A standardised, generic, validated approach 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)

Nathan I Cherny, MBBS, FRACP, FRCP
Cancer Pain and Palliative Medicine Service, Department of Medical Oncology, ShaareZedek Medical Center, Jerusalem, Israel

Richard Sullivan, MD, PhD
Kings Health Partners Integrated Cancer Centre
King’s College London, Institute of Cancer Policy
London, United Kingdom

Urania Dafni, ScD
Laboratory of Biostatistics, Division of Public Health, Department of Nursing, University of Athens, Greece.

J. Martijn Kerst, MD, PhD
Department of Medical Oncology Antoni van Leeuwenhoek Hospital/
Netherlands Cancer Institute, Amsterdam, The Netherlands

Correspondence to
Nathan I. Cherny,
Norman Levan Chair of Humanistic Medicine
Associate Professor of Medicine (BGU)
Director, Cancer Pain and Palliative Medicine Service
Department of Medical Oncology
ShaareZedek Medical Center
Jerusalem, Israel 91031
E mail: mcbs@esmo.org

Alberto Sobrero, MD
Department of Medical Oncology
IRCCS San Martino IST
Genova, Italy

Christoph Zielinski, MD
Division of Oncology
Medical University Vienna
Vienna, Austria.

Elisabeth G.E. de Vries, MD, PhD
Department of Medical Oncology,
University Medical Center Groningen,
University of Groningen,
Groningen, The Netherlands

Martine J. Piccart, MD, PhD
Université Libre de Bruxelles, Jules Bordet Institute,
Brussels, Belgium
abstract

The value of any new therapeutic strategy or treatment is determined by the magnitude of its clinical benefit balanced against its cost. Evidence for clinical benefit from new treatment options is derived from clinical research, in particular phase III randomised trials, which generate unbiased data regarding the efficacy, benefit and safety of new therapeutic approaches. To date there is no standard tool for grading the magnitude of clinical benefit of cancer therapies, which may range from trivial (median progression-free survival advantage of only a few weeks) to substantial (improved long term survival). Indeed, in the absence of a standardised approach for grading the magnitude of clinical benefit, conclusions and recommendations derived from studies are often hotly disputed and very modest incremental advances have often been presented, discussed and promoted as major advances or "breakthroughs". Recognising the importance of presenting clear and unbiased statements regarding the magnitude of the clinical benefit from new therapeutic approaches derived from high quality clinical trials the European Society for Medical Oncology (ESMO) has developed a validated and reproducible tool to assess the magnitude of clinical benefit for cancer medicines, the ESMO Magnitude of Clinical Benefit Scale (ESMO-MCBS). This tool uses a rational, structured and consistent approach to derive a relative ranking of the magnitude of clinically meaningful benefit that can be expected from a new anti-cancer treatment. The ESMO-MCBS is an important first step to the critical public policy issue of value in cancer care, helping to frame the appropriate use of limited public and personal resources to deliver cost effective and affordable cancer care. The ESMO-MCBS will be a dynamic tool and its criteria will be revised on a regular basis.
introduction

The value of any new therapeutic strategy or treatment is determined by the magnitude of its clinical benefit balanced against its cost [1]. Value considerations have become increasingly important in an era of rapid expansion of new, expensive cancer medicines and other technologies such as advanced radiotherapy techniques or robotic surgery which provide small incremental benefits [2-5] within the context of cost-constrained health care systems[6]. This is especially true in Europe where the costs of care delivery [6] and cancer outcomes [7-9] vary substantially across Europe with the latter being influenced by the level of economic development [9, 10]. In some instances discrepant outcomes between countries in Europe can be attributed to inordinate delays, sometimes of years, in making highly effective treatments available at an affordable cost to the patient [11, 12].

Whereas costs of procurement and out of pocket expenditures vary from country to country, the magnitude of clinical benefit, as derived from well-designed clinical trials, is a relative constant. Consequently, meaningful discussion of value and relative value are predicated on an understanding of the magnitude of clinical benefit [1]. Clinical benefit in this context refers to the added benefit compared to a control which, in most cases, is the best current standard care.

Evidence for clinical benefit from new treatment approaches is derived from comparative outcome studies, most commonly phase III randomised clinical trials, which generate ostensibly unbiased data regarding the efficacy, benefit and safety of new therapeutic approaches. The potential benefits of a new treatment can be summarised as either living longer and/or living better, evaluated in clinical studies through the treatment effect on overall survival (OS) and/or quality of life (QoL), and their surrogates (Table 1). In studies of interventions with curative intent in which mature survival data is not yet available disease-free survival (DFS), recurrence-free survival (RFS), event-free survival (EFS), distant disease free survival (DDFS), and time to recurrence (TTR), are used as surrogate measures. The validity of this approach, though not uncontroversial [13], is relatively well supported by data derived from a wide range of solid tumour settings including in colon [14], gastric [15],
lung [16] and breast [17] cancers. In studies evaluating therapies in non-curative settings, progression-free survival (PFS), and time to progression (TTP) provide information about biological activity and may indicate benefit for some patients [18, 19] however they are not reliable surrogates for improved survival [18, 20-23] or QoL [23, 24].

To date there is no standard tool for grading the magnitude of clinical benefit of cancer therapies [25, 26], which may range from trivial (median PFS advantage of only a few weeks) to substantial (improved long term survival). Indeed, in the absence of a standardised approach for grading the magnitude of clinical benefit, conclusions and recommendations derived from studies are often hotly disputed [25] and very modest incremental advances have often been presented, discussed and promoted as major advances or "breakthroughs" [5, 25-29]. Overestimating or overstating the benefits from new intervention can cause harm: It confounds public policy decision making [29], undermines the credibility of oncology research reporting [26, 29, 30], harms patients who choose to undertake treatments based on exaggerated expectations that may subject them to either risk of adverse effects, inconvenience or substantial personal costs [26, 28] and in the public domain they fuel sometimes inappropriate hype or disproportionate expectations about novel treatments [31, 32] and the need to allocate public or personal funds to provide them.

It is important for the Oncology Community to present clear and unbiased statements regarding the magnitude of clinical benefit from new therapeutic approaches supported by credible research. ESMO aims to highlight those treatments which bring substantial improvements to the duration of survival and/or the QoL of cancer patients which need to be distinguished from those whose benefits are more modest, limited or even marginal. To this end, ESMO has undertaken the development of a validated and reproducible tool to assess the magnitude of clinical benefit of anti-cancer interventions, the ESMO Magnitude of Clinical Benefit Scale (ESMO-MCBS). ESMO intends to apply this scale prospectively to each new anti-cancer drug/intervention that will be European Medicines Agency (EMA) approved. Drugs or treatment interventions that obtain the highest scores on the scale will be highlighted in the ESMO guidelines, with the hope that they will be rapidly endorsed by health authorities across the European Union.
background and methodology

An ESMO Task Force to guide the development of the grading scale was established in March 2013. The members of the Task Force co-chaired by Elisabeth de Vries and Martine Piccart, are Richard Sullivan, Nathan Cherny, Urania Dafni, Martijn Kerst, Alberto Sobrero and Christoph Zielinski. A first generation draft scale (ESMO-MCBS v-2) was developed and adapted through a “snowball” method based upon previous work of Task Force members who had independently developed preliminary models of clinical benefit grading. The first generation scale was sent for review by 276 members of the ESMO faculty and a team of 51 expert biostatisticians.

The second generation draft (ESMO-MCBS v-1.0) was formulated based on the feedback from faculty and biostatisticians and the conceptual work of Alberto Sobrero regarding the integration of both hazard ratio (HR), prognosis and absolute differences in data interpretation [33, 34]. The second generation draft was applied in a wide range of contemporary and historical disease settings by members of the ESMO-MCBS Task Force, the ESMO Guidelines Committee and a range of invited experts. Results were scrutinised for face validity, coherence and consistency. Where deficiencies were observed or reported, targeted modifications were implemented and the process of field testing and review was repeated. This process was repeated through 13 redrafts of the scale preceding the current one (ESMO-MCBS v1.0). The final version and fielded testing results were reviewed by selected members of the ESMO faculty and the ESMO Executive Board.

The goal of the ESMO-MCBS evaluation was to assign the highest grade to trials having adequate power for a relevant magnitude of benefit, and to make appropriate grade adjustment to reflect the observed magnitude of benefit. To achieve this goal, a dual rule was implemented; first, taking into account the variability of the estimated HR from a study, the lower limit of the 95% Confidence Interval (CI) for the HR is compared to specified threshold values; and second the observed absolute difference in treatment outcomes is compared to the minimum absolute gain considered as beneficial. Different candidate threshold values for HR and absolute gains for survival, DFS and PFS, adjusted to represent
as accurately as possible the expert opinion of the oncology community, have been explored through extensive simulations. The finally implemented combined thresholds for the HR and the minimum observed benefit that could be considered as deserving the highest grade in both the curative and non-curative setting are outlined in Table 2.

In all forms HR thresholds refer to the lower extreme of the 95% CI (Figure 2). The performance of the evaluation rule based on the lower limit of the 95% CI of HR, was compared to the simpler rule of using a cut-off for the point estimate of HR, in conjunction with the additional rule on the minimum absolute gain in treatment outcome. The simulation results under different HR values and corresponding power, favoured the proposed approach to use the lower limit of the 95% CI which takes into account the variability of the estimate. The correspondence between an HR value and the minimum absolute gain considered as beneficial according to the ESMO-MCBS, is presented by median survival (OS or PFS) for standard treatment, in Figure 2. For example, for a standard treatment median survival of 6 months, an absolute gain of 3 months corresponds to an HR=0.67, while a gain of 1.5 months corresponds to an HR=0.8.

