Changing With the Times:

Hazards in Novel Comparative Efficacy Techniques While Maintaining the Path of HTA Informativeness

Technical appropriateness versus usefulness and informativeness of advanced statistical methods for EU Joint Clinical Assessments and NICE HTA submissions

Moderated by: Julie Roiz

Presenters: Tracy Westley, Miranda Cooper, Maiwenn Al

ISPOR EU ISSUE PANEL

20 November 2024

Today's Panel

Julie Roiz

Senior Vice President, Health Economics & Outcomes Research

Lumanity

Maiwenn Al

Associate Professor

Erasmus School of Health Policy & Management

Netherlands

Miranda Cooper

Statistical Science Director, GMA & Payer, Oncology Biometrics,

AstraZeneca

Tracy Westley

Head of Statistical Services, Modelling & Analytics

Lumanity

Three Perspectives for Time-varying ITC estimates

NMA and CEM technical example

Project impact and planning from the manufacturer perspective

HTA historical context including country complexities

Key: CEM, cost-effectiveness model; HTA, health technology appraisal; NMA, network meta-analysis.

Background for NMA technical example

- Oncology advancements in immunotherapies, with novel biological mechanisms of action, drive our need to consider more nuanced statistical methods to flexibly and maximally capture long-term benefits vs. legacy SoC[1]
- Traditional extrapolation and comparison of time-to-event outcomes assume PH between two treatments, resulting in a singular numeric survival difference, that is, a one HR over follow-up
- To enable the hazard ratios to change over time, emerging methods for comparison, used to inform the lifetime horizons of CEM, are being tested and published in current literature
- These methods extend to ITCs including NMA
- For any of these time-varying methods, detailed HTA guidance for ITC model selection, including transparency of past applications in CEM, is limited
- We demonstrate how various time-varying NMA methods can impact final estimates of treatment cost-effectiveness[2]
- We discuss the complexities of applying these emerging, time-varying ITC methods within the context of informing EU Joint Clinical Assessments and NICE technology appraisals

Key: CEM, cost-effectiveness modelling; HR, hazard ratio; HTA, health technology assessment; ITC, indirect treatment comparison; NMA, network meta-analyses; PH, proportional hazards; SoC, standard of care;

References: 1. Rutherford, MJ. et al. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020 [Available from http://www.nicedsu.org.uk]. 2. Yu Heng Liu. Implementation of non-proportional hazards network meta-analysis results in cost-effectiveness models. 2023. Data on file.

Poll Question

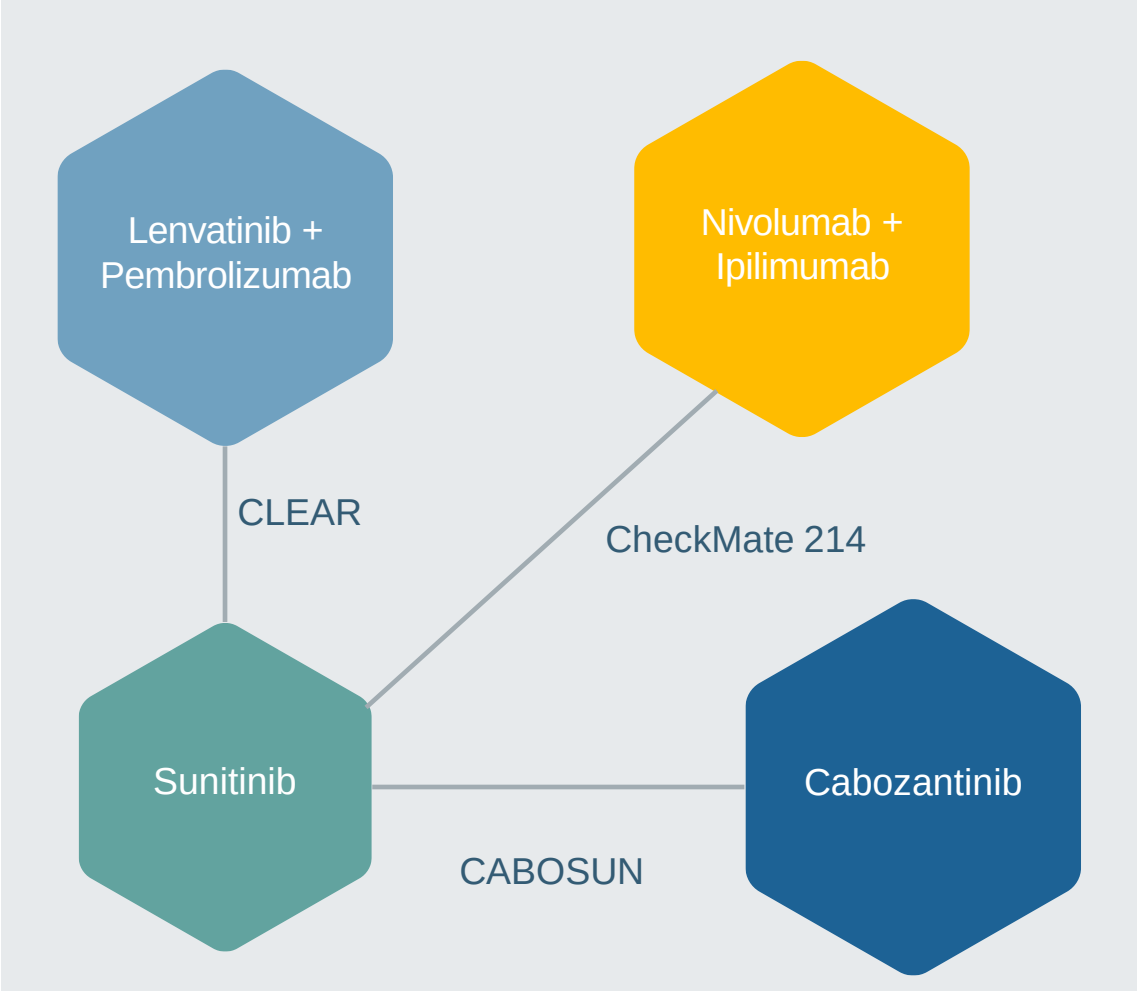
Poll Question: *Have you ever used time-varying ITC?*

