Incorporating External Evidence into Extrapolations: A Debate on the Trade-Offs Between Complexity and Simplicity

Motivation

- Health technology assessment (HTA) often aims to estimate long-term outcomes such as life years (LYs) and quality-adjusted life years (QALYs) with and without evaluated treatments using extrapolation methods
- Why the use of external evidence for extrapolation?
  - Short-term time horizons from phase II-III studies with immature overall survival (OS) data, (e.g., relatively mature progression-free survival [PFS] data but limited OS data available)
  - Single arm regulatory approval studies requiring a clinical practice or historical comparator
  - Baseline risk may not represent populations in practice settings

Current use and recommendations for external evidence in extrapolations

- National Institute for Health and Care Excellence (NICE) recommends using real-world evidence (RWE) for baseline risk estimates but keeping trial-based treatment effects[1]
- NICE and Canada's drug and health technology agency considers expert elicitation methods in absence of data[2]
- No formal recommendations from Institute for Clinical and Economic Review (ICER) but acknowledges the benefits and limitations of RWE and identifies sources during the literature review stage[3]
- Multiple solutions available for incorporating RWE but considerable barriers in place

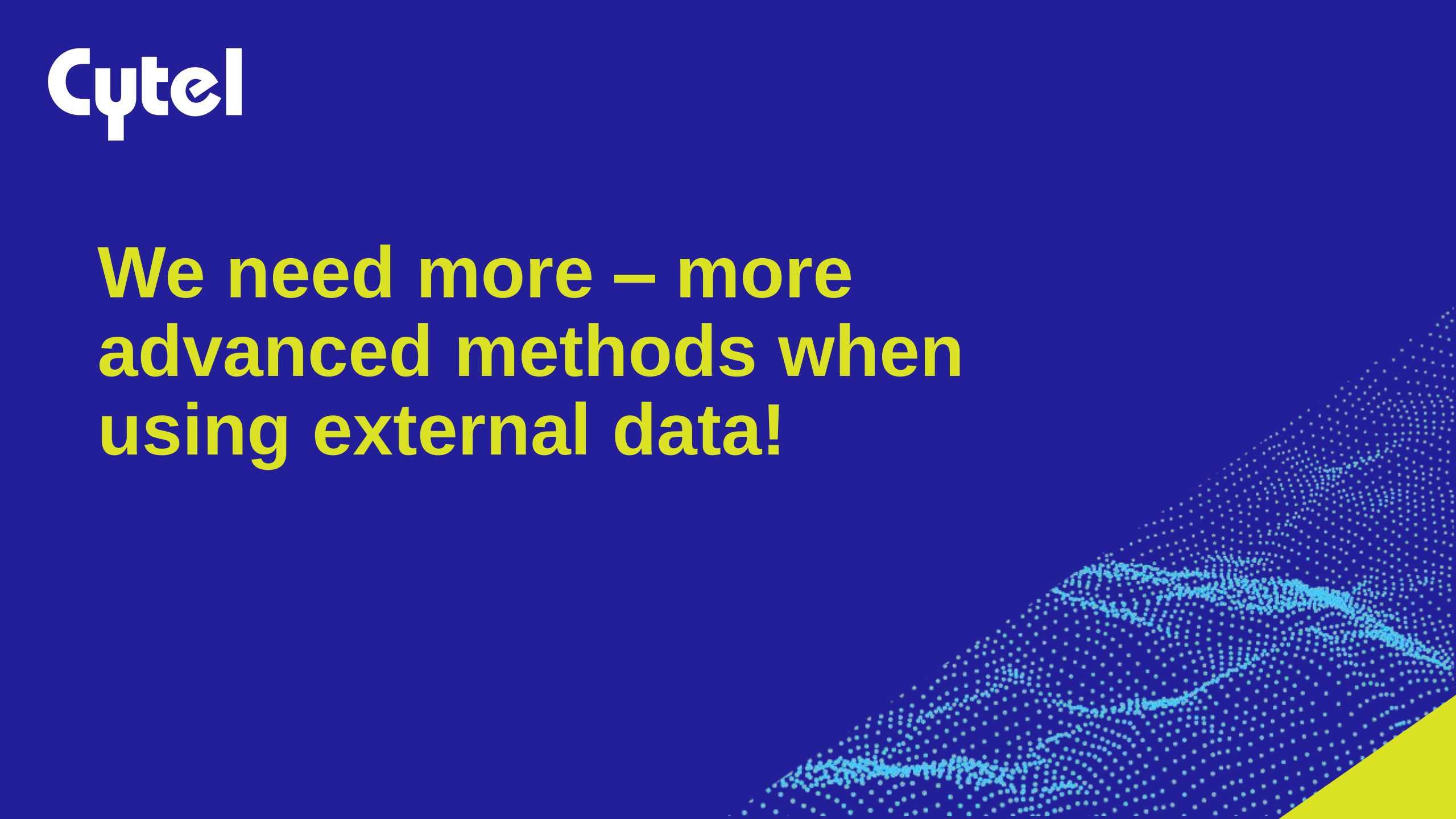
1 https://www.nice.org.uk/corporate/ecd9/chapter/overview;
2 https://www.york.ac.uk/che/research/teehta/elicitation/steer/
3 https://icer.org/our-approach/methods-process/considering-clinical-real-world-and-unpublished-evidence/