**the ESMO magnitude of clinical benefit scale (ESMO-MCBS v1.0) (Appendix I)**

The ESMO Magnitude of Clinical Benefit Scale version 1 (ESMO-MCBS v1.0) has been developed only for solid cancers. Given the profound differences between the curative and palliative settings the tool is presented in two parts. Form 1 is used to evaluate adjuvant and other treatments with curative intent. Form 2 (a, b or c) is used to evaluate non-curative interventions, with form 2a for studies with OS as the primary outcome, form 2b for studies with PFS or TTP as primary outcomes, 2c for studies with QoL, toxicity or response rate as primary outcomes and for non-inferiority studies. Form 2a is prognostically sub-stratified for studies where the control arm produced OS greater or less than or equal to 1 year and form 2b for studies where the control arm produced PFS greater or less than or equal to 6 months.
eligibility for application of the ESMO-MCBS

The ESMO-MCBS can be applied to comparative outcome studies evaluating the relative benefit of treatments using outcomes of survival, QoL, surrogate outcomes for survival or QoL (DFI, EFS, TTR, PFS and TTP) or treatment toxicity in solid cancers. Eligible studies can have either a randomised or comparative cohort design [35, 36] or a meta-analysis which report statistically significant benefit from any one, or more of the evaluated outcomes. When more than one study has evaluated a single clinical question, results derived from well powered registration trials should be given priority.

Studies with pre-planned subgroup analyses with a maximum of 3 subgroups can be scored. When statistically significant results are reported for more than one subgroup, then each of these should be evaluated separately. Subgroups not showing statistically significant results are not graded. Except for studies that incorporate collection of tissue samples to enable re-stratification based on new genetic or other biomarkers, findings from un-planned (post-hoc) subgroup analysis cannot be graded and they can only be used as foundation for hypothesis generation.

**form 1**

This form is used for adjuvant and neo-adjuvant therapies and for localised or metastatic diseases being treated with curative intent. This scale is graded A, B or C. Grades A and B represent a high level of clinical benefit (Figure 1). The scale makes allowance for early data demonstrating high DFS without mature survival data. Studies initially evaluated based on DFS criteria alone will need to be revaluated when mature survival data is available. Hyper mature data from studies that were un-blinded after compelling early results with subsequent access to the superior arm are contaminated, subsequently late intention to treat (ITT) follow-up data are not evaluable [37, 38]. Pathological complete remission from neo-adjuvant therapies is not included as a criteria for clinical benefit because of lack of consistent evidence that it is a valid surrogate for survival in clinical studies [39-42].
forms 2

These forms are used for studies of new agents or approaches in the management of cancers without curative intent. This scale is graded 5, 4, 3, 2, 1, where grades 5 and 4 represent a high level of proven clinical benefit (Figure 1).

2a: This version is used for therapies evaluated using a primary outcome of OS. The form is stratified by median OS of the control arm ≤12 months and >12 months. Preliminary grading takes into consideration HR and median survival gain as well as late survival advantage and is reported on a 4 point scale. When there is differential grading between the median and late survival gain, the higher score prevails. Preliminary scores can be upgraded by 1 point when the experimental arm demonstrates improved QoL or delayed deterioration in QoL using a validated scale or substantial reduction in grade 3 or 4 toxicity. A score of 5 can only be achieved when optimal survival outcomes are further enhanced by data indicating reduced toxicity or improved quality of life.

2b: This version is used for therapies evaluated using a primary endpoint of PFS or TTP. The form is stratified by median duration of PFS of the control arm ≤6 months and >6 months. The maximal preliminary score is discounted to 3 because PFS and TTP are surrogate outcomes with a less reliable relationship to improved survival or QoL [18, 20-23]. In studies that allow crossover on subsequent therapy, this may be the best available evidence of activity since subsequent therapies may reduce the likelihood of observing survival benefit.

Preliminary scores derived from PFS studies can be upgraded or downgraded depending on secondary outcomes such as toxicity data, improvement in OS or data derived from QoL evaluation. This form incorporates an adverse effect criterion for downgrading in cases of severe toxicity compared to the control arm. If an OS advantage is observed as a secondary outcome, scores are upgraded using the scale on form 2a. In PFS studies that evaluate global QoL, positive findings (as evidenced by statistically significant improvement in global QoL or delayed deterioration in QoL) will upgrade the evaluation by 1 point and, in the absence of survival advantage, the absence of QoL advantage will result in a downgrading by 1 point.
This form is used for therapies evaluated in non-inferiority (equivalence) studies and for studies in which the primary outcomes are QoL, toxicity or response rate (RR).

**field testing of ESMO-MCBS**

ESMO-MCBS has been applied in a wide range of solid tumours by members of the ESMO-MCBS Task Force, the ESMO Guidelines Committee and a range of invited experts (Tables 3-12). When discrepancies between graders were observed, this was generally related to either inaccurate data extraction, variable interpretation of the significance and severity of toxicity data, or errors in applying the data to the correct grading criteria.

**Discussion**

**inherent challenges in developing standard clinical benefit scale**

The substantial variability of study designs (crossover, non-crossover, and partial crossover), planned outcomes and reported outcomes inherently challenge the process of developing a unified scale of clinical benefit. This challenge is all the greater in an era in which both researchers and regulatory authorities are employing surrogate outcome indicators as primary end points for both research and registration criteria [5]. A unified scaling approach requires a process of relative weighting of evidence that demands conceptual rigor, careful reviews of the validity and strength of surrogate endpoints and clinical nuance.

**validity of the ESMO-MCBS**

The ESMO-MCBS version 1 (ESMO-MCBS v1.0) provides an objective and reproducible approach that allows comparisons of the magnitude of benefit between studies that incorporate different primary outcomes (OS, PFS, QoL) and different designs through a process of variable weighting of primary outcomes and adjustments for significant secondary outcomes and toxicity.

The development process has been compliant with the criteria for “accountability for reasonableness” which represent the ethical gold-standard for a fair priority setting process in public policy [43, 44]. The validity of the ESMO-MCBS is derived from 1) Clinically relevant and reasonable criteria for prioritisation of different types of benefit, i.e. that cure takes
precedence over deferral of death, direct endpoints such as survival and QoL take precedence over less reliable surrogates such as PFS or RR and that the interpretation of the evidence for benefit derived from indirect primary outcomes (such as PFS, or RR) may be influenced by secondary outcome data, 2) Coherence: Procedural agreements regarding the evidence to be used/not used, how it will be analysed and evaluated, and precautions to minimising bias (including conflict of interest issues) based upon an understanding of the relative strengths and weaknesses of the usual measured outcomes OS and QoL, and their surrogates and rigorous bio statistical review, 3) Wide applicability over a range of solid cancers and a range of prognoses that have been rigorously tested 4) Statistical validity and 5) A transparent process of development with scope for peer review, appeal and revision.

ESMO-MCBS scores for a specific therapy are not generalisable to indications outside the confines of the context in which they have been evaluated. Consequently the ESMO-MCBS score for a particular medication or therapeutic approach may vary depending on the specifics of the indication and may vary between studies.

**limitations of the ESMO-MCBS v1.0**

The ESMO-MCBS can only be applied to comparative research outcomes; it is therefore not applicable when evidence of benefit derives from single arm studies. This limits its utility in the uncommon situation in which registration is granted on the basis of outcomes reported from single arm studies.

The process of relative weighting of evidence and the thresholds for HR and absolute gains involves judgments and subjective considerations which are amenable to dispute and challenge and indeed, this is invited as part of the dynamic process of peer-review and further development.
factors that may skew or alter ESMO-MCBS scores

**Control arm evaluation:** The ESMO-MCBS evaluates data derived from comparative research, either randomised phase II [46] or phase III studies or cohort studies. The validity of the results may be influenced by the quality and design of the study. Design issues are critical insofar as studies that incorporate a relatively weak control arm may generate the impression of exaggerated benefit. This was manifest in studies evaluating treatment options for hormone refractory prostate cancer where one study used mitoxantrone/prednisone as the control arm [47] based on the findings of a phase III study comparing prednisone vs the combination of prednisone and mitoxantrone which demonstrated improved QoL but no survival advantage for the combination therapy [48] and others used prednisone alone [49] or placebo [50].

**Crossover:** Crossover, or subsequent treatment of control arm patients with biologically similar agent, severely compromises the ability to derive reliable data regarding the survival advantage of treatments in phase III studies. This factor may impact on OS results as illustrated by the study of dacarbazine vs ipilimumab in metastatic melanoma [51] in which the evidence for survival advantage was diluted by the crossover provision in the study. In some instances in which strong PFS advantage is seen, crossover of this type will obscure the potential survival benefit of the new treatment. Statistical approaches to estimate longer-term clinical outcomes despite substantial treatment crossover have been developed [52, 53], and applied [54-57]. While these approaches are encouraging they incorporate a range of assumptions and are not universally accepted [58].