A. Never heard of it!
B. I’m aware…haven't really used it
C. I’m a frequent flyer!

Decision problem exploration – 2022 NICE MTA [ID3760], TA858[1,2]

Presentation scope	
Population	Untreated advanced RCC: Intermediate/poor risk subgroup (IMDC criteria)
Interventions vs sunitinib (common comparator to connect the network)	▪ **Lenvatinib + pembrolizumab** (CLEAR)[3] ▪ **Cabozantinib** (CABOSUN)[4] ▪ **Nivolumab + ipilimumab** (CheckMate 214)[5]
Outcome	Progression-free survival
Cost-effectiveness analysis	▪ Incremental cost per QALY ▪ NHS and PSS perspective ▪ Three-state PSM

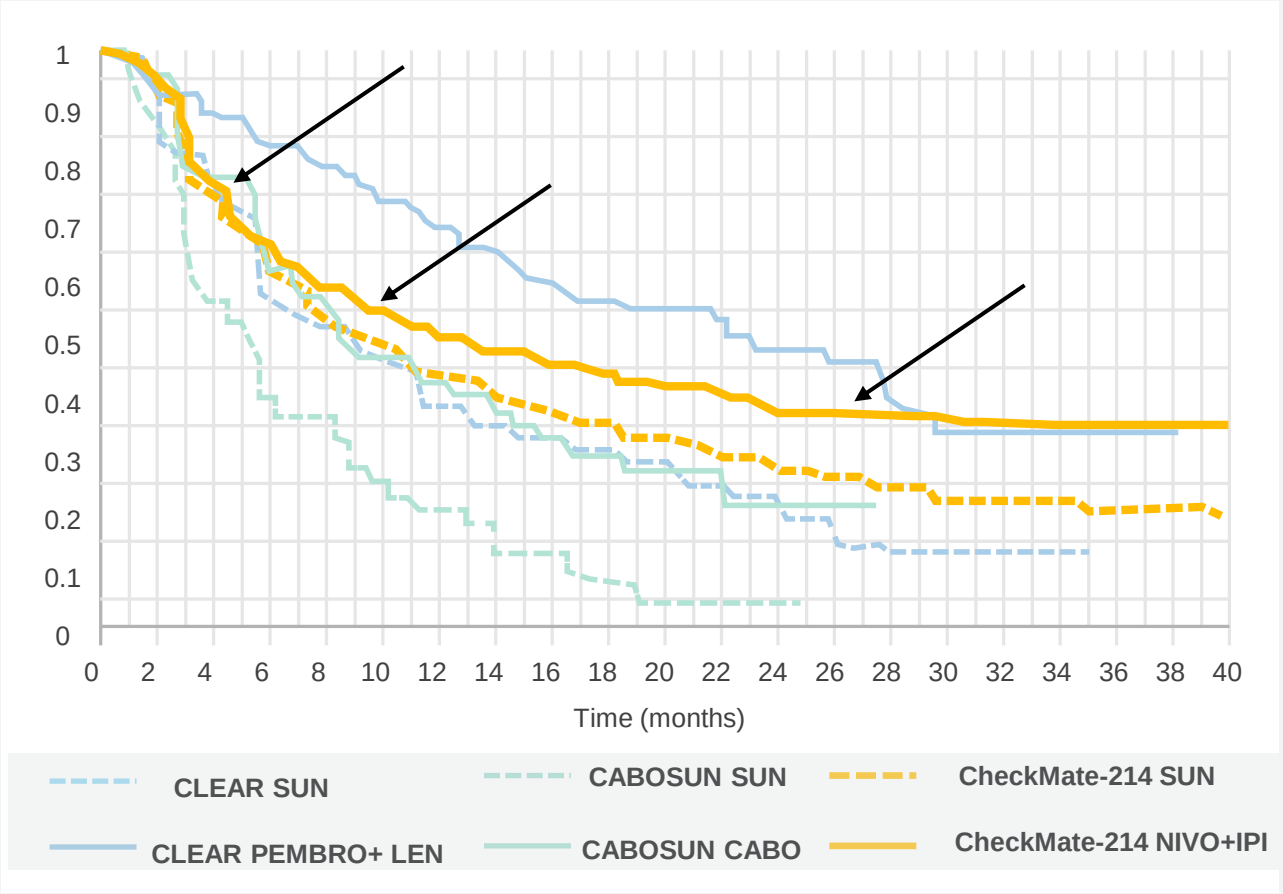
ID3760: Network diagram for the intermediate/poor risk subgroup (PFS, OS and ORR)



