MSR146

Proportional hazards in survival analysis: A review of the variation in determining and presenting PH in NICE STAs

Cheah, Z[1]; Westerberg, E[2]; Ho, M[1]; Simons, C[1]; Moura, A[2]

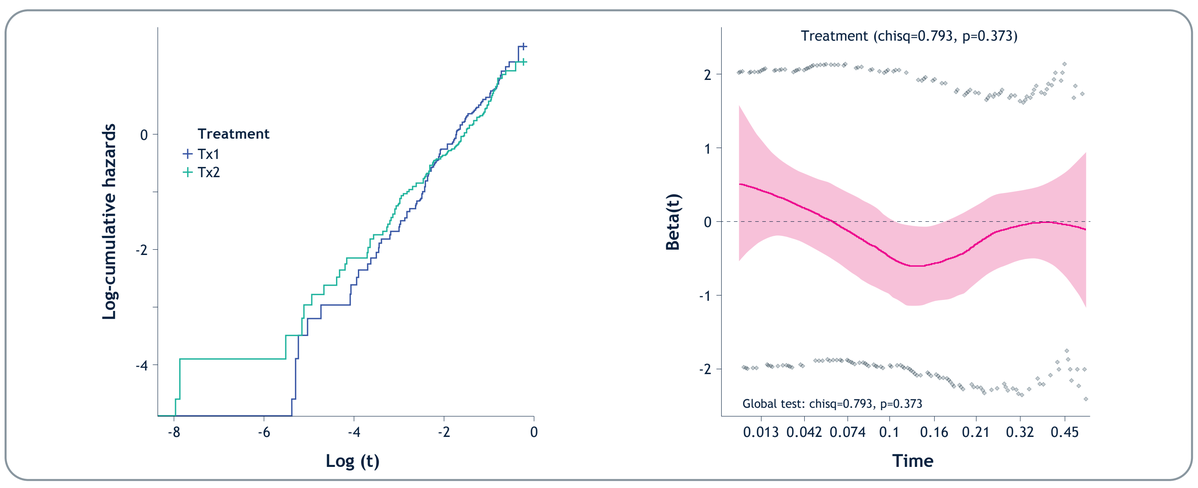
[1]OPEN Health HEOR & Market Access, London, UK
[2]OPEN Health HEOR & Market Access, Rotterdam, The Netherlands

OPEN HEALTH

INTRODUCTION

- Technology appraisals (TAs) submitted to the National Institute for Health and Care Excellence (NICE) describe cost-effectiveness analyses conducted on medical therapies, including survival analysis carried out on time-to-event data.
- A key assumption in survival analysis is the proportional hazards assumption (PHA), which assumes that the hazard ratio between two comparators does not vary with time. The results of this test inform the type of survival analysis that can be carried out on data, which is why the PHA is a crucial one that should be tested.
- NICE technical support document (TSD) 14[1] contains guidance on the methods that should be used for testing the PHA. The recommended PHA testing methods are shown in Figure 1, with both plots generated using simulated data.

Figure 1. Log-cumulative hazards (left) plot and Schoenfeld residuals (right) plot used to test the PHA between two comparators. The Grambsch-Therneau test value is shown at the top and in the bottom right corner of the Schoenfeld residuals plot.



OBJECTIVES

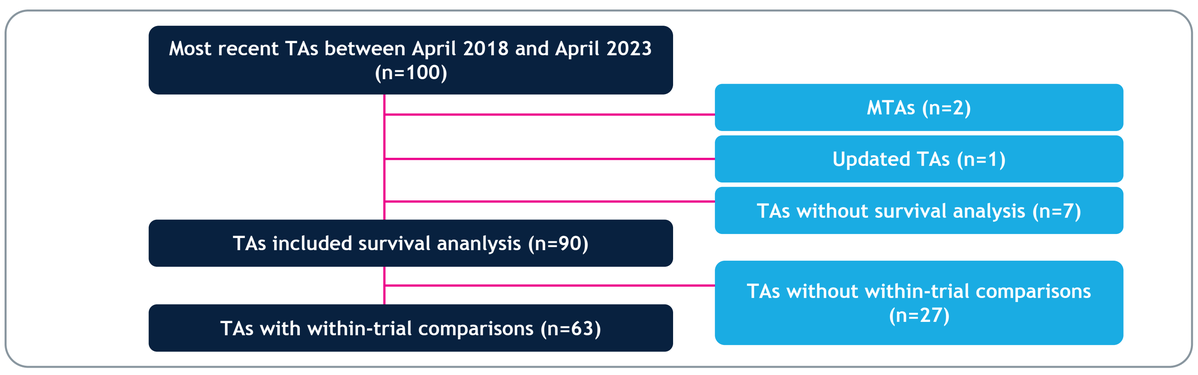
This work aims to explore:

- The variation in reporting of methods of testing for PH in company submissions in TAs
- Statistical methods used to deal with the results from PH testing
- Acceptability of these methods to the NICE external assessment/evidence research group (EAG/ERG) committee

METHODS

- All publicly available NICE single technology assessments (STAs) in oncology that were published on the NICE website from April 2018 up to April 13, 2023, were identified.
  - If an STA contained the word ‘cancer’, ‘carcinoma’, ‘myeloma’, ‘chemotherapy’, ‘lymphoma’, ‘leukaemia’, or ‘melanoma’, it was flagged as an oncology STA in Excel.
  - STAs that did not contain these words were recorded as ‘Not available’ (N/A) and were checked to ensure that no oncology STAs had been missed by applying the algorithm above.