Speakers

- R. Brett McQueen (PhD) Assistant Professor, University of Colorado Anschutz Medical Campus
- Bart Heeg (PhD) Vice President HEOR Cytel
  - Methods developer perspective
- Dawn Lee, Associate Professor of Health Economics and Health Policy at PenTAG, University of Exeter as leader on this submission
  - Model reviewer for HTA perspective
- Ash Bullement (MSc) Analyst at Delta Hat and PhD student at ScHARR, University of Sheffield
  - Advisor and consultant preparing HTA submissions perspective

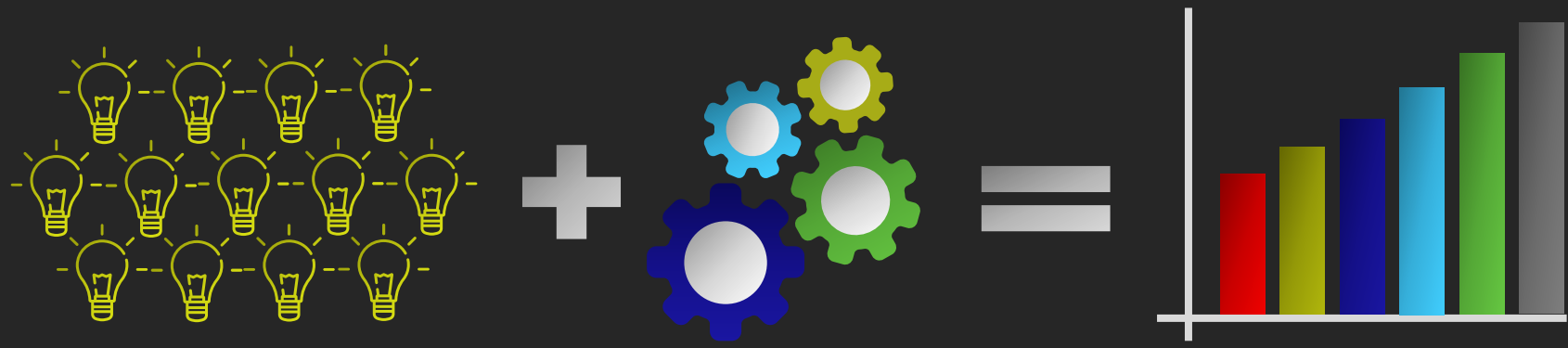
Cytel

We need more – more advanced methods when using external data!