**Unbalanced crossover:** In other instances, unbalanced crossover may exaggerate differences in survival. For instance in the PEAK study comparing FOLFOX6 with either bevacizumab or panitumumab among the patients with KRAS wild type tumours, only 38% of those in the bevacizumab arm received any EGFR antibody in subsequent therapy [59]. Although this study showed a survival advantage of 9.9 months over a baseline of 24.3 months for patient initiated on treatment with panitumumab, it remains unclear as to
whether this was affected by the sequence of treatments or if it was the result that more
than half of the patients in the bevacizumab arm were never exposed to an EGFR antibody.

**Follow up reports:** In some studies first reports are followed up with subsequent further
relevant data analysis. This is particularly true when mature survival data was not available
in studies with a primary outcome of PFS or DFS and in studies that have incorporated post
hoc stratification based on refined appreciation of tumour biology that may impact on
outcome evaluation.

Both of these phenomena were observed in the three publications reporting the findings
from the same study on FOLFOX4 +/- panitumumab for the first-line treatment of KRAS wild
type colorectal cancer [60-62]. The study, which did allow for crossover to other EGFR
antibodies, had PFS as a primary endpoint. The initial publication demonstrated a modest
PFS advantage with non-significant median OS gain [60]. The subsequent publication of
mature data demonstrated a significant OS gain [61] with the greatest benefit restricted to
patients with KRAS, NRAS, BRAF wild type tumours [62]. Almost identical data maturation
was observed in the CRYSTAL study evaluating FOLFIRI +/- cetuxumab in the same clinical
setting [63-65].

Maturation of survival data also increased the ESMO-MCBS score of vemurafenib in the
treatment of metastatic melanoma [66, 67] from ESMO MCBS 3 based on PFS to 4, based on
OS.

**using data from the ESMO-MCBS**

The ESMO-MCBS incorporates a structured, rational and valid approach to data
interpretation and analysis that reduces the tendency to have judgments affected by bias or
uninformed and/or idiosyncratic data interpretation. Consequently, application of the scale
reduces the likelihood that statements of clinical benefit will be distorted by either
overestimation or overstatement on one extreme or, nihilism at the other. This structured
and disciplined approach to deriving estimates of clinically meaningful benefit from
published data can be used in a range of settings.
Public policy applications: Grading derived from the ESMO-MCBS provides a backbone for value evaluations for cancer medicines. Medicines and therapies that fall into the ESMO-MCBS A+B for curative therapies and 4+5 for non-curative therapies should be highlighted for accelerated assessment of value and cost effectiveness. While a high ESMO-MCBS score does not automatically imply high value (that depends on the price), the scale can be utilised by to frame such considerations[68] and can help public policy-makers advance “accountability for reasonableness” in resource allocation deliberations[43, 44].

Formulation of Clinical Guidelines: The prevailing current practice of the National Comprehensive Cancer Network (NCCN), American Society of Clinical Oncology (ASCO), ESMO and the National Cancer Institute(NCI)in their guidelines is to grade the “level of evidence” supporting the efficacy of therapeutic interventions; grading the evidence as very high when derived from meta-analyses of well conducted phase III studies, or from large well conducted phase III studies relative to lower levels such as that derived from non-randomised studies, anecdote or expert clinical opinion. A major shortcoming of this approach is that it may result in a high level of evidence irrespective of the actual magnitude of the benefit observed, even if the magnitude of the benefit is very limited[69]. This discrepancy has been highlighted by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group which was formed in 2000 to improve the quality of guideline formulation. The GRADE working group emphasised that a particular quality of evidence does not necessarily imply a particular strength of recommendation[70, 71]. They have developed and championed a widely endorsed approach emphasising appropriate framing of research and guideline questions[72], evaluation of the strength of recommendations that incorporates evaluation of the balance between desirable and undesirable outcomes (estimated effects), and the confidence in the magnitude effect of the interventions on important outcomes[73].

This recommendation can be accomplished by describing both the level of benefit and the level of evidence for recommended therapeutic interventions. For cancer therapies, the ESMO-MCBS scale provides a clear, well-structured and validated mechanism to indicate the magnitude of benefit in addition to the level of evidence that can inform both national and international (e.g. ESMO) guidelines.
Clinical decision making: The data encapsulated in ESMO-MCBS scoring can help clinicians to weigh the relative merits of competing relevant therapeutic options in situations in which there is no direct comparative data comparing the available therapeutic options. The grading may also be of benefit in explaining the relative merit of therapeutic options to patients and their families. This information may be especially helpful when treatments incorporate substantial out of pocket costs and the real “value” of the treatment needs to be candidly addressed to avoid over investment or sacrifice of limited financial resources to pay for treatments that have only limited magnitude of benefit.

Editorial decisions and commentaries: The ESMO-MCBS may be of use to editors, peer reviewers and commentators in considering the clinical significance of research findings from randomised clinical studies, cohort studies and meta-analyses with statistically significant positive findings.

Relevance to the ASCO initiatives

ASCO has undertaken two initiatives to help promote the value in cancer care. The first was a working group to propose new thresholds for the approval of cancer medications[74]. For each of four conditions (metastatic colon cancer, metastatic breast cancer, non-small cell lung cancer and pancreatic cancer) they have proposed thresholds for meaningful clinical benefit improvement defined in terms of minimal increases in OS (absolute and HR) and also thresholds for minimal increases in surrogate indicators including 1 year survival and PFS. Interestingly, in non-curative therapies the ASCO recommended thresholds for survival benefit correlate very closely to the thresholds for ESMO-MCBS score of 4-5 (in form 2a) and the recommended thresholds for PFS correlate closely with the thresholds for ESMO-MCBS score of 3-4 which is the highest attainable when the primary outcome is PFS (in form 2b). Secondly ASCO has developed a Value in Cancer Care Task Force that has been charged with the challenge of developing a framework for evaluating value in oncology. While concurring with ESMO that the evaluation of net clinical benefit is key element in the evaluation of value, ASCO has not yet described their proposed approach to evaluate the magnitude of clinical benefit. A key challenge for the future will be to establish whether
there can be harmonisation between the different approaches to value in Europe and the USA.

**Conclusion**

ESMO is committed to promoting rational, responsible and affordable cancer care, the importance of organisational integrity, and the promotion of best use of limited health care resources. The ESMO-MCBS v-1.0 was born out of these considerations. It represents a first version of a well validated tool to stratify the magnitude of clinical benefit for new anticancer treatments and is applicable over a full range of solid tumours. Based on the data derived from well-structured phase III clinical trials or meta-analyses, the tool uses a rational, structured and consistent approach to derive a relative ranking of the magnitude of benefit that can be anticipated from any new treatment. The ESMO-MCBS is an important first step to the major ongoing task of evaluating value in cancer care which is essential for appropriate uses of limited public and personal resources for affordable cancer care. The ESMO-MCBS will be a dynamic tool and its criteria will be revised on a regular basis pending peer reviewed feedback and developments in cancer research and therapies.

**Acknowledgements (Appendix II)**

The authors wish to acknowledge the support and contribution of the ESMO Executive Board, the ESMO Faculty, Members of the ESMO Guidelines Committee, expert biostatisticians and the logistic and organisational support provided by ESMO Staff Nicola Latino.

**funding**

This project was funded by ESMO.

**disclosure**

The authors have declared the following:

Nathan I Cherny: No conflicts of interest to declare.

Urania Dafni: Lecture Fees - Amgen
Elisabeth G.E. de Vries: Research grants from Roche/Genentech, Amgen, Novartis are made available to my hospital the UMCG in Groningen.
J.Martijn Kerst: No conflicts of interest to declare.

Martine J. Piccart: Board member - PharmaMar; Consultant (honoraria) - Amgen, Astellas, AstraZeneca, Eli Lilly, GSK, Invivis, MSD, Novartis, Pfizer, Roche-Genentech, Sanofi Aventis, Symphogen, Synthon, Verastem; Research grants to Institute: Most companies; Speakers bureau/stock ownership: None.
Alberto Sobrero: Advisory Board and Symposia Satellite with; Amgen, Bayer, Celgene, Merck, Roche and Sanofi.
Richard Sullivan: No conflicts of interest to declare.
Christoph Zielinski: Advisory Boards - Roche, Celgene, Bristol Myers-Squibb; Lecture fees - Amgen, Bristol Myers-Squibb
references

17. Gill S, Sargent D. End points for adjuvant therapy trials: Has the time come to accept disease-free survival as a surrogate end point for overall survival? Oncologist 2006; 11: 624-629.
40. Burki TK. Pathological complete response is no surrogate for survival. Lancet Oncol 2014; 15: e111.


103. Schwartzberg LS, Rivera F, Karthaus M et al. PEAK: a randomized, multicenter phase II study of panitumumab plus modified fluorouracil, leucovorin, and oxaliplatin (mFOLFOX6) or bevacizumab plus mFOLFOX in patients with previously untreated, unresectable, wild-type KRAS exon 2 metastatic colorectal cancer. J Clin Oncol 2014; 32: 2240-2247.
Table 1: Potential benefits of a new treatment

Living longer

- Improved OS
- Improved surrogate of OS
  - DFS (when OS data is immature in adjuvant setting)
  - Improved PFS

Living better

- Improved quality of life
- Improved surrogate of quality of life
  - Improved PFS
- Reduced toxicity
Figure 1: Visualisation of ESMO-MCB scores for curative and non-curative setting.