Figure 2. TA selection process



Abbreviations: MTA, multiple technology appraisal; TA, technology appraisal.

- Information on PH tests conducted (and their reporting), their adherence to NICE TSD 14 guidelines[1], as well as EAG/ERG opinion on the testing methods used (with relation to NICE guidelines) were extracted and analysed descriptively by one reviewer, with a sample reviewed by a second reviewer.
  - All authors listed contributed to either the extraction or reviewing of data.
- Out of all the STAs identified, the 100 most recent STAs were screened (2020–2023) and selected.
- Requests for unavailable supporting documents were not made.

RESULTS

PHA testing in company submissions

- Out of the 63 TAs that conducted a within-trial comparison, 79.4% (n=50) of TAs mentioned accounting for the PHA within the submission document.
  - Out of the 13 that did not mention accounting for the PHA within the submission document, 30.8% (n=4) presented PHA test results in their original company submission.
- Out of the 50 TAs that mentioned accounting for the PHA within the submission document, 84.0% (n=42) of submissions presented PHA test results in the original company submission document.
  - Despite mention of accounting for the PHA within the original submission document, 10.0% (n=5) did not present any PHA test results.
  - The remaining 6.0% (n=3) of submissions submitted PHA test results upon EAG/ERG request (Figure 3).

Types of PHA tests in company submissions

- Out of all 63 TAs that conducted a within-trial comparison, 22.2% (n=14) of TAs reported results from all three PHA tests recommended in NICE TSD DSU 14.[1]
- 27.0% (n=17) of submissions presented results from log-cumulative hazard plots only, 25.4% (n=16) of submissions presented both log-cumulative hazard and Schoenfeld residual plots, 6.3% (n=4) of submissions presented both log-cumulative hazard plots and the Grambsch-Therneau test results, and 1.6% of submissions presented either only Grambsch-Therneau test values or Schoenfeld residual plots only (n=1) (Figure 4).

Figure 3. TAs that mentioned and went on to present PHA test results in their submission

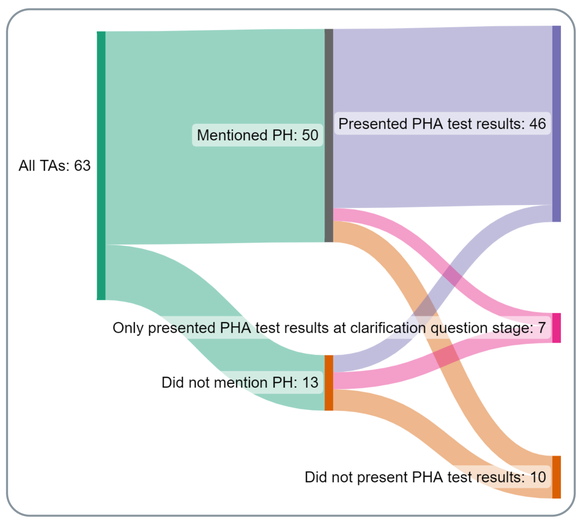
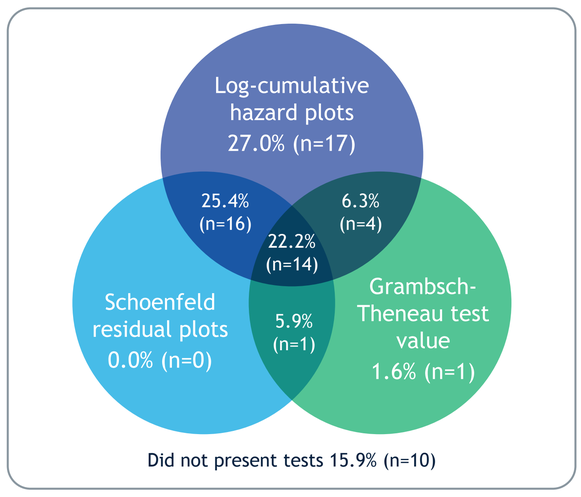


Figure 4. Proportion of the type of PHA test results presented, including test results provided upon EAG/ERG request