Key: IMDC, International Metastatic Renal Cell Carcinoma; MTA, multiple technology assessment; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; NMA, network meta-analyses; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; PSS, Personal Social Services; PSM, partitioned survival model; QALY, quality-adjusted life year; RCC, renal cell carcinoma. **References**: 1. Fleeman N et al Lenvatinib plus pembrolizumab for untreated advanced renal cell carcinoma [ID3760]: A Multiple Technology Appraisal. LRiG, University of Liverpool, 2018. Data on file. 2. NICE MTA Overview | Lenvatinib with pembrolizumab for untreated advanced renal cell carcinoma | Guidance | NICE. 2023. 8 November 2024. 3. Motzer R et al. *NEJM*. 2021 Apr 8;384(14):1289-300. 4. Choueiri, T.K. et al. *Eur J Cancer*, 94, pp.115-125. 5. Motzer, R.J. et al. *J. Immunother. Cancer*, 8(2).

AG discussion of proportional hazards violation

- AG discussed a proportional hazards violation for progression-free survival in the combined intermediate- and poor-risk subgroup of CheckMate-214
- Time-constant HR is not an appropriate estimator of all time points across the trial follow-up periods. (This would apply both the trial-reported HR and for HRs estimated from the NMA.)
- The companies provided results from fractional polynomial NMAs, but the AG considered these '**highly uncertain**'
- AG cautioned that the estimates from these flexible modelling techniques can be '**unintuitive and difficult to interpret**'
- AG discussed uncertainty for both PH and fractional polynomial NMA; they considered the results of proportional hazards NMAs to be '**less uncertain'** and '**could be used for decision making**'



Key: AG, Assessment Group; HR, hazard ratio; NMA, network meta-analyses; PH, proportional hazards

Applying NICE MTA [ID3760] Evidence

Our enhanced methods: Conducting multiple time-varying NMA results to a reconstructed CEM

Actions

- Bayesian NMA of fixed effects models to estimate PFS:[2–4]
  1) Multivariate parametric survival models[5]
  2) Fractional polynomial (FP) models[6]
  3) Restricted cubic spline NMA[7]
  4) (Time-constant) Cox PH estimates[8]
- Reconstructed patient-level data from published KM graphs[1]
- For each model there are unique processes and decisions; design choices aimed to be consistent across model
- NMA model selection was based on recommended criteria for survival extrapolation[9], including visual fit to the published KM graphs, statistical parsimony, biological plausibility, and NMA model diagnostics (e.g. model convergence)
- Sensitivity analyses to model choice was assessed (results not presented)

Supportive assumptions

- Published trial results and available MTA inputs were prioritized, with pragmatic assumptions to address evidence gaps
- For CLEAR, digitized PFS KM were sourced from the overall population as evidence for the intermediate/poor risk subgroup was not available (PFS HR difference of 0.39 vs. 0.42)
- Time to treatment discontinuation was assumed to be equal to PFS
- The type of NMA method involves different selection of the reference treatment curve (trial-specific vs pooled across network)
- Some FP models resulted in implausibly large HRs at the start of follow-up (leading to zero survival); For the first month, no treatment effect applied from the FP models

Key: FP, fractional polynomial; HR, hazard ratio; KM, Kaplan–Meier; MTA, multiple technology assessment; NICE, National Institute for Health and Care Excellence; NMA, network meta-analyses; PFS, progression-free survival

References:: 1. Liu N, et al. *BMC Med Red Meth.* 2021 Jun 1;21(1):111. 2. Cope S et al. *Value Health*. 2023 Apr 1;26(4):465-76. 3. Dias, S, et al NICE DSU Technical Support Document 2: 2011;. http://www.nicedsu.org.uk. 4. Sturtz S, et al . *J Stat Software*. 2005 Jan 7;12:1-6. 5. Cope S, et al. *Research Syn Methods*. 2020 May;11(3):443-56. 6. Jansen JP. *BMC Med Res Methodol*. 2011 May 6;11:61. 7. Freeman SC, et al *Res Synth Meth* 2017;8(4):451–64. 8. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological). 1972 Jan;34(2):187-202. 9. Latimer, N. NICE DSU Technical Support Document 14: 2011. http://www.nicedsu.org.uk.