**Curative** - Evaluation form 1: for new approaches to adjuvant therapy or new potentially curative therapies

**Non-curative** - Evaluation forms 2a, b or c: for therapies that are not likely to be curative
Figure 2: Use of threshold HR in the ESMO-MCBS exemplified for HR threshold of 0.65.
Figure 3: The correspondence between an HR value and the minimum absolute gain in months (mth) considered as beneficial according to the ESMO-MCBS by median survival (OS or PFS) for control.
Table 2: Maximal Preliminary Scores

Treatments with Curative Intent (Form 1)
>5% improvement of survival at ≥ 3 years follow-up
Improvements in DFS alone HR <0.60 (primary endpoint) in studies without mature survival data

Treatments with Non Curative Intent (Form 2)

Primary outcome OS (Form 2a)
Control ≤ 12 months
HR ≤ 0.65 AND Gain ≥ 3 months OR
Increase in 2 year survival alone ≥ 10%

Control > 12 months
HR ≤ 0.70 AND Gain ≥ 5 months OR
Increase in 3 year survival alone ≥ 10%

Primary outcome PFS (Form 2b)
Control ≤ 6 months
HR ≤ 0.65 AND Gain ≥ 1.5 months

Control > 6 months
HR ≤ 0.65 AND Gain ≥ 3 months
<table>
<thead>
<tr>
<th>Medication (New vs control)</th>
<th>Trial name</th>
<th>Setting</th>
<th>Primary outcome</th>
<th>PFS control</th>
<th>PFS gain</th>
<th>PFS HR</th>
<th>OS control</th>
<th>OS gain</th>
<th>OS HR</th>
<th>QoL</th>
<th>Toxicity</th>
<th>ESMO-MCBS ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erlotinib vs carboplatin gemcitabine</td>
<td>OPTIMAL, CTONG-0802</td>
<td>1st line stage IIIb or IV non-squamous, with EGFR mutation</td>
<td>PFS</td>
<td>4.6 mth</td>
<td>8.5 mth</td>
<td>0.16 (0.10–0.26)</td>
<td>19.5 mth</td>
<td>NS</td>
<td>12% less serious adverse events</td>
<td>4 [75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erlotinib vs platinum-based chemotherapy doublet</td>
<td>EURTAC</td>
<td>1st line stage IIIb or IV non-squamous, with EGFR mutation</td>
<td>PFS (crossover allowed)</td>
<td>5.2 mth</td>
<td>4.5 mth</td>
<td>0.37 (0.25–0.54)</td>
<td>NS</td>
<td>15% less severe adverse reactions</td>
<td>4 [76]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gefitinib vs carboplatin + paclitaxel</td>
<td>IPASS</td>
<td>1st line stage IIIb or IV adenocarcinoma, with EGFR mutation</td>
<td>PFS (crossover allowed)</td>
<td>6.3 mth</td>
<td>3.3 mth</td>
<td>0.48 (0.34-0.67)</td>
<td>NS</td>
<td>Improved</td>
<td>Reduced toxicity</td>
<td>4 [77, 78]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afatinib vs Cisplatin + pemetrexed</td>
<td>LUX-Lung 3</td>
<td>1st line stage IIIb or IV adenocarcinoma with EGFR mutation (Del19/L858R)</td>
<td>PFS (crossover allowed)</td>
<td>6.9 mth</td>
<td>6.7 mth</td>
<td>0.47 (0.34–0.65)</td>
<td>0.47 (0.34–0.65)</td>
<td>Improved</td>
<td>Improved</td>
<td>4 [79, 80]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crizotinib vs chemotherapy</td>
<td>1st line stage IIIb or IV non-squamous, with ALK mutation</td>
<td>PFS (crossover allowed)</td>
<td>3.0 mth</td>
<td>4.7 mth</td>
<td>0.49 (0.37–0.64)</td>
<td>NS</td>
<td>Improved</td>
<td>1% increased toxic death</td>
<td>4 [81]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crizotinib vs cisplatin + pemetrexed</td>
<td>1st line stage IIIb or IV non-squamous, with ALK mutation</td>
<td>PFS</td>
<td>7.0 mth</td>
<td>3.9 mth</td>
<td>0.45 (0.35-0.60)</td>
<td>NS</td>
<td>Improved</td>
<td>4 [82]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pemetrexed vs placebo</td>
<td>Stage IIIb or IV disease maintenance after responding to 4 cycles platinum doublet</td>
<td>PFS stratified for histology (non-squamous)</td>
<td>2.6 mth</td>
<td>1.9 mth</td>
<td>0.47 (0.37–0.60)</td>
<td>10.3 mth</td>
<td>5.2 mth</td>
<td>0.70 (0.56–0.88)</td>
<td>Improved</td>
<td>4 [83]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisplatin pemetrexed vs cisplatin gemcitabine</td>
<td>1st line stage IIIb or IV (non-squamous)</td>
<td>OS (non-inferiority)</td>
<td>10.4 mth</td>
<td>1.4 mth</td>
<td>0.81 (0.70–0.94)</td>
<td>NS</td>
<td>Less grade 3+ toxicity neutropenia anaemia thrombocytopenia</td>
<td>4 [84]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotherapy +/- palliative care</td>
<td>Stage IV non-small cell ECOG&lt;2</td>
<td>QoL</td>
<td>8.9 mth</td>
<td>2.7 mth</td>
<td>HR for death in control arm 1.7 (1.14-2.54)</td>
<td>Improved</td>
<td>NS</td>
<td>4 [85]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paclitaxel/carboplatin +/- bevacizumab</td>
<td>1st line stage IIIb or IV, non-squamous</td>
<td>OS</td>
<td>10.3 mth</td>
<td>2.0 mth</td>
<td>0.79 (0.67-0.92)</td>
<td>NS</td>
<td>NS</td>
<td>2 [86]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erlotinib vs placebo</td>
<td>SATURN</td>
<td>Stage IIIb or IV disease maintenance after responding to 4-6 cycles platinum doublet</td>
<td>PFS</td>
<td>11.1 wk</td>
<td>1.2 wk</td>
<td>0.71 (0.62–0.82)</td>
<td>11.0 mth</td>
<td>1.0 mth</td>
<td>0.81 (0.70-95)</td>
<td>Improved</td>
<td>1 [87]</td>
<td></td>
</tr>
<tr>
<td>Medication</td>
<td>Trial name</td>
<td>Setting</td>
<td>Primary outcome</td>
<td>PFS control</td>
<td>PFS gain</td>
<td>PFS HR</td>
<td>OS control</td>
<td>OS gain</td>
<td>OS HR</td>
<td>QoL</td>
<td>Toxicity</td>
<td>ref</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>---------</td>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>------------</td>
<td>---------</td>
<td>-------</td>
<td>-----</td>
<td>----------</td>
<td>-----</td>
</tr>
<tr>
<td>Chemotherapy +/- trastuzumab</td>
<td>HERA</td>
<td>(Neo)adjuvant HER-2 positive tumours</td>
<td>DFS</td>
<td>2 years DFS 77.4%</td>
<td>8.40%</td>
<td>0.54 (0.43-0.67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T-DM1 vs lapatinib + capcitabine</td>
<td>EMILIA</td>
<td>2nd line metastatic after trastuzumab failure</td>
<td>PFS and OS</td>
<td>6.4 mth</td>
<td>3.2 mth</td>
<td>0.65 (0.55-0.77)</td>
<td>25 mth</td>
<td>6.8 mth</td>
<td>0.68 (0.55-0.86)</td>
<td>Delayed deterioration</td>
<td>5 [89, 90]</td>
</tr>
<tr>
<td></td>
<td>Trastuzumab + chemotherapy +/- pertuzumab</td>
<td>CLEOPATRA</td>
<td>1st line metastatic</td>
<td>PFS</td>
<td>12.4 mth</td>
<td>6 mth</td>
<td>0.62 (0.52-0.84)</td>
<td>40.8 mth</td>
<td>15.7 mth</td>
<td>0.68 (0.56-0.84)</td>
<td>No improvement</td>
<td>4 [91-94]</td>
</tr>
<tr>
<td></td>
<td>Lapatinib +/- trastuzumab</td>
<td>EGF104900</td>
<td>3rd line metastatic</td>
<td>PFS</td>
<td>2 mth</td>
<td>1 mth</td>
<td>0.73 (0.57-0.93)</td>
<td>9.5 mth</td>
<td>4.5 mth</td>
<td>0.74 (0.57-0.97)</td>
<td></td>
<td>4 [95, 96]</td>
</tr>
<tr>
<td></td>
<td>Capecitabine +/- lapatinib</td>
<td></td>
<td>2nd line metastatic after trastuzumab failure</td>
<td>PFS</td>
<td>4.4 mth</td>
<td>4 mth</td>
<td>0.49 (0.34-0.71)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3 [97]</td>
</tr>
<tr>
<td></td>
<td>Eribulin vs other chemotherapy</td>
<td>EMBRACE</td>
<td>3rd line metastatic after anthracycline and taxane</td>
<td>OS</td>
<td></td>
<td></td>
<td></td>
<td>10.6 mth</td>
<td>2.5 mth</td>
<td>0.81 (0.66-0.99)</td>
<td></td>
<td>2 [98]</td>
</tr>
<tr>
<td></td>
<td>Paclitaxel +/- bevacizumab</td>
<td></td>
<td>1st line metastatic</td>
<td>PFS</td>
<td>5.9 mth</td>
<td>5.8 mth</td>
<td>0.60 (0.51-0.70)</td>
<td></td>
<td></td>
<td>NS</td>
<td>No improvement</td>
<td>2 [24]</td>
</tr>
<tr>
<td></td>
<td>Exemestane +/- everolimus</td>
<td>BOLERO-2</td>
<td>Metastatic after failure of aromatase inhibitor (with PFS&gt;6 mth)</td>
<td>PFS</td>
<td>4.1 mth</td>
<td>6.5 mth</td>
<td>0.43 (0.35-0.54)</td>
<td></td>
<td></td>
<td>NS</td>
<td>No improvement</td>
<td>2 [99]</td>
</tr>
<tr>
<td>Medication</td>
<td>Trial name</td>
<td>Setting</td>
<td>Primary outcome</td>
<td>PFS control</td>
<td>PFS gain</td>
<td>PFS HR</td>
<td>OS control</td>
<td>OS gain</td>
<td>OS HR</td>
<td>QoL</td>
<td>Toxicity</td>
<td>ESMO-MCBS</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>--------------</td>
<td>----------------------------------------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------------</td>
<td>-----</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Best standard non chemotherapy or radiotherapy treatment +/- radium-223</td>
<td>ALSYMPCA</td>
<td>Castration refractory and bone pain</td>
<td>OS</td>
<td>11.3 mth</td>
<td>3.6 mth</td>
<td>0.70 (0.55-0.88)</td>
<td>Improved</td>
<td>5</td>
<td>[100]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prednisone +/- abiraterone</td>
<td></td>
<td>Castration refractory after docetaxel</td>
<td>OS</td>
<td>10.9 mth</td>
<td>3.9 mth</td>
<td>0.65 (0.44-0.77)</td>
<td>4</td>
<td>[49]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enzalutamide vs placebo</td>
<td>AFFIRM</td>
<td>Castration refractory after docetaxel</td>
<td>OS</td>
<td>13.6 mth</td>
<td>4.8 mth</td>
<td>0.63 (0.43-0.75)</td>
<td>Improved</td>
<td>4</td>
<td>[50]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enzalutamide vs placebo</td>
<td>PREVAIL</td>
<td>Castration refractory pre docetaxel</td>
<td>PFS and OS</td>
<td>3.2 mth</td>
<td>&gt;12 mth</td>
<td>0.19 (0.15-0.23)</td>
<td>Improved</td>
<td>3</td>
<td>[101]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Docetaxel (Q7 or Q21) prednisone vs mitoxantrone + prednisone</td>
<td></td>
<td>Castration refractory</td>
<td>OS</td>
<td>16.5 mth</td>
<td>2.4 mth (Q21)</td>
<td>0.76 (0.52-0.94)</td>
<td>Improved</td>
<td>3</td>
<td>[102]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cabazitaxel+ prednisone vs mitoxantrone + prednisone</td>
<td>TROPIC</td>
<td>Castration refractory after docetaxel</td>
<td>OS</td>
<td>12.7 mth</td>
<td>2.4 mth</td>
<td>0.70 (0.59-0.83)</td>
<td>2</td>
<td>[47]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table 6: Field testing ESMO-MCBS v1.0: Colorectal Cancer**

