |
| 5 years Tam vs Tam→AI (meta-analysis) | EBCTCG | Postmenopausal HR+ | DFS and OS | 10 years | 19.0 | 2.0% | 0.82 (0.75–0.91) | Overall mortality 17.5% | 2.9 | 0.82 (0.73–0.91) | C | 59 |
| Tam→AI vs upfront AI (meta-analysis) | EBCTCG | Postmenopausal HR+ | DFS and OS | 7 years | 14.5 | 0.7% | 0.9 (0.81–0.99) | Breast cancer mortality 9.3% | 1.1 | 0.89 (0.78–1.03) | NEB | 59 |
| Tam→AI vs upfront AI (meta-analysis) | EBCTCG | Postmenopausal HR+ | DFS and OS | 7 years | 14.5 | 0.7% | 0.9 (0.81–0.99) | Overall mortality 14.5% | 0.9 | 0.96 (0.86–1.07) | NEB | 59 |

Chart blanks—relevant variables not available in manuscript.
*HR values in parentheses refer to 95% CI.
A, anastrozole; AI, aromatase inhibitor; DFS, disease-free survival; E, exemestane; ESMO-MCBS, The European Society for Medical Oncology-Magnitude of Clinical Benefit Scale; HR+, hormone receptor-positive; L, letrozole; NEB, no evaluable benefit; Node+, node-positive; OS, overall survival; Tam, tamoxifen.
### Table 8: Extended endocrine therapy

| Study | Trial name | Setting | Primary outcome | Median follow-up | DFS control (%) | DFS gain (%) | DFS HR* | OS control (%) | OS gain (%) | OS HR* | ESMO-MCBS V.1.1 | Reference |
|-------|------------|---------|-----------------|------------------|-----------------|--------------|---------|----------------|-------------|---------|-----------------|-----------|
| Letrozole 5 years vs placebo | MA-17 | Postmenopausal HR+ after 5 years tamoxifen (all) | DFS | 30 months | 89.8% | 4.6 | 0.58 (0.45–0.76) | 95.0% | 0.4 | 0.82 (0.57–1.19) | A | 65 66 |
| Letrozole 5 years vs placebo | MA-17 | Postmenopausal HR+ after 5 years tamoxifen (Node+) | DFS | 30 months | 91.0% | 4.0 | 0.66 (0.48–0.91) | 94.0% | 1.0 | 0.97 (0.73–1.28) | A | 67 |
| Letrozole 5 years vs placebo | MA-17R | Postmenopausal HR+ after 5 years tamoxifen | DFS | 6.3 years | 91.0% | 4.0 | 0.66 (0.48–0.91) | 94.0% | 1.0 | 0.97 (0.73–1.28) | NEB | 67 |
| Anastrozole for 3 years vs placebo | ABCSG6a | Postmenopausal HR+ after 5 years tamoxifen | DFS | >5 years | 7.1% | 4.7 | 0.62 (0.40–0.96) | 88.3% | 1.4 | 0.98 (0.59–1.34) | NEB | 68 |
| Exemestane vs placebo | NSABP-B33 | Postmenopausal HR+ after 5 years tamoxifen | DFS | 30 months | 89.0% | 2.0 | 0.68 (p<0.07) | NEB | 69 |
| Tamoxifen | ATLAS | Postmenopausal HR+ after 5 years tamoxifen | BC recurrence and BC mortality | 7.6 years | 7.6% | 3.7 | 0.75 (0.62–0.90) | BC mortality 15% | 0.71 (0.58–0.88) | C | 70 |