Resultant ICERs by NMA model type (numerically rounded)

Changing the PFS* NMA inputs

- **Across NMA models**

ICER: Pembro + Len vs. CABO, range (max-min) ≈160K per QALY

ICER : Pembro + Len vs. NIVO + IPI, range (max-min) ≈900K per QALY

NMA model	ICER vs CABO	ICER vs NIVO + IPI
Proportional hazards (time-constant)	£563K	£1.4 mil
Restricted cubic spline (two-knot)	£577K	£1.0 mil (minimum)
1st-order fractional polynomial	£512K	£1.7 mil
2nd-order fractional polynomial	£421K (minimum)	£1.8 mil
Generalized gamma (parametric)	£525K	£1.2 mil
Log-normal (parametric)	£582K (maximum)	£1.9 mil (maximum)

Notes: OS NMA results did not change across PFS modeling and were assumed proportional. In the next iteration of this CEM exercise, we will also model OS with different time-varying hazards models. We expect larger impact on the CEM life-years and quality-adjusted life-years which may stabilize final ICER results.

Key: ICER, incremental cost-effectiveness ratio; NMA, network meta-analyses; PFS, progression-free survival; QALY, quality-adjusted life-year.

Technical/project next steps

Continuing the exercise

01

Written dissemination of the statistical methods, including the alternate choices and their impact in the CEM

02

As life-year estimates also rely on OS, we will add time-varying OS NMA results to this exercise in PFS. This will help us show additional meaningful changes of the incremental LYs and QALYs.

03

Consideration of additional time-varying ITC methods, such as mixture cure models to address updated, longer term trial results

04

Potentially compare with ITC techniques that adjust for population differences, such as ML-NMR of time-to-event data.

Key: CEM, cost-effectiveness modelling; ITC, indirect treatment comparison; LY, life-years; ML-NMR, multilevel network meta-regression; NMA, network meta-analyses; OS, overall survival; PFS, progression-free survival.

Changing With the Times: Hazards in Novel Comparative Efficacy Techniques While Maintaining the Path of HTA Informativeness

Panellist:

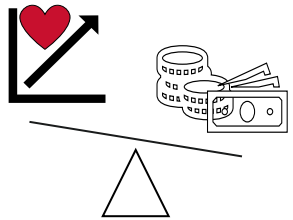
Miranda Cooper, GMA & Payer, Oncology Biometrics, AstraZeneca, Macclesfield, UK

Conflicts of Interest

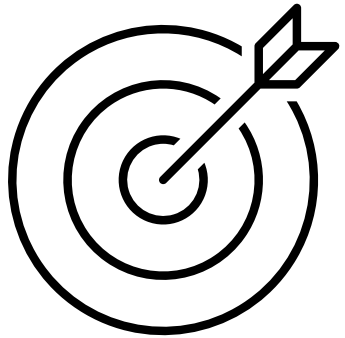
I am an employee of AstraZeneca and the views expressed here are my own and not those of AstraZeneca

Clinical benefit

Clinical and Cost-effectiveness



Value-based



Objective: Reduce uncertainty for decision maker

Diversity of HTA

5 (20%)	13 (52%)	7 (28%)	HTA model by country* Fontrier (2022)

*Includes count for assessment at EU level

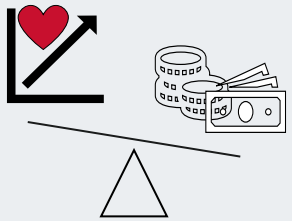
Fontrier, A. M., Visintin, E., & Kanavos, P. (2022). Similarities and differences in health technology assessment systems and implications for coverage decisions: evidence from 32 countries. PharmacoEconomics-open, 1-14.

Our Approach to ITC has large downstream impact

Comparative effectiveness approach

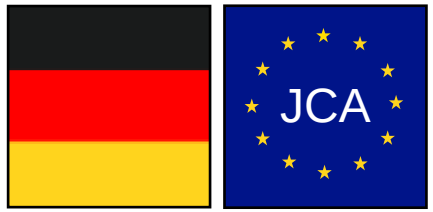
Clinical benefit

Clinical and cost-effectiveness



Value-based

Examples:



Number of distinct ITC types submitted in England is **at least twice** that submitted in France, Germany Italy and Spain (Macabeo 2024)

Acceptance of current ITC methods found to range from **0% (France)** to **47% (England)** (Macabeo 2024)

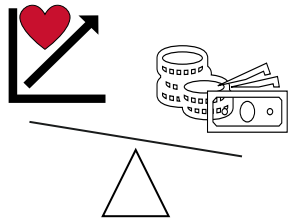
Diversity in the requirements of the decision maker means that a harmonised approach to ITCs is challenging.

Comparative efficacy is the basis of HTA approval across EU and non-EU countries regardless of HTA model

Macabeo, B., et al. (2024). The Acceptance of Indirect Treatment Comparison Methods in Oncology by Health Technology Assessment Agencies in England, France, Germany, Italy, and Spain. PharmacoEconomics-Open, 8(1), 5-18.

