Can we satisfactorily measure the clinical value of new classes of oncology agents with a single summary measure?

Clare Jones, Giles Monnickendam, Mingshu Zhu, Jan McKendrick. PRMA Consulting, Fleet, Hampshire, UK

Background

- The hazard ratio (HR) and difference in median survival are the summary measures of survival experience most commonly specified as endpoints in oncology clinical trials. Landmark survival is also used, although less frequently.
- Most value frameworks use these summary measures of either overall survival (OS) or progression-free survival (PFS) to represent patient benefit in the assessment of the value of new oncology treatments.
- However, if the proportional hazards (PH) assumption does not hold, the interpretation of these summary statistics can become problematic and fail to adequately capture the expected benefit to patients.
- If PH does not apply, Cox model estimates of the HR cannot be considered as a simple average of the ratio of hazards over time. The HR becomes a measure of the between-group difference in survival with indefinite meaning.
- Furthermore, the survival gain of patients at any point in the survival distribution, as measured by median or landmark survival differences, may not reflect survival gains across the full patient population.
- As new classes of oncology therapy are introduced and evaluated against comparators from established classes, the routine assumption of PH may no longer be appropriate. As a result, value frameworks may score therapies on an incomplete picture of actual clinical benefit.

Objective

The objective of this work was to test how satisfactorily the scoring methodology of the ASCO value framework[1] and the ESMO Magnitude of Clinical Benefit Scale[2] capture and represent patient benefit when the PH assumption is not valid.

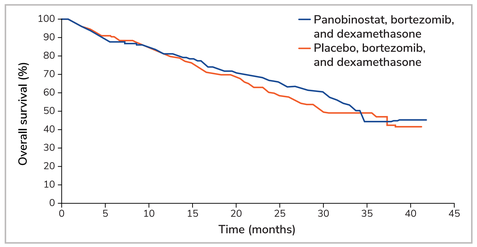
Methods

We identified oncology drugs from the published list of FDA-approved drugs from January 2011 to December 2016[3].

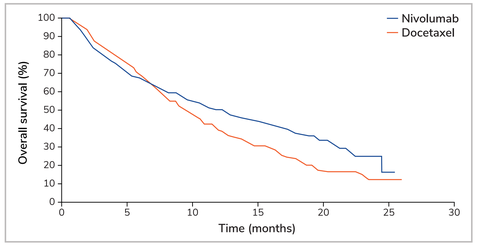
We reviewed published Kaplan–Meier curves and associated summary statistics for survival endpoints from the pivotal trials supporting regulatory approval.

We selected two archetypal examples of distinctly non-PH survival curves for OS:

Example 1: OS for panobinostat in combination with bortezomib and dexamethasone, versus bortezomib and dexamethasone alone, for the treatment of relapsed or refractory multiple myeloma[4]: *survival curves that that initially diverge before re-converging*

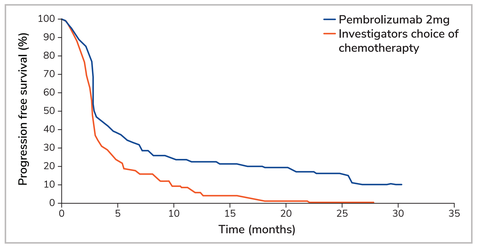


Example 2: OS for nivolumab compared with docetaxel for second-line treatment of advanced non-small cell lung cancer (NSCLC)[5]: *survival curves that track closely together for a period before diverging*



Additionally, we selected an example of non-PH in PFS:

Example 3: PFS for pembrolizumab compared with chemotherapy for the treatment of ipilimumab-refractory melanoma[6]: *steady divergence of survival curves, followed by the emergence of a plateau in one arm of the trial*



The overall scoring methodology for each value framework was reviewed; for each product the ASCO and ESMO scoring for survival benefit was completed and results considered in the context of the survival experience represented by Kaplan–Meier curves.

Results

Table 1 Overview of drug/trial characteristics

	Panobinostat OS (San-Miguel, 2014)	Nivolumab OS	Pembrolizumab PFS
Primary endpoint	PFS	OS	PFS & OS (co-primary)
Indication	Relapsed/refractory multiple myeloma	Second-line non-squamous NSCLC	Ipilimumab refractory melanoma
Treatment	Panobinostat plus bortezomib and dexamethasone	Nivolumab	Pembrolizumab 2mg/kg
Comparator	Placebo plus bortezomib and dexamethasone	Docetaxel	Investigator's choice of chemotherapy
Median survival	33.6 vs 30.4 months	12.2 vs 9.4 months	3.0 vs 2.9 months
HR (95% CI)	0.87 (0.69–1.10)	0.73 (0.59–0.89)	0.58 (0.46–0 .73)

ASCO value framework scoring methodology[2]

- Clinical benefit is scored using one of OS, PFS, or response rate (RR), in that order of priority, regardless of whether the endpoint is the primary outcome measure.
- OS and PFS are always measured by the HR if this is available. For OS, the HR is subtracted from 1 and multiplied by 100; for PFS the HR is subtracted from 1 and multiplied by 80.
- A tail-of-the-curve bonus of up to 20 points (16 points for PFS) is added for an improvement of more than 50% in survival proportion at twice the median survival of the comparator, assuming survival at this point exceeds 20%.