*Table 8: Extended endocrine therapy*

| Study | Trial name | Setting | Primary outcome | Median follow-up (months) | DFS control (%) | DFS gain (%) | DFS HR* | OS control (%) | OS gain (%) | OS HR* | ESMO-MCBS V.1.1 | Reference |
|-------|------------|---------|-----------------|-----------------|-----------------|--------------|---------|----------------|-------------|---------|-----------------|-----------|
| Exemestane+OFS vs Tam+OFS | SOFT-TEXT | All | DFS | 96 | 82.8 | 4.0 | 0.77 (0.67–0.90) | 91.8 | 2.1 | 0.80 (0.66–0.96) | C | 62 88 |
| Tam vs Tam+OFS vs exemestane+OFS | SOFT | All | DFS Tam+OFS | 96 | 78.9 | 4.3 | 0.76 (0.62–0.93) | 91.5 | 1.8 | 0.67 (0.48–0.92) | C | 62 87 |
| Tam vs Tam+OFS vs exemestane+OFS | SOFT | All | DFS E+OFS | 96 | 78.9 | 7.0% | 0.65 (0.53–0.81) | 91.5 | 0.6 | 0.85 (0.62–1.15) | NEB | 62 87 |
| Tam vs Tam+OFS vs exemestane+OFS | SOFT | No chemo | DFS Tam+OFS | 96 | 87.4 | 3.2 | 0.76 (0.52–1.12) | 91.5 | 1.6 | 0.74 (0.51–1.09) | NEB | 62 87 |
| Tam vs Tam+OFS vs exemestane+OFS | SOFT | No chemo | DFS E+OFS | 96 | 87.4 | 5.2 | 0.58 (0.38–0.88) | NEB | 62 87 |
| Tam vs Tam+OFS vs exemestane+OFS | SOFT | Past- chemo | DFS Tam+OFS | 96 | 71.4 | 5.3 | 0.76 (0.60–0.97) | 85.1 | 4.3 | 0.59 (0.42–0.84) | B | 62 87 |
| Tam vs Tam+OFS vs exemestane+OFS | SOFT | Past- chemo | DFS E+OFS | 96 | 71.4 | 9.0 | 0.82 (0.64–1.07) | 85.1 | 2.1 | 0.79 (0.57–1.09) | NEB | 62 87 |
| Anastrozole+OFS vs Tam+OFS | ABCSG-12 | Premenopausal HR+ | DFS | 94.4 | NA | – | 1.08 (0.81–1.44) | 96.3 | –2.1 | 1.63 (1.05–2.52) | NEB | 64 71 72 |

*Table 9: Ovarian function suppression in premenopausal women*
Table 10  Adjuvant bone-modifying agents

| Study | Trial name | Setting | Primary outcome | Median follow-up (months) | DFS control (%) | DFS gain (%) | DFS HR* (95% CI) | OS control (%) | OS gain (%) | OS HR* (95% CI) | ESMO-MCBS V.1.1 Reference |
|-------|------------|---------|-----------------|---------------------------|-----------------|--------------|-----------------|----------------|-------------|-----------------|--------------------------|
| Clodronate vs placebo | NSABP-B-34 | Adjuvant clodronate | DFS | 90.7 | NA | Nil | 0.91 (0.78–1.07) | NA | Nil | 0.84 (0.65–1.05) | NEB 88 |
| Ibandronate vs placebo | GAIN | HR+ Node+ | DFS | 38.7 | NA | Nil | 0.94 (0.77–1.16) | NA | Nil | 0.96 (0.71–1.31) | NEB 89 |
| Denosumab vs placebo | ABCSG-18 | Postmenopausal women on AI | Time-to-first clinical fracture | NA | NA | NA | 79.3% | 3.6 | 0.77 (0.56–1.00) | NS |
| Clodronate vs placebo | NSABP-B-34 | Adjuvant clodronate | Time-to-first bone metastases | 5.6 years | NA | NA | 79.3% | 3.6 | 0.77 (0.56–1.00) | NS |
| Zoledronate vs placebo | ABCSG-12 | Premenopausal with OFS | DFS | 94.4 months | 3.40 | 0.77 (0.60–0.99) | 94.5% | 2.2 | 0.66 (0.43–1.02) | NEB 64 71 72 |
| Zoledronate vs placebo | AZURE/BIG01-04 | DFS (all patients) | 84 months | NA | 0.94 (0.82–1.06) | NA | 0.93 (0.81–1.07) | NEB 93 94 |
| Zoledronate vs placebo | AZURE/BIG01-04 | DFS 5 years+menopausal at diagnosis | NA | – | 0.77 (0.63–0.96) | NA | – | 0.81 (0.63–1.04) | NEB 93 94 |
| Adjuvant bisphosphonate (meta-analysis) | EBCTCG | With hormonal therapy | DFS and OS | 5.6 years | All | Breast cancer mortality 18.4% | 1.8 | 0.91 (0.83–0.99) | C 60 |
| Adjuvant bisphosphonate (meta-analysis) | EBCTCG | With hormonal therapy | DFS and OS | 5.6 years | All | All-cause mortality 22.3% | 1.5 | 0.92 (0.85–1.00) | p<0.06 | NEB 60 |
| Adjuvant bisphosphonate (meta-analysis) | EBCTCG | With hormonal therapy | DFS and OS | 5.6 years | Postmenopausal | Breast cancer mortality 18% | 3.3 | 0.82 (0.73–0.93) | B 60 |
| Adjuvant bisphosphonate (meta-analysis) | EBCTCG | With hormonal therapy | DFS and OS | 5.6 woman years | Premenopausal | Breast cancer mortality 20.7% | –0.1 | – | – | NEB 60 |