Clinical benefit

Clinical and
Cost-effectiveness



Value-based

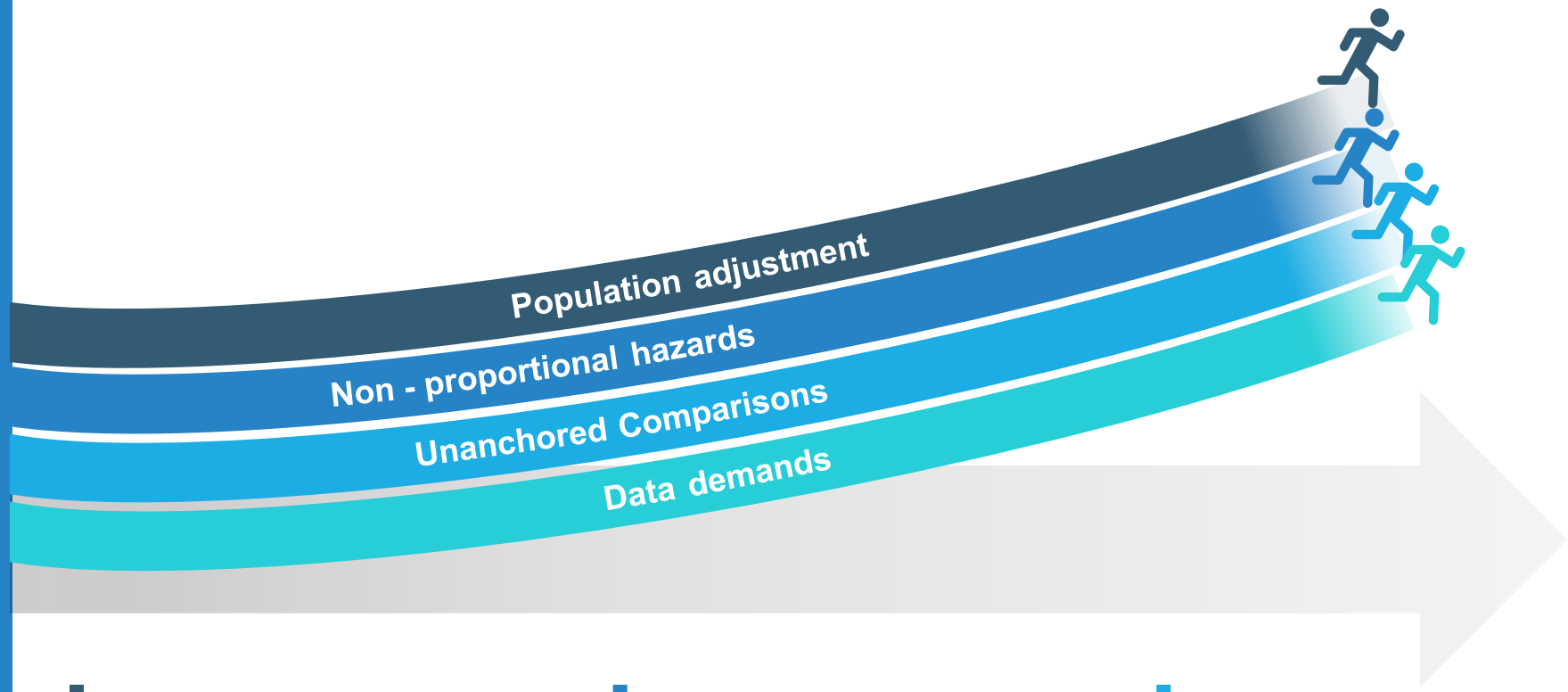
Diversity of HTA

Objective: Reduce uncertainty for decision maker

Plan accordingly

Consider how NPH fit into the bigger evidence picture

Fitting NPH into the bigger (evidence) picture



Treatments continue to advance and there is increased demand for comparative efficacy which have brought new challenges

Methods for comparative efficacy have advanced at a similar pace (Ades 2024)

