

## 01. There are 6 forms

### Evaluation form 1a: for RCTs evaluating new approaches to new potentially curative therapies

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.

### Evaluation form 1b: for single arm therapies with curative intent and de-escalation studies

### Evaluation form 2a: for therapies that are not likely to be curative with primary endpoint of OS with separate sheets for:

- IF median OS with the standard treatment is <12 months
- IF median OS with the standard treatment ≥12 - <24 months
- IF median OS with the standard treatment ≥24 - <36 months
- IF median OS with the standard treatment ≥36 months

### Evaluation form 2b: for therapies that are not likely to be curative with primary endpoint PFS with separate sheets for:

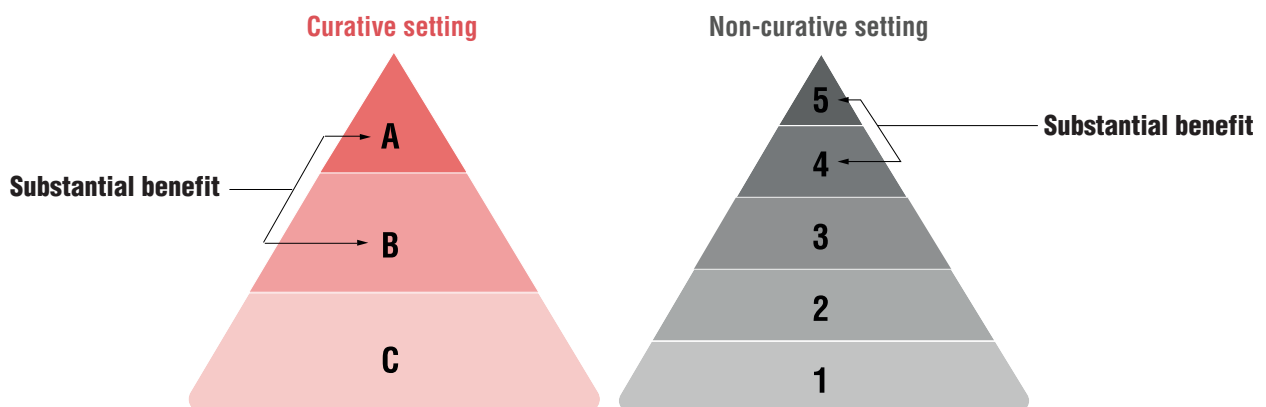
- IF median PFS with standard treatment <6 months
- IF median PFS with standard treatment ≥6 - <12 months
- IF median PFS with standard treatment ≥12 months

### Evaluation form 2c: for therapies that are not likely to be curative with primary endpoint other than OS or PFS or equivalent (non-inferiority) studies

### Evaluation form 3: for single-arm studies in “orphan diseases” and for diseases with “high unmet need” when primary outcome is PFS or ORR

## 02. ESMO-MCBS scores

The highest possible grades of the ESMO-MCBS:H are A in the curative setting, and 5 for non-curative indications. Grades of A and B in the curative setting and 5 and 4 in the non-curative setting indicate substantial clinical benefit. New therapies demonstrating substantial clinical benefit justify rapid consideration for reimbursement.



### 3. Eligibility for application of the ESMO-MCBS:H

The ESMO-MCBS:H can be applied to comparative outcome studies evaluating in haematological malignancies the relative benefit of treatments using endpoints of survival, QoL and conventional surrogate endpoints (for example, DFS, DFI, EFS, PFS, RFS, TTR, TTP) or treatment toxicity.

- Eligible studies can have either a randomised or comparative cohort design or a meta-analysis that report statistically significant benefit from anyone, or more of the evaluated outcomes.
- Single arm studies with curative intent, including de-escalation studies, and studies in non-curative settings that have resulted in licensing can be evaluated.
- When more than one study has evaluated a single clinical question, results derived from well powered registration trials should be given priority.
- Evidence of benefit derived from meta-analyses can be graded only for meta-analyses and systemic reviews compliant with PRISMA standards<sup>1</sup>. These include requirements for:
  - a. Plausible question based on randomised evidence using an exhaustive review of relevant studies
  - b. Evaluation of consistency across studies regarding population of interest
  - c. Relevant patient characteristics and control arm, coupled with lack of bias (publication, selective reporting)
  - d. Exploration of heterogeneity and clear description of limitations.

---

### 4. Analysis of phase III trials

- Adequately powered studies showing statistically significant improvement in the primary outcome (defined by  $P < 0.050$  or less if that is a predefined threshold).
- Careful analysis of the “control arm” and identification of endpoints.
- Check subgroup analysis:
  - a. Studies with pre-planned subgroup analyses with a maximum of 3 subgroups can be graded (provided there is adjustment for multiple comparisons).
  - b. When statistically significant results are reported for any subgroup, then each of these should be graded separately.
  - c. Subgroups not showing statistically significant results are not graded.
  - d. Except for studies that incorporate collection of biologic 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.
  - e. Claims of benefit based on analyses contravening these statistical constraints are not scoreable (even when they are the basis of regulatory authority approval).