*HR values in parentheses refer to 95% CI.
AI, aromatase inhibitors; DFS, disease-free survival; ESMO-MCBS, The European Society for Medical Oncology-Magnitude of Clinical Benefit Scale; HR+, hormone receptor-positive; NA, not applicable; NEB, No evaluable benefit; Node+, node-positive; NS, not statistically significant; OFS, ovarian function suppression; OS, overall survival; SNA, scoring not applicable.
Apart from the scores for double HER2 blockade, the derived scores were more commonly endorsed as reasonable than unreasonable. There was no consensus about the grading for double HER2 blockade (unreasonable 32%; reasonable 29%): many respondents expounded that the scores for the APHINITY and ExteNET studies, derived from the relative benefit gain in DFS but with very small absolute benefit, were excessively high. In situations when the primary outcome of the study was DFS, and a robust DFS benefit was observed (in terms of both relative and absolute benefits) but without significant OS benefit, a proportion of reviewers expressed that a grade of NEB under represented the clinical value of prolonged interim time without disease, treatment and toxicity.

**DISCUSSION**

The validity of the ESMO-MCBS is predicated on adherence to the public policy ethical standard of ‘accountability for reasonableness’ and the field testing of the scale over a large range of clinical trials is an important part of the development process. This study, applying the ESMO-MCBS V.1.1 to 59 individual trials and 6 meta-analyses, has demonstrated that form 1 of the ESMO-MCBS can be applied to systemic adjuvant therapy trials. Moreover, apart from a few specific exceptions, the generated grades were considered reasonable by experts in the ESMO Breast Cancer Faculty, largely reflecting standard clinical practice.

Applying the scale and interpreting the results was, in most instances, straightforward. A small number of studies did not incorporate all critical data in accordance with CONSORT standards. In some instances HRs were published without CIs, some meta-analyses did not include absolute gain data for OS and some studies report the HR to reflect increased recurrence risk (eg, MA-21). Furthermore, even with long-term follow-up, some studies never published follow-up of their mature survival data. Since magnitude of benefit grades derived from OS gain at maturity is often less than that derived from DFS, the non-publication of mature OS results occasionally resulted in disproportionally high scores in some studies. This is well illustrated in two examples: no mature survival data were ever published for the ITA study by Boccardo et al which evaluated switching from tamoxifen to an aromatase inhibitor and the MA21 study that evaluated the addition of paclitaxel to an anthracycline. Consequently, these were among the few studies in their respective classes to score an A, while all others for which mature survival data were available scored C or NEB. We note that this anomaly could be misinterpreted to suggest superiority, or even manipulated with delays or even non-reporting of mature OS data to avoid downgrading.

We note that the ESMO-MCBS is agnostic to DFS type and does not distinguish between DFS, invasive DFS (iDFS) and distant DFS (DDFS) that is also called ‘distant metastasis-free survival’. In recent years, there has been a shift to more accurate end points such as invasive iDFS or DDFS, which are better surrogates for OS benefit, since they emphasise events that are more closely related to cancer mortality (ie, invasive relapse or distant metastases). This underscores the importance of new initiatives to introduce standardisation in the definitions and application of these end points.

A key aim of this study was to identify shortcomings in the current version of form 1 which will be addressed in future versions of the scale. This field testing and peer-review process identified six shortcomings in form 1. All of these shortcomings have been reviewed by the ESMO-MCBS Working Group and initiatives are underway to address each of them as part of the forthcoming revisions to be incorporated in the next version of the scale (V.2.0).

1. **HR thresholds for DFS are excessively lenient.** The experience of this field testing indicates that trials initially graded on the basis of DFS in initial publications, commonly attained lower scores when mature OS data were available and that in many cases the OS gains were not significant. This indicates that the relative benefit thresholds for grade B and C (lower limit of the 95% CI of the HR 0.65–0.85 and >0.85, respectively) are excessively lenient. Consequently, we recommend lowering of the HR thresholds for grades B and C.