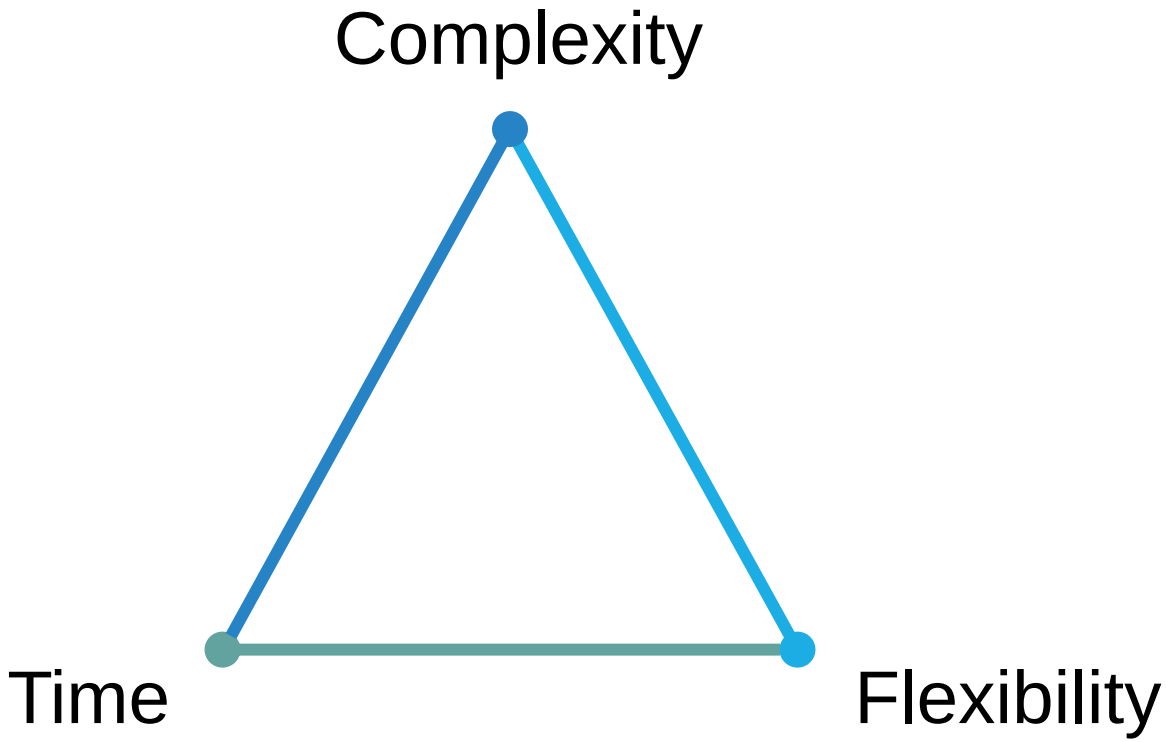
Generally, these advancements have been developed to address analysis challenges in isolation and are associated with individual complexities and assumptions

Addressing them together compounds complexity and scope

Even if we could, should we?

Ades, A. E. et al. (2024). Twenty years of network meta-analysis: Continuing controversies and recent developments. Research Synthesis Methods

Finding the right balance



How do we plan for NPH?

Increasing demand for pre-specification but recent guidance offers flexibility in approach

There is no clear guidance on the best method for NPH analysis, or under which conditions should one method be preferred over another. This complicates pre-specification.

An everything approach to analysing NPH TTE outcomes not viable for HTD **or** for assessors/HTA to appraise – **information overload**.

Need for HTD and assessors to avoid familiarity bias, there is more than one way to meet analysis requirements.

Prioritise the key challenges and select one or two methods that can be justified to address these

Erasmus School of
Health Policy
& Management

Non-proportional hazards

How do different European reimbursement agencies take time-varying hazard ratios into account

Maiwenn Al

Case: tafasitamab

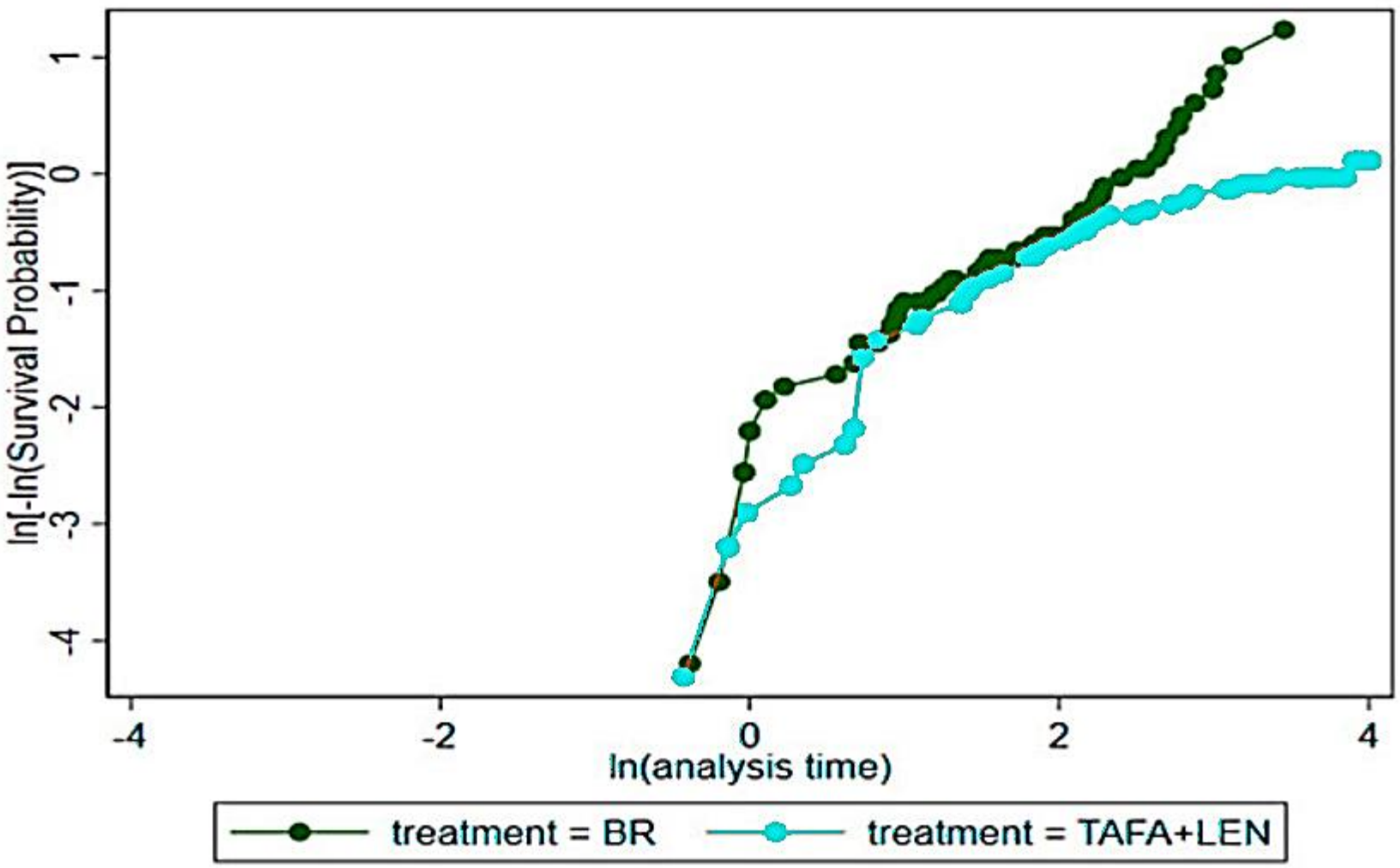
Intervention:	tafasitamab + lenolidamide (TAFA+LEN)
Population:	relapsed/refractory large B-cell lymphoma (not eligible for ASCT)
Evidence:	**single arm**, phase 2 (L-MIND)
Outcomes:	overall response rate, PFS, OS