<table>
<thead>
<tr>
<th>Medication</th>
<th>Trial name</th>
<th>Setting</th>
<th>Primary outcome</th>
<th>PFS control</th>
<th>PFS gain</th>
<th>PFS HR</th>
<th>OS control</th>
<th>OS gain</th>
<th>OS HR</th>
<th>QoL</th>
<th>Toxicity</th>
<th>ESMO-MCBS</th>
<th>ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOLFOX4 +/- panitumumab</td>
<td>PRIME</td>
<td>1st line metastatic (Post hoc KRAS, NRAS BRAF WT)</td>
<td>PFS</td>
<td>7.9 mth</td>
<td>2.3 mth</td>
<td>0.72 (0.58-0.90)</td>
<td>20.2 mth</td>
<td>5.8 mth</td>
<td>0.28 (0.62-0.99)</td>
<td>4</td>
<td>[62]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panitumumab + mFOLFOX6 vs bevacizumab +mFOLFOX6</td>
<td>PEAK</td>
<td>1st line metastatic (KRAS-WT)</td>
<td>PFS</td>
<td>NS</td>
<td></td>
<td></td>
<td>24.3 mth</td>
<td>9.9 mth</td>
<td>0.62 (0.44-0.89)</td>
<td>4*</td>
<td>[103]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOLFIRI +/- cetuximab</td>
<td>CRYSTAL</td>
<td>1st line metastatic stratified for KRAS-WT (Post hoc KRAS, NRAS WT)</td>
<td>PFS</td>
<td>8.4 mth</td>
<td>3.0 mth</td>
<td>0.56 (0.41-0.76)</td>
<td>20.2 mth</td>
<td>8.2 mth</td>
<td>0.59 (0.54-0.88)</td>
<td>4</td>
<td>[65]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cetuximab vs best supportive care</td>
<td></td>
<td>Refractory metastatic KRAS-WT</td>
<td>OS</td>
<td>1.9 mth</td>
<td>1.8 mth</td>
<td>0.4 (0.30-0.54)</td>
<td>4.8 mth</td>
<td>4.7 mth</td>
<td>0.35 (0.41-0.74)</td>
<td>4</td>
<td>[104]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOLFOX4 +/- panitumumab</td>
<td>PRIME</td>
<td>1st line metastatic KRAS-WT</td>
<td>PFS</td>
<td>8 mth</td>
<td>1.6 mth</td>
<td>0.80 (0.66-0.97)</td>
<td>19.4 mth</td>
<td>4.4 mth</td>
<td>0.68 (0.70-0.98)</td>
<td>4</td>
<td>[60, 61]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOLFIRI +/- cetuximab</td>
<td>CRYSTAL</td>
<td>1st line metastatic stratified for KRAS-WT</td>
<td>PFS</td>
<td>8.4 mth</td>
<td>1.5 mth</td>
<td>0.70 (0.56-0.87)</td>
<td>20 mth</td>
<td>3.5 mth</td>
<td>0.60 (0.67-0.95)</td>
<td>3</td>
<td>[63, 64]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILF +/- bevacizumab</td>
<td></td>
<td>1st line metastatic</td>
<td>OS</td>
<td>15.6 mth</td>
<td>4.7 mth</td>
<td>0.66 (0.54-0.81)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>[105]</td>
<td></td>
</tr>
<tr>
<td>FOLFIRI +/- panitumumab</td>
<td></td>
<td>2nd line metastatic KRAS-WT</td>
<td>PFS</td>
<td>3.9 mth</td>
<td>2 mth</td>
<td>0.73 (0.59-0.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>[106]</td>
</tr>
<tr>
<td>FOLFOX+/- bevacizumab vs bevacizumab alone</td>
<td>E3200</td>
<td>2nd line metastatic after FOLFIRI</td>
<td>OS</td>
<td>10.8 mth</td>
<td>2.1 mth</td>
<td>0.75 (0.63-0.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>[107]</td>
</tr>
<tr>
<td>Panitumumab, vs best supportive care</td>
<td></td>
<td>3rd line metastatic stratified for KRAS</td>
<td>PFS</td>
<td>7.3 wk</td>
<td>5 wk</td>
<td>0.45 (0.34-0.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>[108]</td>
</tr>
<tr>
<td>FOLFIRI bevacizumab vs FOLFIRI + bevacizumab</td>
<td></td>
<td>1st line metastatic</td>
<td>PFS</td>
<td>9.7 mth</td>
<td>2.4 mth</td>
<td>0.75 (0.62-0.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>[109]</td>
</tr>
<tr>
<td>TAS-102 vs placebo</td>
<td>CONCOURSE</td>
<td>3rd line or beyond metastatic</td>
<td>OS</td>
<td>5.3 mth</td>
<td>1.8 mth</td>
<td>0.68 (.058-0.81)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>[110]</td>
</tr>
<tr>
<td>Regorafenib vs placebo</td>
<td>CORRECT</td>
<td>3rd line metastatic</td>
<td>OS</td>
<td>5 mth</td>
<td>1.4 mth</td>
<td>077 (0.64-0.94)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>[111]</td>
</tr>
<tr>
<td>2nd line chemotherapy +/- bevacizumab</td>
<td>ML18147</td>
<td>2nd line beyond progression on bevacizumab</td>
<td>OS</td>
<td>9.6 mth</td>
<td>1.5 mth</td>
<td>0.81 (0.69-0.94)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>[112]</td>
</tr>
<tr>
<td>FOLFIRI +/- aflibercept</td>
<td>VELOUR</td>
<td>2nd line after oxaliplatin based treatment</td>
<td>OS</td>
<td>4.7 mth</td>
<td>2.2 mth</td>
<td>0.76 (0.66-0.87)</td>
<td>12 .1 mth</td>
<td>1.5 mth</td>
<td>0.82 (0.71-0.94)</td>
<td>1</td>
<td>[113]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOLFIRI +/- Ramucirumab</td>
<td>RAISE</td>
<td>2nd line metastatic after bevacizumab, oxaliplatin, fluoropyrimidine</td>
<td>OS</td>
<td>11.7 mth</td>
<td>1.6 mth</td>
<td>0.84 (0.73-0.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>[114]</td>
</tr>
</tbody>
</table>