2. **Lack of absolute gain constraint on DFS scoring can generate inappropriately high scores when absolute gain is very small.** Expert peer reviewers concerned that grades accrued on the basis of relative benefit when the observed absolute benefit is very small were unreasonably high. This was highlighted in their critique of scores generated in the APHINITY and ExteNET trials. This could be corrected by applying the ‘dual rule’ whereby grade criteria include both relative and absolute benefit thresholds in a manner that is consistent with all other forms of the ESMO-MCBS V.1.1.

3. **The clinical benefit derived from DFS gain is not credited when OS gain is not verified.** In many instances, gains derived from DFS were not credited when there was no significant gain in mature OS. When substantially improved DFS does not result in improved OS, the grading of NEB undervalued the time gained without need for medical treatment, which may itself be a valued outcome independent of OS.

4. **Need to define OS maturity in adjuvant studies.** According to the ESMO-MCBS V.1.1, surrogate scores prevail if mature OS data are not yet available. Maturity is generally defined as the time point where most of the anticipated events will have occurred. In a non-curative setting, when all patients are expected to die, conventionally it is when the median survival of both arms is reached. However, in the adjuvant setting, when the number of anticipated events may vary according to the tumour type and stage, this convention does not apply. Consequently, evaluating maturity of survival data in this setting requires familiarity with the specific clinical scenario and it is conceivable that in some instances this may be source of reasonable disagreement even between experts. ESMO-MCBS instructions for use
should include guidelines for OS maturity. For example, 5 years for subtypes at high risk for earlier recurrence (such as triple negative and HER2-positive/endocrine unresponsive subtypes) and at least 8 years for endocrine responsive tumours (including HER2-positive/endocrine-responsive). 57

5. Lack of capacity to grade single-arm de-escalation studies in the curative setting. A recent single-arm phase 2 study reported excellent outcomes for node-negative HER2-positive breast cancers smaller than 2 cm treated with the combination of paclitaxel and trastuzumab (without an anthracycline). 58 These type of studies are often used to evaluate de-escalation strategies. Form 1 is unable to grade these studies.

6. Lack of consideration of toxicity in the curative setting. The current version of form 1 does not consider toxicity. The shortcoming of this approach is illustrated by the ExteNET study that scores an ‘A’ for the hormone-positive subgroup despite very substantial toxicity secondary to the neratinib, which resulted in a 27.6% discontinuation rate. 59 While we appreciate that patients may be willing to make short-term toxicity trade-off to improve cure rate, it is not clear that this approach applies also for long-term toxicity such as peripheral neuropathy or secondary cancers (especially when improvement in cure rate may be small). We support the proposition, initially made by patient advocacy groups, that ESMO-MCBS scores in form 1 should be annotated to indicate acute and/or long-term toxicities.

CONCLUSIONS

In a time of exponential growth in the costs of cancer care, tools to assist physicians and regulatory bodies in evaluating new therapeutic options are critical. This study reinforces the validity of the ESMO-MCBS approach to adjuvant therapies insofar as the scoring of adjuvant approaches in early breast cancer largely reflects standard clinical practice. This field testing has identified six shortcomings that have been reviewed by the ESMO-MCBS Working Group and that form the foundation for amendments to be incorporated into future iterations of the ESMO-MCBS.

Author affiliations

1. Oncology, Shaare Zedek Medical Center, Jerusalem, Israel
2. Department of Medical Oncology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
3. National and Kapodistrian University of Athens, Athens, Greece
4. Frontier Science Foundation-Hellas, Athens, Greece
5. Université Libre de Bruxelles, Institut Jules Bordet, Bruxelles, Belgium
6. ESMO-MCBS Working Group, European Society for Medical Oncology, Viganello, Switzerland
7. Breast Unit, Champalimaud Clinical Center/Champalimaud Foundation, Lisbon, Portugal

Acknowledgements The authors wish to thank and acknowledge our oncology colleagues listed who participated in the field testing for ‘reasonable’ scorings using form 1 of the ESMO-MCBS and who have agreed to place their names in this publication (online supplementary appendix 1). We also thank those who wished to remain anonymous.

Contributors Conception of the work: all authors. Funding acquisition: not applicable. Data collection and data analysis: all authors. Manuscript writing/editing: all authors. Final approval: all authors.