Comparative effectiveness – ITC

Comparators:	- R-GemOx (rituximab – gemcitabine – oxaliplatine) - BR (bendamustine – rituximab)
Observational cohort:	- Data from health records (N. America, Europe, Asia-Pacific) - In- and exclusion similar to L-MIND - Propensity score 1:1 matching - PH not assessed
Results median OS:	- TAFA+LEN 31.6, BR 9.9, R-GemOx 11.0 months
HR vs TAFA:	- R-GemOx HR 0.42, BR HR 0.42

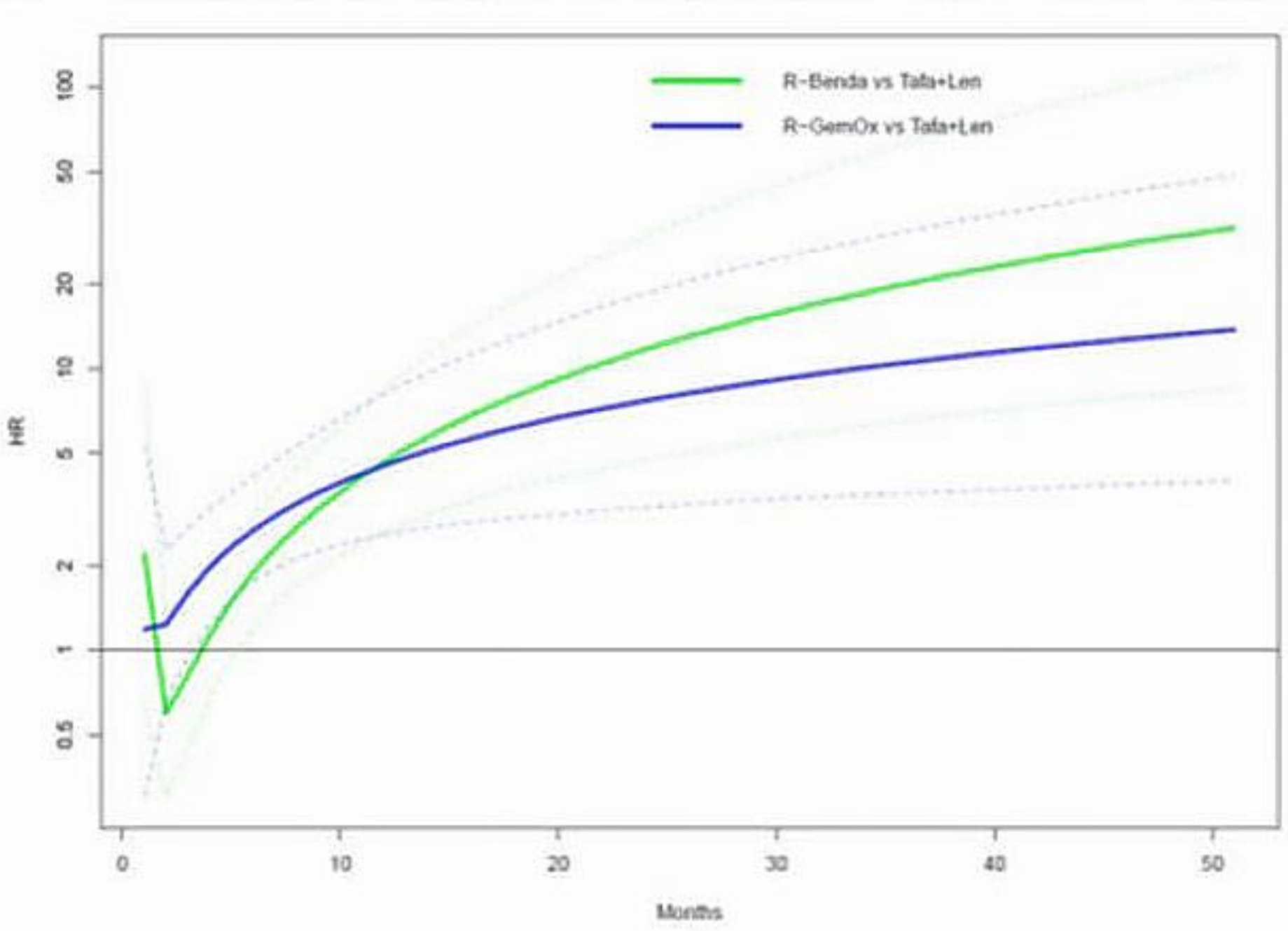
Clin Cancer Res 2022;28:4003–17 10.1158/1078-0432.CCR-21-3648