*unbalanced crossover
<table>
<thead>
<tr>
<th>Medication</th>
<th>Trial name</th>
<th>Setting</th>
<th>Primary outcome</th>
<th>PFS control</th>
<th>PFS gain</th>
<th>PFS HR</th>
<th>OS control</th>
<th>OS gain</th>
<th>OS HR</th>
<th>QoL</th>
<th>Toxicity</th>
<th>ESMO-MCBS</th>
<th>ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paclitaxel or topotecan liposomal doxorubicin +/- bevacizumab</td>
<td>AURELIA</td>
<td>Recurrent platinum resistant</td>
<td>PFS (crossover allowed)</td>
<td>3.4 mth</td>
<td>3.3 mth</td>
<td>0.48 (0.38-0.60)</td>
<td>Improved</td>
<td>4</td>
<td>[115, 116]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paclitaxel and carboplatin (5 or 6 cycles) +/- bevacizumab till 18 cycles or progression</td>
<td>ICON7</td>
<td>High risk, early stage post resection or advanced ovarian or primary peritoneal</td>
<td>PFS stratified for stage and risk of progression</td>
<td>(All) 22.4 mth</td>
<td>1.7 mth</td>
<td>0.81 (0.70-0.94)</td>
<td>NS</td>
<td>1</td>
<td>[117]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(high risk) 14.5 mth</td>
<td>3.6 mth</td>
<td>0.73 (0.60-0.90)</td>
<td>28.8 mth</td>
<td>7.8 mth</td>
<td>0.64 (0.48-0.85)</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gemcitabine and carboplatin +/- bevacizumab</td>
<td>OCEANS</td>
<td>Recurrent platinum sensitive</td>
<td>PFS (crossover allowed)</td>
<td>8.4 mth</td>
<td>4 mth</td>
<td>0.48 (0.39-0.61)</td>
<td>Improved</td>
<td>3</td>
<td>[118]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paclitaxel and carboplatin (6 cycles) +/- bevacizumab continual till 10 months or progression</td>
<td>GOG 218</td>
<td>Incompletely resected stage III and stage IV</td>
<td>PFS (crossover allowed)</td>
<td>10.3 mth</td>
<td>Bevacizumab continual 3.9 mth</td>
<td>0.72 (0.63-0.82)</td>
<td>NS</td>
<td>3</td>
<td>[119]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liposomal doxorubicin +/- trabectedin</td>
<td>OVA-301</td>
<td>2nd line metastatic</td>
<td>PFS stratified for platinum sensitivity/resistance</td>
<td>(sensitive) 7.5 mth</td>
<td>1.7 mth</td>
<td>0.73 (0.56-0.95)</td>
<td>Not improved</td>
<td>2</td>
<td>[120]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olaparib vs placebo</td>
<td>BRCA ovarian cancer in remission</td>
<td>PFS</td>
<td>4.3 mth</td>
<td>6.9 mth</td>
<td>0.18 (0.10-0.31)</td>
<td>NS</td>
<td>Not improved</td>
<td>2</td>
<td>[121]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medication</td>
<td>Trial name</td>
<td>Setting</td>
<td>Primary outcome</td>
<td>PFS control</td>
<td>PFS gain</td>
<td>PFS HR</td>
<td>OS control</td>
<td>OS gain</td>
<td>OS HR</td>
<td>QoL</td>
<td>Toxicity</td>
<td>ESM0-MCBS</td>
<td>ref</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------</td>
<td>----------------------------------------------</td>
<td>----------------------------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>-------------</td>
<td>---------</td>
<td>-------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
</tr>
<tr>
<td>Pazopanib vs sunitinib</td>
<td>COMPARZ</td>
<td>1st line metastatic RCC with clear cell component</td>
<td>PFS non inferiority</td>
<td>9.5 mth</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>[122]</td>
</tr>
<tr>
<td>Temsirolimus vs interferon vs combined</td>
<td></td>
<td>1st line poor-prognosis metastatic RCC</td>
<td>OS</td>
<td>7.3 mth</td>
<td>3.3 mth</td>
<td>0.73 (0.58-0.92)</td>
<td></td>
<td></td>
<td>Reduced</td>
<td></td>
<td>4</td>
<td>[123]</td>
<td></td>
</tr>
<tr>
<td>Sunitinib vs interferon</td>
<td></td>
<td>1st line metastatic</td>
<td>PFS crossover allowed</td>
<td>5mth</td>
<td>6mth</td>
<td>0.42 (0.32-0.54)</td>
<td>21.8 mth</td>
<td>4.6 mth</td>
<td>NS</td>
<td>Improved</td>
<td>4</td>
<td>[124][125]</td>
<td></td>
</tr>
<tr>
<td>Axitinib vs sorafenib</td>
<td>AXIS</td>
<td>Previously treated metastatic RCC</td>
<td>PFS</td>
<td>4.7 mth</td>
<td>2.0 mth</td>
<td>0.66 (0.55-0.81)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>[126]</td>
<td></td>
</tr>
<tr>
<td>Sorafenib vs placebo</td>
<td>TARGET</td>
<td>2nd line locally advanced or metastatic RCC</td>
<td>OS</td>
<td>2.8 mth</td>
<td>2.7 mth</td>
<td>0.44 (0.35-0.55)</td>
<td>15.9 mth</td>
<td>3.4 mth</td>
<td>0.77 (0.63-0.95)</td>
<td></td>
<td>3</td>
<td>[127]</td>
<td></td>
</tr>
<tr>
<td>Everolimus vs placebo</td>
<td>RECORD1</td>
<td>2nd or 3rd line after TKI metastatic RCC</td>
<td>PFS crossover allowed</td>
<td>1.9 mth</td>
<td>2.1 mth</td>
<td>0.30 (0.22-0.40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>[128]</td>
<td></td>
</tr>
<tr>
<td>Pazopanib vs placebo</td>
<td></td>
<td>2nd line locally advanced or metastatic RCC</td>
<td>PFS crossover allowed</td>
<td>4.2 mth</td>
<td>5 mth</td>
<td>0.46 (0.34-0.62)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>[129]</td>
<td></td>
</tr>
<tr>
<td>Interferon +/- bevacizumab</td>
<td>AVOREN</td>
<td>1st line metastatic RCC with clear cell</td>
<td>PFS</td>
<td>5.4 mth</td>
<td>4.6 mth</td>
<td>0.63 (0.52-0.75)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>[130]</td>
<td></td>
</tr>
<tr>
<td>Interferon +/- bevacizumab</td>
<td>CALGB 90206</td>
<td>1st line metastatic RCC with clear cell</td>
<td>OS amended to PFS</td>
<td>5.2 mth</td>
<td>3.3 mth</td>
<td>0.71 (0.66-0.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>[131]</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: Field testing ESMO-MCBS v1.0: Renal Cell Cancer
Table 9: Field testing ESMO-MCBS v1.0: Sarcoma