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

Competing interests SP-S reports institutional financial support for her advisory role from AstraZeneca, Pfizer, Novartis, Roche, Teva, NanoString, EMDed reports institutional financial support for her advisory role from Daiichi Sankyo, Merck, NSABP, Pfizer, Sanofi, Synthon and institutional financial support for clinical trials or contracted research from Amgen, AstraZeneca, Bayer, Chugai Pharma, CytoMx Therapeutics, G1 Therapeutics, Genentech, Nordic Nanovector, Radius Health, Regeneron, Roche, Synthon; MJF reports scientific board member for Oncylitics, consultant honoraria from AstraZeneca, Cameli-IDS, Crescendo Biologics, Debiopharm, G1 Therapeutics, Genentech, Huya, Immunomedics, Lilly, Menarini, MSD, Novartis, Odonate, Periphigen, Pfizer, Roche, Seattle Genetics, research grants to institute AstraZeneca, Lilly, MSD, Novartis, Pfizer, Radius, Roche-Genentech, Servier, Synthon; FC reports institutional financial support for her advisory role from Astellas/Medivation, AstraZeneca, Célgene, Daiichi-Sankyo, Eisai, GE Oncology, Genentech, GlaxoSmithKline (GSK), Merck-Sharp, Merus BV, Novartis, Pfizer, Pierre-Fabre, Roche, Sanofi, Teva.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. All data freely available.

