Network Meta-Analysis in the Presence of Non-proportionality: A Review of NICE Submissions

Jean-Gabriel Le Moine, Emma Hawe, Shaun Abeysinghe

RTI Health Solutions, Manchester, United Kingdom

BACKGROUND

- Survival network meta-analyses (NMAs) are a key component of National Institute for Health and Care Excellence (NICE) submissions in oncology when head-to-head evidence is not available for all comparators.[1]
- Traditionally, NMAs of survival endpoints have used methods to combine hazard ratios (HRs). This relies on the assumption that treatment effect is proportional over time. This assumption, while widely used, may be unrealistic, particularly when evidence comes from many different randomized controlled trials.
- In 2011, Jansen et al.[2] introduced a method based on fractional polynomials (FPs) to conduct an NMA of survival endpoints in absence of proportional hazard (PH).
- To date, the NICE provides limited guidance regarding the most appropriate method to apply when conducting a survival NMA, especially in presence of non-PHs.[3-5]

OBJECTIVE

- Our primary objective was to identify recent NICE submissions that included NMAs of survival endpoints and to assess use of HRs and FP methods and their acceptability.
- Secondary objective was to assess how different published results were for the same comparisons when both methods were applied.

METHODS

- A review of NICE's technology appraisals (TAs) was conducted to identify submissions that included NMAs of survival endpoints. Submissions for oncology products published or updated since June 2016 were considered of interest. A 10% quality check of extracted data was performed by an independent researcher.
- Information about PH assumption testing, NMA methods, and Evidence Review Group (ERG) comments were extracted in a standardised data extraction table.
- A targeted literature review was conducted to identify articles published since 2011 that applied and compared both HR-based and FP-based NMA of survival endpoints.

RESULTS

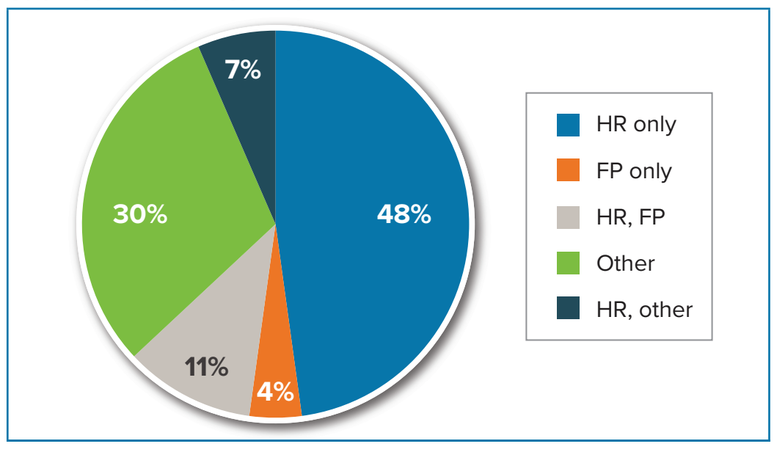
- Searches for relevant TAs identified 78 TAs published or updated since June 2016 regarding oncology products. Of these, 46 (59%) included some form of indirect comparison.
- Results from HR-based NMAs were reported in 65% (30/46) of the included TAs, while FP approaches were used in 15% (7/46) of the TAs (Figure 1), alone (4%, 2/46) or with HR-based methods (11%, 5/46).
- In 37% of the identified TAs (17/46), additional methods other than HR and FP were applied. These included meta-analyses of shape and scale parameters, restricted mean survival time analyses, matched-adjusted indirect comparisons, simulated treatment comparisons, and naïve indirect comparisons.
- The PH assumption was tested in approximately 72% (33/46) of the TA submissions. Methods used to assess PH assumptions were mostly visual inspection of the Kaplan-Meier (KM) curves and log-log plots of cumulative survival versus time, as well as the Q-Q curves, or the Grambsch and Therneau test.
- ERG's comments suggested that, when the PH assumption is not met, an FP-based NMA methodologically is accepted. Selected comments regarding PH assumption and FP approach are presented in Table 1.
- Our targeted search to identify comparisons of results of HR versus FP NMA of survival endpoints found a limited number of publications.
- It was found that the consistency of results generated by HR-versus FP-methods were indication-specific.
  - Analysis of patients with gastric cancer found that FP-based approaches consistently showed reduced survival benefit for treatments compared with traditional HR NMA approaches.[6]
  - Analyses of patients with advanced Soft Tissue Sarcoma found small differences for some treatment comparisons between HR and FP analyses.[7]
  - In first-line metastatic renal cell cancer, median survival was typically lower with FP than PH, with overestimation of long-term progression free survival when PH was incorrectly assumed.[8]
  - An analysis of two case studies found more treatments in the HR NMA showed improvement in survival of at least 3 months compared with FP.[9]