<table>
<thead>
<tr>
<th>Medication</th>
<th>Trial name</th>
<th>Setting</th>
<th>Primary outcome</th>
<th>PFS control</th>
<th>PFS gain</th>
<th>PFS HR</th>
<th>OS control</th>
<th>OS gain</th>
<th>OS HR</th>
<th>QoL</th>
<th>Toxicity</th>
<th>ESMO-MCBS</th>
<th>ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imatinib 1 year vs placebo</td>
<td>ACOSOG Z9001</td>
<td>Adjuvant for GIST</td>
<td>RFS stratified for risk</td>
<td>1 year RFS</td>
<td>83%</td>
<td>13%</td>
<td>0.35 (0.22-0.53)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>[132]</td>
</tr>
<tr>
<td>3 vs 1 year imatinib</td>
<td>SSG XVIII</td>
<td>Adjuvant for high risk GIST</td>
<td>5 year RFS</td>
<td>48%</td>
<td>18%</td>
<td>0.46 (0.32-0.65)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>[133]</td>
<td></td>
</tr>
<tr>
<td>Sunitinib vs placebo</td>
<td></td>
<td>Advanced GIST 2nd line after imatinib</td>
<td>TTP crossover allowed</td>
<td>6.4 wk</td>
<td>16.9 wk</td>
<td>0.33 (0.23-0.47)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>[134]</td>
<td></td>
</tr>
<tr>
<td>Regorafenib vs placebo</td>
<td>GRID</td>
<td>3rd line after imatinib and sunitinib</td>
<td>PFS crossover allowed</td>
<td>0.9 mth</td>
<td>3.7 mth</td>
<td>0.27 (0.19-0.39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>[135]</td>
<td></td>
</tr>
<tr>
<td>Pazopanib vs placebo</td>
<td>PALETTE</td>
<td>Previously treated non-GIST metastatic soft tissue sarcoma</td>
<td>PFS</td>
<td>1.6 mth</td>
<td>3 mth</td>
<td>0.31 (0.24-0.40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>[136]</td>
<td></td>
</tr>
<tr>
<td>Ridaforolimus vs placebo</td>
<td>SUCCEED</td>
<td>Sarcoma after response or stable disease with 1st line treatment</td>
<td>PFS</td>
<td>14.6 wk</td>
<td>3.1 wk</td>
<td>0.72 (0.61-0.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>[137]</td>
<td></td>
</tr>
<tr>
<td>Medication</td>
<td>Trial name</td>
<td>Setting</td>
<td>Primary outcome</td>
<td>PFS control</td>
<td>PFS gain</td>
<td>PFS HR</td>
<td>OS control</td>
<td>OS gain</td>
<td>OS HR</td>
<td>QoL</td>
<td>Toxicity</td>
<td>ESMO-MCBS</td>
<td>ref</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>---------</td>
<td>----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>------------</td>
<td>---------</td>
<td>--------</td>
<td>-----</td>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
</tr>
<tr>
<td>Ipilimumab +/- glycoprotein 100 vaccine vs vaccine alone</td>
<td></td>
<td>Previously treated metastatic</td>
<td>OS</td>
<td>6.4 mth</td>
<td>3.7 mth</td>
<td>0.69 (0.56-0.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Vemurafenib vs dacarbazine</td>
<td>BRIM-3</td>
<td>1st line or 2nd line after IL-2 metastatic with BRAF V600E mutation</td>
<td>PFS and OS</td>
<td>1.6 mth</td>
<td>4.7 mth</td>
<td>0.26 (0.20-0.33)</td>
<td>9.7 mth</td>
<td>3.9 mth</td>
<td>0.70 (0.57-0.87)</td>
<td>4</td>
<td>[66, 67]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trametinib vs dacarbazine or paclitaxel</td>
<td>METRIC</td>
<td>Unresectable or metastatic with BRAF V600E mutation</td>
<td>PFS (crossover allowed)</td>
<td>1.5 mth</td>
<td>3.3 mth</td>
<td>0.45 (0.33-0.63)</td>
<td>6 mth: 67%</td>
<td>14.00%</td>
<td>0.70 (0.57-0.87)</td>
<td>Improved</td>
<td>4*</td>
<td>[139, 140]</td>
<td></td>
</tr>
<tr>
<td>Dabrafenib +/- trametinib</td>
<td></td>
<td>1st line unresectable or metastatic with BRAF V600E mutation</td>
<td>Toxicity, PFS</td>
<td>5.8 mth</td>
<td>3.6 mth</td>
<td>0.30 (0.25-0.62)</td>
<td></td>
<td></td>
<td></td>
<td>12% reduction skin cancer</td>
<td>4</td>
<td>[141]</td>
<td></td>
</tr>
<tr>
<td>Dabrafenib vs dacarbazine</td>
<td></td>
<td>1st line unresectable or metastatic with BRAF V600E mutation</td>
<td>PFS (crossover allowed)</td>
<td>2.7 mth</td>
<td>2.1 mth</td>
<td>0.30 (0.18-0.51)</td>
<td></td>
<td></td>
<td></td>
<td>Improved</td>
<td>4</td>
<td>[142, 143]</td>
<td></td>
</tr>
<tr>
<td>Dabrafenib + trametinib vs vemurafenib</td>
<td></td>
<td>1st line unresectable or metastatic with BRAF V600E mutation</td>
<td>OS</td>
<td>7.3 mth</td>
<td>4.1 mth</td>
<td>0.69 (0.53-0.89)</td>
<td>1 year: 65%</td>
<td>7%</td>
<td>0.69 (0.53-0.89)</td>
<td>17% reduction skin cancer</td>
<td>4*</td>
<td>[144]</td>
<td></td>
</tr>
<tr>
<td>Vemurafenib +/- cobimetinib</td>
<td></td>
<td>1st line unresectable or metastatic with BRAF V600E mutation</td>
<td>PFS</td>
<td>6.2 mth</td>
<td>3.7 mth</td>
<td>0.51 (0.39-0.68)</td>
<td>9 mth: 73%</td>
<td>8%</td>
<td></td>
<td>9% reduction skin cancer</td>
<td>4*</td>
<td>[145]</td>
<td></td>
</tr>
<tr>
<td>Dacarbazine +/- nivolumab</td>
<td></td>
<td>1st line unresectable or metastatic BRAF-V600-WT</td>
<td>OS</td>
<td>2.2 mth</td>
<td>2.9 mth</td>
<td>0.43 (0.34-0.56)</td>
<td>10.8 mth</td>
<td>6+ mth</td>
<td>0.42 (0.25-0.73)</td>
<td></td>
<td>4*</td>
<td>[146]</td>
<td></td>
</tr>
<tr>
<td>Dacarbazine +/- ipilimumab</td>
<td></td>
<td>1st line metastatic</td>
<td>OS (crossover allowed)</td>
<td></td>
<td></td>
<td></td>
<td>3 years survival: 12.2%</td>
<td>8.60%</td>
<td>0.33 (0.24-0.53)</td>
<td></td>
<td></td>
<td>3</td>
<td>[51, 147]</td>
</tr>
</tbody>
</table>

* immature survival data
Table 11: Fieldtesting ESMO-MCBS v-1.0: Pancreatic Cancer

<table>
<thead>
<tr>
<th>Medication</th>
<th>Trial name</th>
<th>Setting</th>
<th>Primary outcome</th>
<th>PFS control</th>
<th>PFS gain</th>
<th>PFS HR</th>
<th>OS control</th>
<th>OS gain</th>
<th>OS HR</th>
<th>QoL</th>
<th>Toxicity</th>
<th>ESMO-MCBS</th>
<th>ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOLFIRINOX vs gemcitabine</td>
<td>1st line advanced or metastatic, good PS</td>
<td>OS (crossover allowed)</td>
<td>OS</td>
<td>6.8 mth</td>
<td>4.4 mth</td>
<td>0.57</td>
<td>0.45-0.73</td>
<td>Delayed deterioration</td>
<td>5     [148]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gemcitabine +/- Nab paclitaxel</td>
<td>1st line advanced or metastatic, good PS</td>
<td>OS</td>
<td>OS</td>
<td>6.7 mth</td>
<td>1.8 mth</td>
<td>0.72</td>
<td>0.61-0.83</td>
<td>5% gain at 24 mth</td>
<td>3       [149]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gemcitabine +/- erlotinib</td>
<td>1st line advanced or metastatic</td>
<td>OS</td>
<td>OS</td>
<td>5.9 mth</td>
<td>0.3 mth</td>
<td>0.82</td>
<td>0.69-0.99</td>
<td></td>
<td>1       [150].</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 12: Field testing ESMO-MCBS v1.0: Gastro-oesophageal Cancer

<table>
<thead>
<tr>
<th>Medication</th>
<th>Trial name</th>
<th>Setting</th>
<th>Primary outcome</th>
<th>PFS control</th>
<th>PFS gain</th>
<th>PFS HR</th>
<th>OS control</th>
<th>OS gain</th>
<th>OS HR</th>
<th>QoL</th>
<th>Toxicity</th>
<th>ESMO-MCBS</th>
<th>ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery +/- perioperative epirubicin, cisplatin, 5FU</td>
<td>ISRCTN 93793971</td>
<td>Gastric or distal oesophagus stage II-III</td>
<td>OS</td>
<td>5 years</td>
<td>23%</td>
<td></td>
<td>13%</td>
<td>0.66 (0.53-0.81)</td>
<td></td>
<td>A</td>
<td>[151]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgery +/- perioperative cisplatin/5FU</td>
<td></td>
<td>Gastric or distal oesophagus stage II-III</td>
<td>OS</td>
<td>5 years</td>
<td>24%</td>
<td></td>
<td>14%</td>
<td>0.69 (0.50-0.95)</td>
<td></td>
<td>A</td>
<td>[152]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramucirumab vs placebo</td>
<td>REGARD</td>
<td>2nd line gastro-oesophageal or gastric cancer after cisplatin/5FU</td>
<td>OS</td>
<td>3.2 mth</td>
<td>2 mth</td>
<td></td>
<td>0.78 (0.60-0.99)</td>
<td></td>
<td>2</td>
<td></td>
<td>[153]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix I

**ESMO Magnitude of Clinical Benefit Scale v1.0**

**Form 1: for new approaches to adjuvant therapy or new potentially curative therapies**

<table>
<thead>
<tr>
<th>Name of study:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Study drug:</td>
<td>Indication:</td>
</tr>
<tr>
<td>First author:</td>
<td>Year:</td>
</tr>
<tr>
<td>Name of evaluator:</td>
<td></td>
</tr>
</tbody>
</table>

**Grade A**

- >5% improvement of survival at ≥ 3 years follow-up
- Improvements in DFS alone (primary endpoint) (HR <0.65) in studies without mature survival data

**Grade B**

- ≥ 3% but ≤ 5% improvement at ≥ 3 years follow-up
- Improvement in DFS alone (primary endpoint) (HR 0.65 - 0.8) without mature survival data
- Non inferior OS or DFS with reduced treatment toxicity or improved Quality of Life (with validated scales)
- Non inferior OS or DFS with reduced treatment cost as reported study outcome (with equivalent outcomes and risks)

**Grade C**

- <3% improvement of survival at ≥ 3 years follow-up
- Improvement in DFS alone (primary endpoint) (HR >0.8) in studies without mature survival data

**Magnitude of clinical benefit grade (highest grade scored)**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Form 2a: for therapies that are not likely to be curative with primary endpoint of OS

| Name of study: |
| Study drug: | Indication: |
| First author: | Year: | Journal: |
| Name of evaluator: |

**IF median OS with the standard treatment is < 1 year**

**Grade 4**

HR ≤ 0.65 AND Gain ≥ 3 months

Increase in 2 year survival alone ≥ 10%

**Grade 3**

HR ≤ 0.65 AND Gain 2.5-2.9 months

Increase in 2 year survival alone 5 - <10%

**Grade 2**

HR > 0.65-0.70 OR Gain 1.5-2.4 months

Increase in 2 year survival alone 3 - <5%

**Grade 1**

HR > 0.70 OR Gain <1.5 months

Increase in 2 year survival alone <3%

**Preliminary magnitude of clinical benefit grade (highest grade scored)**

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
</table>

Mark with X if relevant
Quality of Life assessment / grade 3-4 toxicities assessment*

<table>
<thead>
<tr>
<th>Does secondary endpoint quality of life show improvement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Are there statistically significantly less grade 3-4 toxicities impacting on daily well-being*</td>
<td></td>
</tr>
</tbody>
</table>

*This does not include alopecia, myelosuppression, but rather chronic nausea, diarrhoea, fatigue, etc.