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

ORCID iD

Shani Paluch-Shimon http://orcid.org/0000-0002-6677-2199

REFERENCES

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer 2013;49:1374–403.
2. Ferlay J et al. GLOBOCAN 2008 v1.2, cancer incidence and mortality worldwide: IARC CancerBase No. 10. Lyon, France: International Agency for Research on Cancer, 2010.
3. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the global burden of disease study 2015. Lancet 2016;388:1659–72.
4. Mariotto AB, Yabroff KR, Shao Y, et al. Projections of the cost of cancer care in the United States: 2010-2020. J Natl Cancer Inst 2011;103:117–28.
5. Chen M, Sullivan R, Dafni U, et al. A standardised, generic, validated framework to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for medical oncology magnitude of clinical benefit scale (ESMO-MCBS). Annals of Oncology 2015;26:1547–73.
6. Schnipper LE, Davidson NE, Wollins DS, et al. American Society of clinical oncology statement: a conceptual framework to assess the value of cancer treatment options. J Clin Oncol 2015;33:2563–77.
7. Chen NY, Dafni U, Banco J, et al. ESMO-Magnitude of clinical benefit scale version 1.1. Annals of Oncology 2017;28:2340–66.
8. Gianni L, Pienkowski T, Im Y-H, et al. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. Lancet Oncol 2012;13:25–32.
9. Cardoso F, Kyriakides O, Oheo G, et al. Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2019;30:1674.
10. Tell MA, Gradishar WJ, Ward JH. NCCN guidelines updates: breast cancer. J Natl Compr Canc Netw 2019;17:552–5.
11. Bear HD, Tang G, Rastogi P, et al. Neoadjuvant plus adjuvant bevacizumab in early breast cancer (NSABP B-40 [NRG Oncology]):
secondary outcomes of a phase 3, randomised controlled trial. *Lancet Oncol* 2015;16:1037–48.
12. Bonilla L, Ben-Aharon I, Vidal L, et al. Dose-Dense chemotherapy in nonmetastatic breast cancer: a systematic review and meta-analysis of randomized controlled trials. *J Natl Cancer Inst* 2010;102:1845–54.
13. Palmieri G, Jones A. The 2011 EBCCTG polychemotherapy overview. *Lancet* 2012;379:390–2.
14. Hutchins LF, Green SJ, Ravdin PM, et al. Randomized, controlled trial of cyclophosphamide, methotrexate, and fluorouracil versus cyclophosphamide, doxorubicin, and fluorouracil with and without tamoxifen for high-risk, node-negative breast cancer: treatment results of intergroup protocol INT-0102. *J Clin Oncol* 2005;23:8313–21.
15. Levine MN, Pritchard KI, Bramwell VHC, et al. Randomized trial comparing cyclophosphamide, epirubicin, and fluorouracil with cyclophosphamide, doxorubicin, and fluorouracil followed by paclitaxel for early breast cancer: phase III trial. *J Natl Cancer Inst* 2005;97:177–87.
16. Jones SE, Savin MA, Holmes FA, et al. Phase III trial comparing doxorubicin plus cyclophosphamide as adjuvant chemotherapy for node-negative breast cancer: results of NSABP B-28. *J Clin Oncol* 2005;23:3868–96.
17. Bunnell M, Levine MN, Chapman J-AW, et al. Cyclophosphamide, epirubicin, and fluorouracil versus dose-dense epirubicin and cyclophosphamide followed by paclitaxel versus doxorubicin and cyclophosphamide followed by paclitaxel in node-positive or high-risk node-negative breast cancer. *J Clin Oncol* 2010;28:77–82.
18. Mamoouas EP, Bryant J, Lembersky B, et al. Paclitaxel after doxorubicin plus cyclophosphamide as adjuvant chemotherapy for node-positive breast cancer: update of national cancer Institute of Canada clinical Trials Group trial MA5. *J Clin Oncol* 2005;23:5166–70.
19. Martin M, Segui MA, Antoné A, et al. Adjuvant docetaxel for high-risk, node-negative breast cancer. *N Engl J Med* 2010;363:2200–10.
20. Martin M, Rodriguez-Lescure A, Ruiz A, et al. Randomized phase 3 trial of fluorouracil, epirubicin, and cyclophosphamide alone or followed by paclitaxel for early breast cancer. *J Natl Cancer Inst* 2006;98:805–11.
21. Jones S, Holmes FA, O’Shaughnessy J, et al. Docetaxel with cyclophosphamide is associated with a 7-year overall survival benefit compared with doxorubicin and cyclophosphamide: 7-year follow-up of US oncology research trial 9735. *J Clin Oncol* 2009;27:1177–83.
22. Blum JL, Flynn PJ, Youthers G, et al. Anthracyclines in early breast cancer: the ABC Trials-USOR 06-090, NSABP B-31 and NCCTG N9831. *J Clin Oncol* 2017;35:2647–55.
23. Citron ML, Berry DA, Cirrincione C, et al. Randomized trial of dose-dense versus conventionally scheduled and sequential versus concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: first report of intergroup trial I/USOR 07132, Cancer and leukemia group B trial 9741. *J Clin Oncol* 2003;21:1431–9.
24. Moebus V, Jackisch C, Lueck H-J, et al. Intense dose-dense sequential chemotherapy with epirubicin, paclitaxel, and cyclophosphamide compared with conventionally scheduled chemotherapy in early primary breast cancer: mature results of an age phase III study. *J Clin Oncol* 2010;28:2874–80.
25. Del Mastro L, De Placido S, Bruzzi P, et al. Fluorouracil and doxorubicin and cyclophosphamide chemotherapy in early breast cancer: results of the ATAC (Arimidex, tamoxifen alone or in combination) trial efficacy and safety update analyses. *J Natl Cancer Inst* 2004;96:1832–31.
26. Rea G. Abstract GS1-01: increasing the dose density of adjuvant chemotherapy by shortening intervals between courses or by sequential drug administration significantly reduces both disease recurrence and breast cancer mortality: an EBCCTG meta-analysis of 21,000 women with node-positive breast cancer: interim results. *J Clin Oncol* 2015;33:863–72.
27. Petrelli F, Cabiddo M, Coinu A, et al. Adjuvant dose-dense chemotherapy in breast cancer: a systematic review and meta-analysis of randomized trials. *Breast Cancer Res Treat* 2015;151:251–9.
28. Masuda N, Lee S-J, Ohtani S, et al. Adjuvant capcitabine for breast cancer after preoperative chemotherapy. *N Engl J Med* 2017;376:249–59.
29. von Minckwitz G, Schneeweis A, Lobli S, et al. Neoadjuvant carboplatin in patients with triple-negative and HER2-positive early breast cancer (GeparSerocto: GGBG tide-66): a randomised phase 2 trial. *Lancet Oncol* 2014;15:747–56.
30. Lobli S, O’Shaughnessy J, Untch M, et al. Addition of the PARP inhibitor veliparib plus carboplatin or carboplatin alone to neoadjuvant chemotherapy in triple-negative breast cancer (brightness): a randomised, phase 3 trial. *Lancet Oncol* 2018;19:497–509.
van de Velde CJH, Rea D, Seynaeve C, et al. Adjuvant tamoxifen and exemestane in early breast cancer (team): a randomised phase 3 trial. Lancet 2011;377:321–31.

Thürlimann B, Kershaw A, et al. Breast International Group (BIG) 1-98 Collaborative Group. A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. N Engl J Med 2005;353:2747–57.

Coates AS, Kershaw A, Thürlimann B, et al. Five years of letrozole compared with tamoxifen as initial adjuvant therapy for postmenopausal women with endocrine-responsive early breast cancer: update of the BIG 1-98 trial. J Clin Oncol 2007;25:486–92.