---

### 05. More than one outcome may be applicable

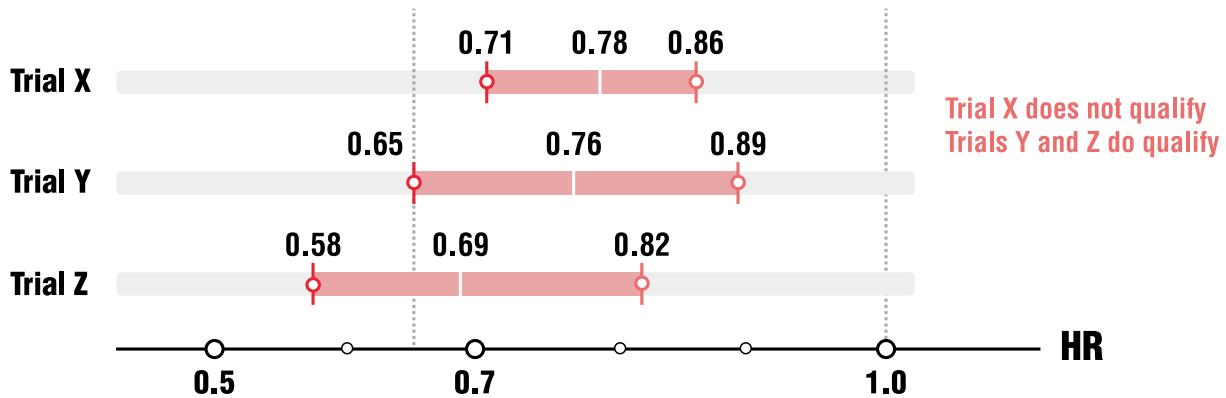
The statistical significance of secondary outcomes is determined by the same criteria as for primary outcomes (i.e. defined by  $P < 0.050$  or less if that is a predefined threshold).

---

DFI, disease free interval; DFS, disease-free survival; EFS, event free survival; ESMO-MCBS:H, European Society for Medical Oncology-Magnitude of Clinical Benefit Scale for Haematological Malignancies; ORR, overall response rate; OS, overall survival; PSF, progression-free survival; QoL, quality of Life; RFS, relapse free survival; TTP, time to progression; TTR, time to relapse.

<sup>1</sup> Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009; 6(7): e1000097.

06. For a required hazard ratio (HR), not the point estimate but the lower limit of 95% confidence interval (CI) estimated based on the observed HR in the trial should encompass the required HR



Example: for threshold set at  $HR \leq 0.65$  it is the lower limit of the 95% CI which has to be  $\leq 0.65$

## 7. In studies with curative intent

- In evaluation of DFS (or RFS, TTP and EFS):
  - Note time point of evaluation (in months or years)
  - Indicate if specific outcome TTP, DFS, iDFS (invasive DFS)
  - Maturity of survival data may be protocol defined or, if not defined, determined by the specific clinical entity. Examples:

Disease	Follow-up for OS data maturity
AML, ALL, high grade lymphoma	5 years
MM, follicular lymphoma	8-10 years

- In cases where OS data maturity has not yet been reached and both OS and DFS are potentially scoreable, the higher score prevails
- Scores are annotated for toxicity:
  - AT: indicates high prevalence of acute transient side effects impacting daily well-being. All curative therapies incorporating autologous or allogeneic bone marrow or stem cell transplant are annotated AT
  - PT: indicates high prevalence of persistent and chronic side effects and late side effects that impact QoL. All curative therapies incorporating allogeneic bone marrow or stem cell transplant are annotated PT due to graft vs host disease

08. In instances when the median of the control arm is reached and the relative benefit gain (HR) is significant, the median of the experimental arm is estimated on the basis of control arm (months) divided by the point estimate of the HR

AML, acute myeloid leukaemia; ALL, acute lymphoblastic leukaemia; AT, acute toxicity; CI, confidence interval; DFS, disease-free survival; HR, hazard ratio; iDFS, invasive disease-free survival; EFS, event free survival; MM, multiple myeloma; OS, overall survival; PT, persistent toxicity; QoL, Quality of Life; RFS, relapse free survival; TTP, time to progression.

**9. In the case of OS in the non-curative setting check for:**

- Reduced toxicity
  - Improvement in QoL
  - Report final adjusted score taken into account toxicity, and QoL when relevant
- 

**10. In case of PFS in the non-curative setting check for:**

- Indicators of toxicity
  - Survival data when also available
  - Early termination with crossover based on planned interim survival analysis
  - Global QoL advantage using validated scale if applicable
  - Report final adjusted grade taken into account toxicity, survival advantage and QoL when applicable
- 

**11. Studies violating the statistical constraints of the ESMO-MCBS:H are not eligible for grading even in circumstances where they are the basis for regulatory body (EMA, FDA, etc) approval: they may be indicated as statistical violations**

---

**12. Studies using parameters that are not evaluable using the ESMO-MCBS:H are indicated not scoreable**