Table 1. ERG Comments on Submissions With FP Methods Since June 2016

Appraisal ID & Title	Last Updated	Summary of Selected ERG Comments on NMA Approach
TA520: Atezolizumab for treating locally advanced or metastatic non-small cell lung cancer after chemotherapy*	May 2018	• HRs based on data from 2 trials relied on a PH assumption. • As demonstrated by the company, this assumption does not hold, so OS and PFS HRs must be interpreted with caution. • The company approach to the ITC was influenced by a range of factors; it is difficult to identify the most appropriate combination of factors to use to generate ITC results. • FP ITC results are difficult to interpret.
TA512: Tivozanib for treating advanced renal cell carcinoma**	Mar 2018	• There are concerns that PH did not hold for PFS and OS. • The company submitted new analyses, using an FP NMA based on a simplified network structure, for OS and PFS. • OS results from the company's FP-based NMA are implausible because they contradicted head-to-head evidence and may be confounded by treatment crossover. • High-level uncertainty around FP-based NMA estimates may outweigh the slight benefits observed in PFS between treatments. • There are several issues with FP-based NMAs submitted for OS and PFS; various checks and additional analyses should be performed to reduce uncertainty in the results.
TA498: Lenvatinib with everolimus for previously treated advanced renal cell carcinoma**	Jan 2018	• FP was added to the submission after a pre-meeting with ERG showing that 2/3 trials violated the PH assumption. • The company tested only a second-order FP approach; further scenarios were conducted by ERG. • ERG's replication of the NMA did not match the company's results, so additional exploratory analyses were conducted.
TA492: Atezolizumab for untreated locally advanced or metastatic urothelial cancer when cisplatin is unsuitable*	Dec 2017	• To enable the formation of a network for NMA, the company employed a simulated treatment comparison to "predict" a matching atezolizumab arm for each comparator study. • Resulting comparisons of atezolizumab against each comparator were then included in an NMA. • The company selected an FP model approach for the NMA because higher-order FP models do not require the assumption of PH. • The approach to NMA was well-suited to the data format available to the company, which consisted of individual patient data for atezolizumab and aggregate population data for the comparators.
TA465: Olaratumab in combination with doxorubicin for treating advanced soft tissue sarcoma**	Aug 2017	• Recommended to compare estimates from reconstructed data and published HR, as a means of validating the reconstruction approach used for HR. • We agree that the PH assumption is not upheld for two studies and that the method using FP is more appropriate.
TA428: Pembrolizumab for treating PD-L1-positive non-small cell lung cancer after chemotherapy**	Sep 2017	• The company used HR-based results in the model, as a constant hazard assumption was found reasonable for OS. • The PH assumption was violated for PFS. The company suggested it was still conservative to use HR-based results. • ERG commented that NMA with FP models would have been a more adequate approach due to the PFS KM curves violating the PH assumption. • In response to the ERG comments, the company updated its model with FP results.
TA366: Pembrolizumab for advanced melanoma not previously treated with ipilimumab**	Sep 2017	• The meta-analysis was not robust and did not reflect a changing HR over time. • ERG concluded that the FP-based NMA results cannot be considered a reliable method for estimating the long-term relative effectiveness for pembrolizumab versus ipilimumab, and hence for either of the other comparators in the evidence network. • Populations in control arms across trials were not comparable. • The analysis to account for treatment switching in key trial was not adequate, and treatment effect for pembrolizumab compared with vemurafenib may have been overestimated.

*FP-based NMA; **HR and FP–based NMA. ITC = Indirect treatment comparison; PFS = progression-free survival.

Figure 1. Indirect Comparison Methods Used Across Identified Submissions for NICE TAs



DISCUSSION

- A range of approaches was used in submissions to consider deviations from non-PH; however, the information about PH is not always reported in publications.
  - In absence of either reported tests or KM curves, it is not possible to assess PH.
  - In a recent review,[10] most publications used Cox PH regression to analyse survival endpoints and failed to validate the PH assumption.
- The survival endpoints must be represented by reproducible KM curves and preferably a table detailing the number of patients at risk to enable FP analysis.
  - Bias may be introduced by considering only trials that publish KM curves.
- Should non-PH be present, the application of HR NMA methods may bias the results.
- This review focused on the standard HR and FP approaches. NMA is a rapidly expanding field, and other approaches are emerging, including Ouwens et al.,[11] and extensions of this include treatment covariate interactions. More recently, Freeman et al.[12] proposed an approach that utilises restricted cubic splines, which may be extended to accommodate issues such as non-PH.

CONCLUSIONS

- HR NMAs remain the most commonly used approach in submissions, and the assumption of non-PH was often (though not always) tested using available KM data.
- For several submissions where some evidence of non-PH was identified, either the manufacturer performed or the ERG requested that a non-PH analysis be undertaken.
- Given the differences in some comparisons when applying the HR and FP analyses, performing analyses that appropriately account for the possibility of non-PH would be recommended.
- Where performing FP analysis reduces the evidence base, performing both HR and FP analysis may be beneficial.
- Insufficient data to evaluate the PH assumption should be identified as a limitation.
- Where possible, non-PH should be assessed over the trial period; however, this doesn't validate the existence of PH after the trial follow-up, and HR applied to extrapolated curves might provide implausible results.
- Simulation work to consider a broader range of scenarios when applying traditional HR NMA to approaches that do not rely on PH would be of use to evaluate the impact of non-PH in NMAs and to evaluate the different techniques that are emerging.

REFERENCES

Please see handout for references.

CONTACT INFORMATION

Jean-Gabriel Le Moine
Associate Director, Data Analysis and Design Strategy
RTI Health Solutions

Phone: +44(0)161.447.6006
E mail: jglemoine@rti.org

Presented at: ISPOR Europe 2018; 10-14 November 2018; Barcelona, Spain