Regan MM, Neven P, Giobbie-Hurder A, et al. Assessment of letrozole and tamoxifen alone and in sequence for postmenopausal women with steroid hormone receptor-positive breast cancer: the big 1-98 randomised trial at 8 years median follow-up. Lancet Oncol 2011;12:1101–8.

Coombs RC, Hall E, Gibson LJ, et al. A randomised trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer. N Engl J Med 2004;350:1081–92.

Coombs RC, Kilburn LS, Snowden CF, et al. Survival and safety of exemestane versus tamoxifen after 2-3 years’ tamoxifen treatment (intergroup Exemestane study): a randomised controlled trial. Lancet 2007;369:599–70.

Kaufmann M, Jonat W, Hilfrich J, et al. Improved overall survival in postmenopausal women with early breast cancer after anastrozole initiated after treatment with tamoxifen compared with continued tamoxifen: the ARNO 95 study. J Clin Oncol 2007;25:2664–70.

Mouridsen H, Giobbie-Hurder A, et al. BIG 1-98 Collaborative Group. Letrozole therapy alone or in sequence with tamoxifen in women with breast cancer. N Engl J Med 2009;361:766–76.

Early Breast Cancer Trials’ Collaborative Group (EBCTCG). Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. Lancet 2015;386:1341–52.

Pagani O, Regan MM, Francis PA, et al. Exemestane with ovarian suppression in premenopausal breast cancer. N Engl J Med 2014;371:107–18.

Francis PA, Regan MM, Fleming GF, et al. Tailoring adjuvant endocrine therapy for premenopausal breast cancer. N Engl J Med 2018;379:122–37.

Francis PA, Regan MM, Fleming GF, et al. Adjuvant ovarian suppression in premenopausal breast cancer. N Engl J Med 2015;372:436–46.

Gnant M, Millettich B, Stoeger H, et al. Zoledronic acid combined with adjuvant endocrine therapy of tamoxifen versus anastrozole plus ovarian function suppression in premenopausal early breast cancer: final analysis of the Austrian breast and colorectal cancer Study Group trial 12. Ann Oncol 2015;26:313–20.

Goss PE, Ingle JN, Martino S, et al. Randomized trial of letrozole following tamoxifen as extended adjuvant therapy in receptor-positive breast cancer: updated findings from NCIC CTG MA.17. J Natl Cancer Inst 2005;97:1262–71.

Goss PE, Ingle JN, Pritchard KI, et al. Extending aromatase-inhibitor therapy to 10 years versus stopping at 5 years after adjuvant tamoxifen to 10 years versus stopping at 5 years after adjuvant exemestane as extended adjuvant therapy after 5 years of adjuvant endocrine therapy for premenopausal breast cancer. Lancet 2015;386:1353–61.

Hudis CA, Barlow WE, Costantino JP, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the steep system. J Clin Oncol 2007;25:2127–32.

Gourgou-Bourgade S, Cameron D, Poortmans P, et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECEN initiative (definition for the assessment of Time-to-event endpoints in cancer trials). Annals of Oncology 2015;26:2637–40.

Robertson AG, Booth CM, Eisenhauer EA. Disease-Free survival as an end-point in the treatment of solid tumours – perspectives from clinical trials and clinical practice. Eur J Cancer 2014;50:2298–302.

Pan H, Gray R, Braybrooke J, et al. 20-Year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. N Engl J Med 2017;377:1836–46.

Tolansky SM, Barry WT, Dang CT, et al. Adjuvant paclitaxel and trastuzumab for node-negative, HER2-positive breast cancer. N Engl J Med 2015;372:134–44.

Spicceri JA, Wang M, Martino S, et al. Weekly paclitaxel in the adjuvant treatment of breast cancer. N Engl J Med 2008;358:1663–71.

Sparano JA, Zhao F, Martino S, et al. Long-Term follow-up of the E1199 phase III trial evaluating the role of taxane and schedule in operable breast cancer. J Clin Oncol 2015;33:2353–60.

Eiermann W, Pienkowski T, Crow J, et al. Phase III study of doxorubicin/cyclophosphamide with concomitant versus sequential docetaxel as adjuvant treatment in patients with human epidermal growth factor receptor 2-normal, node-positive breast cancer. JCO 2001;29:2367–75.