ASCO scores

Panobinostat OS

- Panobinostat achieved a clinical benefit score of 13 based on the published OS data.
- The score is based only on HR, which does not differentiate the different phases of the survival curve with change in the nature of the treatment effect over time:
  - a positive effect of panobinostat which begins to emerge at around 12 months (≈80% patients still alive).
  - no suggestion of further incremental treatment benefit for panobinostat after around 35 months (≈45% patients still alive).
- Analyses with more mature OS data[7] resulted in a score of 6 based on a less favorable HR (0.94, 95% CI 0.78-1.14) despite a greater difference in median OS (40.3 vs 35.8 months).

Nivolumab OS

- Nivolumab achieved a clinical benefit score of 27.
- However, this score may understate the value of the treatment to the group of patients who survive past 7 months where there is a pronounced positive clinical benefit for nivolumab; the HR is influenced by the lack of early treatment differences.

Pembrolizumab PFS

- If the ASCO score was based on PFS data alone (i.e., assuming no OS data were reported), pembrolizumab 2mg/kg would have achieved a clinical benefit score of 34 if the tail of the curve bonus was added the score would be 50.
- However, with OS data (co-primary endpoint), the score achieved was 14, with HR (0.86, 95% CI 0.67-1.10).
- The score does not appropriately differentiate between the potential clinical benefit for patients who may progress early or later.
  - The survival curves suggest there are some patients who progress early and gain little incremental benefit from treatment with pembrolizumab.
  - The plateau that develops in the pembrolizumab PFS curve at around 7 months indicates that patients who progress later derive substantial incremental PFS benefit from treatment.

ESMO MCBS scoring methodology

- A complex hierarchical scoring methodology is used, which grades non-curative treatments based on survival endpoint (e.g., OS, PFS) and a combination of median survival with HR or landmark survival (2 or 3 years).
- Median survival of the comparator treatment and endpoint (OS or PFS) determine the specific grading scale to be used.
- Grading of OS (non-curative treatment) is based on a mix of median difference, HR and landmark survival. High grades require both HR and median criteria to be met, or alternatively landmark survival alone. At lower grades, a "best of" approach from median, HR, or landmark survival determines scores.
- Grading of PFS (non-curative treatment) is less stringent and requires both HR and median criteria to be met for the high score, and the low score is assigned if HR alone is not met.

ESMO MCBS scores

Panobinostat

- If we assume OS is the primary endpoint, panobinostat would have achieved a preliminary MCBS grade of 1 based on median OS of >1 year for the comparator group and a HR >0.7; the grade based on the primary endpoint of PFS would be 3 [median survival: 12.0 vs 8.1 months, HR (0.63, 95% CI 0.52–0.76)].
- Scoring is driven entirely by HR reflecting average benefit and therefore does not differentiate the experience of patients early and late in the survival curve, before and after the convergence of OS Survival curves.
- Analyses with more mature OS data[7] also resulted in a grade of 1.

Nivolumab

- Nivolumab achieved a preliminary MCBS grade of 4, based on median OS of more than 1 year for the comparator group and a difference in 2 year survival of more than 10% (from visual inspection).
- This score is driven by the difference in survival rates at 2 years and captures the value of nivolumab for patients who can expect to survive long enough to respond to immuno-oncologic treatment; however this is only one point of the survival curve and does not reflect the proportion of patients with a poorer prognosis, who are unlikely to achieve an incremental survival benefit.

Pembrolizumab PFS

- Pembrolizumab achieved a preliminary MCBS grade of 2, based on median PFS ≤6 months for the comparator group, HR for PFS ≤0.65 but median PFS difference <1.5 months.
- This score, the result of a low median gain combined with a high HR, could be considered an appropriate "averaging" of value across patients early and late in the survival distribution.
- The grade based on OS (co-primary endpoint) is 3.

Conclusion

- In general, summary measures of survival or value frameworks that reduce survival benefit to single scores will not capture important variations in the magnitude of benefit between patient subgroups and can also change markedly with maturing survival data.
- The ESMO framework appears to offer greater flexibility, relying on the combination of several summary measures to determine scores; however, where hazards are not proportional for OS, the ESMO framework can be driven by a combination of median gain and HRs or by landmark survival analyses at a single time point that are not representative of the survival experience of large proportions of patients.
- In contrast, the ASCO framework consistently prioritizes the HR if this is available, capturing an average patient experience that cannot be easily interpreted and does not represent average treatment effect in the absence of PH. The ASCO tail-of-curve bonus compensates in cases where there is evidence of long-term survival benefit that has a limited impact on HR, due to a small number of observations of long survival duration where data are immature. However, it is applied based on passing a threshold (not graduated) and can have a marked impact on total scores.
- Decision makers should be aware of these limitations of single summary measures when making treatment decisions - value assessment would benefit from more comprehensive and flexible methods that are able to capture the idiosyncrasies of the more complex relationships between survival functions that occur with non-PH.

References

1. Schnipper *et al.*, (2016) J Clin Oncol 34: 2925-34
2. Cherny *et al.*, (2015) Ann Oncol 26: 1547-73
3. FDA. https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm279174.htm (accessed on 28 September 2016)
4. San-Miguel *et al.*, (2014) Lancet Oncol 15: 1195-206
5. Borghaei *et al.*, (2015) N Engl J Med 373: 1627-39
6. Hamid *et al.*, (2016) Annals of Oncology 27: 1107O-1107O
7. San-Miguel *et al.*, (2016) Lancet Haematol 3: e506-e515