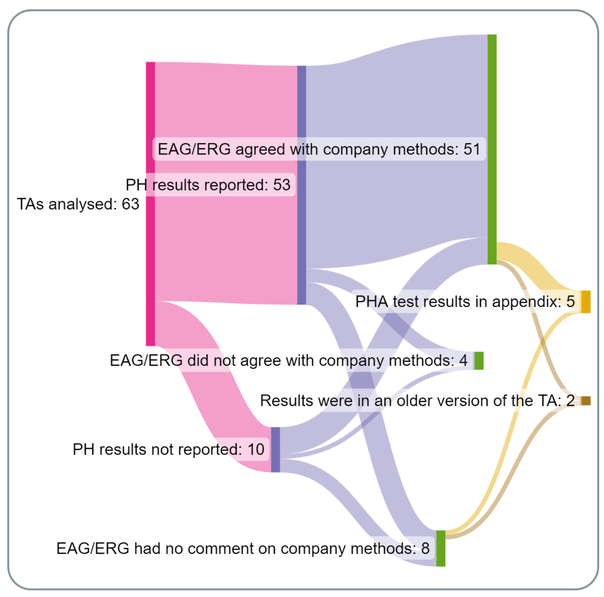


- No submissions ever presented Schoenfeld residual plots by themselves.
- 15.9% (n=10) of submissions did not present any PHA test results.
  - Despite not presenting any PHA test results, 50.0% (n=5) mentioned accounting for the PHA in their submission.
  - Out of the five TAs above, 40.0% (n=2) presented PHA testing results in the company submission appendix.
  - It was unclear whether the remaining 60.0% (n=3) of TAs presented PHA test results; the EAG/ERG did not make any request for PHA test results for these TAs either.

EAG/ERG critique on PHA testing

- Despite the discrepancy in PHA test reporting, PH test results were only critiqued by the EAG/ERG in only 6.3% (n=4) appraisals.
- Out of the four appraisals, only one did not present any form of test results in their submission.
- Reasons outlined for disagreement were that the company’s PHA testing assumption was ad-hoc and not guided by any formal procedure, or that there was limited evidence available to prove that the PH was held (e.g., limited data, or PHA testing was only performed on one outcome and not others).
- The EAG/ERG did not provide any negative critique for 90.0% (n=9) of the TAs that did not present any PHA test results in the original company submission.
  - 66.7 (n=6) obtained a positive critique, and the EAG/ERG did not comment on the remaining 33.3% (n=3) of TAs.
  - 83.3% (n=5) of the six TAs that obtained a positive critique had presented their results in an appendix that was not attached to the original submission document (Figure 5).

Figure 5. EAG/ERG critique on TA PHA test results



CONCLUSIONS

- Despite PHA testing being a required step in parametric model selection for comparative trials as outlined in NICE TSD DSU 14[1], around 15.0% of TAs failed to account for this in their submission as they did not present any form of PHA test results whatsoever.
- Around 7.9% (n=5) of TAs had included PHA test results in the company submission appendix, which was not extracted.
- In all instances where no PHA test results were provided in the original company submission or appendix, the EAG/ERG did not request for these results, failing to adhere to NICE guidelines.
- Out of the 63 TAs that conducted a within-trial analysis, 22.2% (n=14) presented results from all three recommended tests.
  - 57.1% (n=36) presented results from at least two recommended tests.
  - 28.6% (n=18) presented results from one test only.
  - 15.9% (n=10) did not present results from any PHA test.
- Log-cumulative hazard plots were the most frequently displayed test result, with 81.0% (n=51) of TAs presenting this test result in the original company submission.
- This may be because log-cumulative hazard plots are the easiest to produce, understand, and interpret when compared to other PHA test results or plots.
- As results of PHA tests influence the model selection process in survival analysis, guidelines outlined by NICE should be enforced to ensure that all survival analyses conducted in submission have been done so with scientific rigour.
- Despite clear recommendations on PHA testing being outlined by NICE in DSU TSD 14[1], the results from this study show that out of 63 TAs that conducted a within-trial comparison, 22.2% adhered to them fully, 61.9% partially, and 15.9% did not adhere to them at all.
- Therefore, increased adherence to NICE TSD DSU 14[1] guidelines for PHA testing during the submission process is required, ensuring that companies carry out their model selection process correctly.
- Additionally, this study found that despite 10 TAs not presenting any PHA test results in the original company submission, only one TA was critiqued by the EAG/ERG.
- Even when companies only partially adhered to NICE guidelines, the EAG/ERG critique on their PHA testing methods was overwhelmingly positive, as only 6.3% (n=4) TAs out of the 63 were negatively critiqued by the EAG/ERG.
- Thus, the EAG/ERG should increase reinforcement of NICE guidelines, doing so by being more critical of TAs that partially adhered or did not adhere at all to PHA testing guidelines.
- In the future, the results on this study could be expanded upon by investigating whether there was an improvement over adherence to NICE guidelines with regards to PHA test result reporting over time.
- Trends relating to non-adherence to NICE guidelines such as the size of the company making the submission, the size of the trial, disease indication, and the EAG/ERG body reviewing the submission can be explored in the future to determine whether there is a relationship between these factors and observed non-adherence.

REFERENCE: 1. Latimer N. NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. *Sheffield Report Decision Support Unit*. 2011;(0). www.nicedsu.org.uk/NICE DSU TSD Survival analysis. Updated March 2013.v2.pdf

Presented at ISPOR Europe 2023