Swain SM, Tang G, Geyer CE, et al. Definitive results of a phase III adjuvant trial comparing three chemotherapy regimens in women with operable, node-positive breast cancer: the NSABP B-38 trial. J Clin Oncol 2013;31:197–204.

von Minckwitz G, Lobi S, Untch M, et al. Survival after neoadjuvant chemotherapy with or without bevacizumab or everolimus for HER2-negative primary breast cancer (BGIG 44-GeparQuinto). Ann Oncol 2014;25:2363–72.

O’Sullivan CC, Bradbury I, Campbell C, et al. Efficacy of adjuvant trastuzumab for patients with human epidermal growth factor receptor 2-positive early breast cancer and tumours ≤ 2 cm: a meta-analysis of the randomized trastuzumab trials. J Clin Oncol 2015;33:2600–8.

Piccart-Gebhart MJ, Holmes E, Baselga J, et al. Adjuvant lapatinib and trastuzumab for early human epidermal growth factor receptor 2-positive breast cancer: results from the randomized phase III trial of lapatinib and/or trastuzumab treatment optimization trial. J Clin Oncol 2016;34:1034–42.

Pagani O, Regan MM, Walley BA, et al. Adjuvant exemestane with ovarian suppression in premenopausal breast cancer. N Engl J Med 2014;371:107–18.

Francis PA, Regan MM, Fleming GF, et al. Adjuvant ovarian suppression in premenopausal breast cancer. N Engl J Med 2015;372:436–46.

Paterson AHG, Anderson SJ, Lembersky BC, et al. Oral cladronate for adjuvant treatment of operable breast cancer (national surgical adjuvant breast and bowel project protocol B-34): a multicentre, placebo-controlled, randomised trial. Lancet Oncol 2012;13:734–42.

von Minckwitz G, Möbus V, Schneeweiss A, et al. German adjuvant intergroup node-positive study: a phase III trial to compare oral ibandronate versus observation in patients with high-risk early breast cancer. J Clin Oncol 2013;31:3531–9.

Gnant M, Pfölker G, Dubsky PC, et al. Phase III trial of docetaxel as adjuvant treatment in patients with human epidermal growth factor receptor 2-positive early breast cancer and tumors ≤ 2 cm: a meta-analysis of the randomized trastuzumab trials. J Clin Oncol 2015;33:2600–8.

Piccart-Gebhart MJ, Holmes E, Baselga J, et al. Adjuvant lapatinib and trastuzumab for early human epidermal growth factor receptor 2-positive breast cancer: results from the randomized phase III trial of lapatinib and/or trastuzumab treatment optimization trial. J Clin Oncol 2016;34:1034–42.

Pagani O, Regan MM, Walley BA, et al. Adjuvant exemestane with ovarian suppression in premenopausal breast cancer. N Engl J Med 2014;371:107–18.

Francis PA, Regan MM, Fleming GF, et al. Adjuvant ovarian suppression in premenopausal breast cancer. N Engl J Med 2015;372:436–46.

Paterson AHG, Anderson SJ, Lembersky BC, et al. Oral cladronate for adjuvant treatment of operable breast cancer (national surgical adjuvant breast and bowel project protocol B-34): a multicentre, placebo-controlled, randomised trial. Lancet Oncol 2012;13:734–42.

von Minckwitz G, Möbus V, Schneeweiss A, et al. German adjuvant intergroup node-positive study: a phase III trial to compare oral ibandronate versus observation in patients with high-risk early breast cancer. J Clin Oncol 2013;31:3531–9.

Gnant M, Pfölker G, Dubsky PC, et al. Adjuvant denosumab in breast cancer (ABCSG-18): a multicentre, randomised, double-blind, placebo-controlled trial. Lancet 2013;386:433–43.

Powles T, Paterson S, Kanis JA, et al. Randomized, placebo-controlled trial of clodronate in patients with primary operable breast cancer. J Clin Oncol 2012;30:3219–24.

Powles T, Paterson S, McCloskey E, et al. Reduction in bone resorption and improved survival with oral cladronate for adjuvant treatment of operable breast cancer [ISRCTN83688026]. Breast Cancer Res Treat 2006;9:R13.

Coleman RE, Marshall H, Cameron D, et al. Breast-Cancer adjuvant therapy with zoledronic acid. N Engl J Med 2011;365:1396–405.

Coleman R, Cameron D, Dodwell D, et al. Adjuvant zoledronic acid in patients with early breast cancer: final efficacy analysis of the azure (big 01/04) randomised open-label phase 3 trial. Lancet Oncol 2014;15:997–1006.