Adjustments
Upgrade 1 level if improved quality of life and/or less grade 3-4 toxicities impacting daily well-being are shown

<table>
<thead>
<tr>
<th>Final adjusted magnitude of clinical benefit grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
</tbody>
</table>
**IF median OS with the standard treatment > 1 year**

<table>
<thead>
<tr>
<th>Grade 4</th>
<th>Mark with X if relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR ≤ 0.70 AND Gain ≥ 5 months</td>
<td></td>
</tr>
<tr>
<td>Increase in 3 year survival alone ≥ 10%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HR ≤ 0.70 AND Gain 3-4.9 months</td>
<td></td>
</tr>
<tr>
<td>Increase in 3 year survival alone 5 - &lt;10%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HR &gt; 0.70-0.75 OR Gain 1.5-2.9 months</td>
<td></td>
</tr>
<tr>
<td>Increase in 3 year survival alone 3 - &lt;5%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HR &gt; 0.75 OR Gain &lt;1.5 months</td>
<td></td>
</tr>
<tr>
<td>Increase in 3 year survival alone &lt;3%</td>
<td></td>
</tr>
</tbody>
</table>

**Preliminary magnitude of clinical benefit grade (highest grade scored)**

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Quality of Life assessment /grade 3-4 toxicities assessment***

<table>
<thead>
<tr>
<th>Does secondary endpoint quality of life show improvement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Are there statistically significantly less grade 3-4 toxicities impacting on daily well-being*</td>
<td></td>
</tr>
</tbody>
</table>

*This does not include alopecia, myelosuppression, but rather chronic nausea, diarrhoea, fatigue, etc.*
**Adjustments**

Upgrade 1 level if improved quality of life and/or less grade 3-4 toxicities impacting daily well-being are shown

<table>
<thead>
<tr>
<th>Final adjusted magnitude of clinical benefit grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

---

Downloaded from http://annonc.oxfordjournals.org/ by guest on May 30, 2015
Evaluation form 2b: for therapies that are not likely to be curative with primary endpoint PFS

<table>
<thead>
<tr>
<th>Name of study:</th>
<th>Study drug:</th>
<th>Indication:</th>
</tr>
</thead>
<tbody>
<tr>
<td>First author:</td>
<td>Year:</td>
<td>Journal:</td>
</tr>
<tr>
<td>Name of evaluator:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IF with median PFS with standard treatment \(< 6\) months

<table>
<thead>
<tr>
<th>Grade</th>
<th>Condition</th>
<th>Mark with X if relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 3</td>
<td>HR (\leq 0.65) AND Gain &gt; 1.5 months</td>
<td></td>
</tr>
<tr>
<td>Grade 2</td>
<td>HR (&lt; 0.65) BUT Gain &lt; 1.5 months</td>
<td></td>
</tr>
<tr>
<td>Grade 1</td>
<td>HR &gt; 0.65</td>
<td></td>
</tr>
</tbody>
</table>

Preliminary magnitude of clinical benefit grade (highest grade scored)

<table>
<thead>
<tr>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
</table>

Toxicity assessment

Is the new treatment associated with a statistically significant incremental rate of:

- «toxic» death > 2%
- cardiovascular Ischemia > 2%

Mark with X if relevant
hospitalization for «toxicity» > 10%

excess rate of severe CHF > 4%

grade 3 neurotoxicity > 10%

severe other irreversible or long lasting toxicity > 2% please specify:

(INcremental rate refers to the comparison versus standard therapy in the control arm)

**Quality of life / grade3-4 toxicities assessment**

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was quality of life (QoL) evaluated as secondary outcome?</td>
<td></td>
</tr>
<tr>
<td>Does secondary endpoint quality of life show improvement</td>
<td></td>
</tr>
<tr>
<td>Are there statistically significantly less grade 3-4 toxicities impacting on daily well-being*</td>
<td></td>
</tr>
</tbody>
</table>

*This does not include alopecia, myelosuppression, but rather chronic nausea, diarrhoea, fatigue, etc.

**Adjustments**

- **a)** Downgrade 1 level if there is one or more of the above incremental toxicities associated with the new drug
- **b)** Upgrade 1 level if improved quality of life or if less grade 3-4 toxicities that bother patients are demonstrated
- **c)** When OS as secondary endpoint shows improvement, it will prevail and the new scoring will be done according to form 2a
- **d)** Downgrade 1 level if the drug ONLY leads to improved QoL assessment does not demonstrate improved QoL

**Final, toxicity and QoL adjusted, magnitude clinical benefit grade**

<table>
<thead>
<tr>
<th>Grade</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Highest magnitude clinic benefit grade that can be achieved Grade 4.
IF median PFS with standard treatment > 6 months

<table>
<thead>
<tr>
<th>Grade 3</th>
<th>Mark with X if relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR ≤ 0.65 AND Gain &gt; 3 months</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 2</th>
<th>Mark with X if relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR &lt; 0.65 BUT Gain &lt; 3 months</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 1</th>
<th>Mark with X if relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR &gt; 0.65</td>
<td></td>
</tr>
</tbody>
</table>

Preliminary magnitude of clinical benefit grade (highest grade scored)

<table>
<thead>
<tr>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
</table>

Toxicity assessment

Is the new treatment associated with a statistically significant incremental rate of:

- «toxic» death > 2%
- cardiovascular Ischemia > 2%
- hospitalization for «toxicity» > 10%
- excess rate of severe CHF > 4%
- grade 3 neurotoxicity > 10%
- severe other irreversible or long lasting toxicity > 2% please specify:

(Incremental rate refers to the comparison versus standard therapy in the control arm)

Quality of life/ grade3-4 toxicities assessment

Was quality of life (QoL) evaluated as secondary outcome?
<table>
<thead>
<tr>
<th>Does secondary endpoint quality of life show improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are there statistically significantly less grade 3-4 toxicities impacting on daily well-being*</td>
</tr>
</tbody>
</table>

*This does not include alopecia, myelosuppression, but rather chronic nausea, diarrhea, fatigue, etc.

**Adjustments**

- e) Downgrade 1 level if there is one or more of the above incremental toxicities associated with the new drug
- f) Upgrade 1 level if improved quality of life or if less grade 3-4 toxicities that bother patients are demonstrated
- g) When OS as secondary endpoint shows improvement, it will prevail and the new scoring will be done according to form 2a
- h) Downgrade 1 level if the drug ONLY leads to improved QoL assessment does not demonstrate improved QoL

**Final, toxicity and QoL adjusted, magnitude clinical benefit grade**

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
</table>

Highest magnitude clinical benefit grade that can be achieved Grade 4.
Evaluation form 2c: for therapies that are not likely to be curative with primary endpoint other than OS or PFS or equivalence studies

| Name of study: |
| Study drug: | Indication: |
| First author: | Year: | Journal: |
| Name of evaluator: |

Primary outcome is Toxicity or Quality of life AND Non-inferiority Studies

**Grade 4**

Reduced toxicity or improved QoL (using validated scale) with evidence for statistical non inferiority or superiority in PFS/OS

**Grade 3**

Improvement in some symptoms (using a validated scale) BUT without evidence of improved overall QoL

Primary outcome is Response Rate

**Grade 2**

RR is increased $\geq 20\%$ but no improvement in toxicity/QoL/PFS/OS

**Grade 1**

RR is increased $< 20\%$ but no improvement in toxicity/QoL/PFS/OS

**Final magnitude of clinical benefit grade**

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
</table>
Appendix II

Acknowledgements

Fabrice Andre, France
Dirk Arnold, Germany
Paolo A. Ascierto, Italy
Stefan Bielack, Germany
Jean Yves Blay, France
Federico Cappuzzo, Italy
Fatima Cardoso, Portugal
Andrés Cervantes, Spain
Fortunato Ciardiello, Italy
Alan Stuart Coates, Australia
Karina Dahl Steffensen, Denmark
Theo M. De Reijke, The Netherlands
Jean-Yves Douillard, France
Reinhard Dummer, Switzerland
Tim Eisen, UK
Enriqueta Felip, Spain
Constantine Gatsonis, USA
Heikki Joensuu, Finland
Ian Judson, UK
Vesa Kataja, Finland
Roberto Labianca, Italy
Jonathan A. Ledermann, UK
Sumithra J. Mandrekar, USA
Stefan Michiels, France
Mansoor Raza Mirza, Denmark
Mustafa Özgüroğlu, Turkey
Chris Parker, UK
Camillo Porta, Italy
Noemi Reguart, Spain
Daniel J Sargent, USA
Elżbieta Senkus, Poland
Cristiana Sessa, Switzerland
Kirsten Sundby Hall, Norway
Josep Tabernero, Spain
Dongsheng Tu, Canada
Johan Vansteenkiste, Belgium