PH diagnostic
TAFA vs BR
-
PFS



https://www.has-sante.fr/upload/docs/application/pdf/2022-05/minjuvi_12042022_avis_economique.pdf

HR over time



https://www.has-sante.fr/upload/docs/application/pdf/2022-05/minjuvi_12042022_avis_economique.pdf

Comparative effectiveness – MAIC

Comparators: - Pola-BR (polatuzumab – bendamustine – rituximab)

Approach:
- Phase 2 study comparing Pola+BR (n=40) vs BR
- IPD data TAFA adjusted to match baseline Pola-BR
- After matching effective sample size TAFA 29 (was 80)

Results:
- PH appeared not to hold
- OS < 4m HR 1.82, > 4m HR 0.41

Adv Ther (2022) 39:2668–2687 10.1007/s12325-022-02094-5

NICE (UK)

Clinical effectiveness **–** vs Pola-BR

- Indirect comparisons suggest TAFA improved PFS and OS, but not always stat. significant (PH not mentioned)
- Results highly uncertain due to complexity and potential biases of methods indirect comparisons

Cost effectiveness

- OS and PFS extrapolations Pola highly uncertain, in part due to time-varying hazard ratio. Estimated OS and PFS not aligned with NICE STA Pola – BR
- Explicit exploration (non-) PH by EAG
- OS and PFS extrapolations TAFA appropriate, but uncertain

Decision Not recommended

ZIN (Netherlands)

Clinical effectiveness – vs Pola-BR and BR (proxy for R-PECC)

- vs BR: Effect of TAFA very large, but also very uncertain, high risk of bias, unclear clinically relevant
- Vs Pola: Effect of TAFA stat. significant **> 4 months**, very uncertain, high risk of bias, not possible to conclude superior, but not inferior
- PH only discussed as delivered by company (split at 4 months)

Cost effectiveness - Not required

Decision Only insured care if price TAFA equivalent to Pola -> price remained higher, so not reimbursed

HAS (France)

***Clinical effectiveness* –** vs R-GemOx

- Documentation about indirect comparison and estimation of HRs insufficient
- Important covariates not included in ITC (e.g. ECOG)

Cost effectiveness

- Due to major methodological issues (clinical effectiveness and quality of life) results CEA invalid
- Explicit exploration (non-) PH

Decision

- TAFA not proven to provide improvement compared to current treatment => ASMR V (only listed if lower price or cost-saving)
- TAFA given early access due to innovation

TLV (Sweden)

***Clinical effectiveness* –** vs R-GemOx and Pola-BR

- Indirect comparisons suggest TAFA improved PFS and OS, but exact size not clear due to indirect comparisons leading to high uncertainty and high risk of bias

Cost effectiveness

- Non-PH not explored, accepted as delivered by company (MAIC: split at 4 months)
- Results uncertain due to issues clinical effectiveness and company's assumptions cure/long term survival

Decision not nationally reimbursed

Summary

Clinical effectiveness

- Highly uncertain due to indirect comparison
- Data suggests TAFA superior
- PH not discussed or even mentioned

Cost effectiveness

- Uncertainty clinical effectiveness leads to high uncertainty about cost-effectiveness
- UK and France explicit exploration (non-) PH, Sweden not

All countries decided not to reimburse TAFA (unless lower price)

Some observations and thoughts

- Indirect treatment comparisons not problematic (methodologically)
- High uncertainty and risk of bias from ITC is problematic

- Clear divide between assessment clinical effectiveness and cost effectiveness
- CEA common to explore PH => independent or joint survival curves

- Will this methodological divide remain in Joint Clinical Assessment?

- If so, will we achieve goal EU HTA regulation to avoid divergent data requests (and administrative burden) to health technology developers?

General Q&A

Acknowledgements

Lumanity Colleagues

Ioana Gulas: Senior Statistician; Toronto, CA

Yu Heng Liu: Health Economist; Sheffield, UK

Ben Kearns: Director, Health Economics; Sheffield, UK

Emma Tyas: Senior Statistician; Sheffield, UK