The simple side of the equation

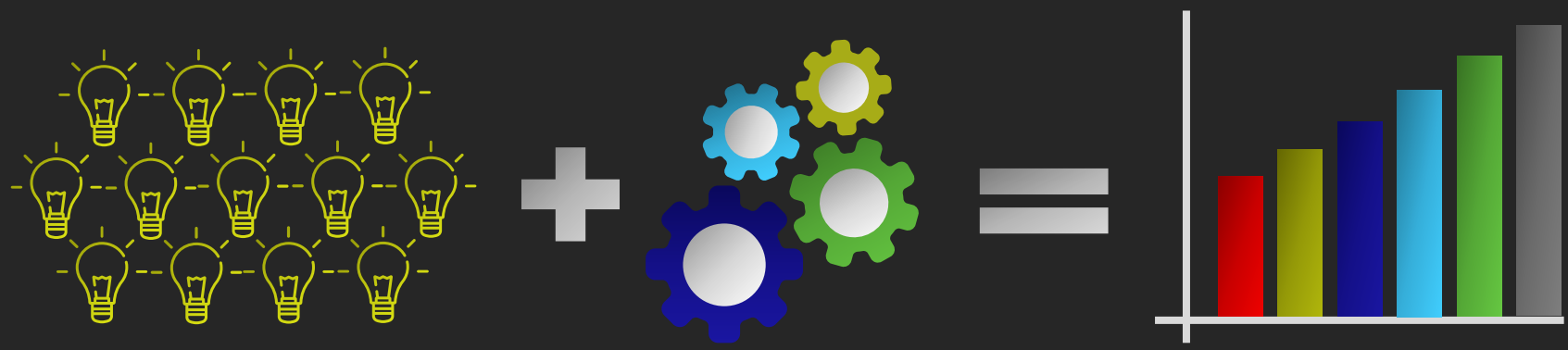
- The use of external data (e.g., real-world evidence) to inform survival extrapolations has been around for decades.
- External data can be used to inform any type of model structure
- Some guidance and methods are available (e.g., TSD 21, Bullement, et al., 2023).
- The use of external data has been:
  - Shown (given the right data) to improve clinical plausibility of survival extrapolations (Chaudhary et al. 2023; Soikkeli et al. 2019)
  - Shown to reduce structural uncertainty (Soikkeli et al. 2019; van Oostrum et al. 2021)
  - Accepted by health technology assessment (HTA) agencies, for instance (but not limited to):
    - Informed fits were approved among others for apalutamide in non-metastatic castrate-resistant prostate cancer in Sweden, Ireland and England.
    - In the Netherlands, cure rates on progression-free survival (PFS) were considered informative for cure rates in overall survival (OS).



Cytel

The more complex side of the equation

- Methods considering external data for survival extrapolations do not need to be complex.
- What is complex though …
  - The number of (scenario) analyses
  - Finding external data that matches the extrapolation population and purposes
  - Novel therapy effect (first in class) in extrapolation period
  - Weighing the importance of external data for the survival extrapolations
  - Generalizability of real-world evidence



There is NICE guidance on external data

TSD 21 Summary recommendations related to external data

- Discuss the potential to couple complex modelling with the range of trial data with external information/data to make more plausible extrapolations
- Incorporation of background mortality should be strongly considered to avoid very poor extrapolations
- Consider other external information (e.g., registry data) to help model long-term survival
- Extrapolating relative treatment effects could involve borrowing information from similar drug classes and other longer-term clinical trial follow-up, and/or eliciting expert opinion. Further research and evaluation is required to explore the viability of this approach.

NICE Guidelines Technical Support Unit. Meta-Analysis. Guideline Methodology Document 1 (Daly et al., 2021)

- Discusses how to model class effects in meta-analysis
- Refer to examples in chronic obstructive pulmonary disease, migraine, pressure ulceration, and for over-active bladder

NICE real world evidence framework

- RCTs are the preferred study design for estimating comparative effects
- Real world evidence should ideally be used to model/extrapolate standard of care

RWE is valuable indeed, but keeping it simple by ignoring complexities might be infeasible or produce clinically implausible survival extrapolations

Generalizability (of real-world evidence (NICE framework))

- Mature RWE is more informative for extrapolations, but means it is likely old, potentially raising generalizability issues
  - Need for complex left / right censoring techniques to create a RWE dataset more generalizable of current practice
- Sample size differs between RWE and trial and this impacts modelled uncertainty
  - How to weigh, or handle uncertainty, if the trial has more patients than the only small UK registry or vice versa
- (Earlier) endpoints not readily available in RWE or measured differently (e.g. PFS)
  - Is the model limited to outcomes generated in RWE or does one use e.g. RWE OS and RCT PFS?
- RWE populations are different (typically broader) than RCT populations
  - How to match and model important effect modifiers and prognostic variables not collected/reported/available from RWE
- If one wants to model effectiveness, one should also consider relative effectiveness and not relative efficacy
  - What about differences in compliance in clinical practice vs clinical trials impacting relative efficacy/effectiveness
  - If local RWE data is used should the relative efficacy be based on European subset of trial patients?
- Combining time-varying HR from the trial with an RWE SoC arm requires some thought

The trial Data

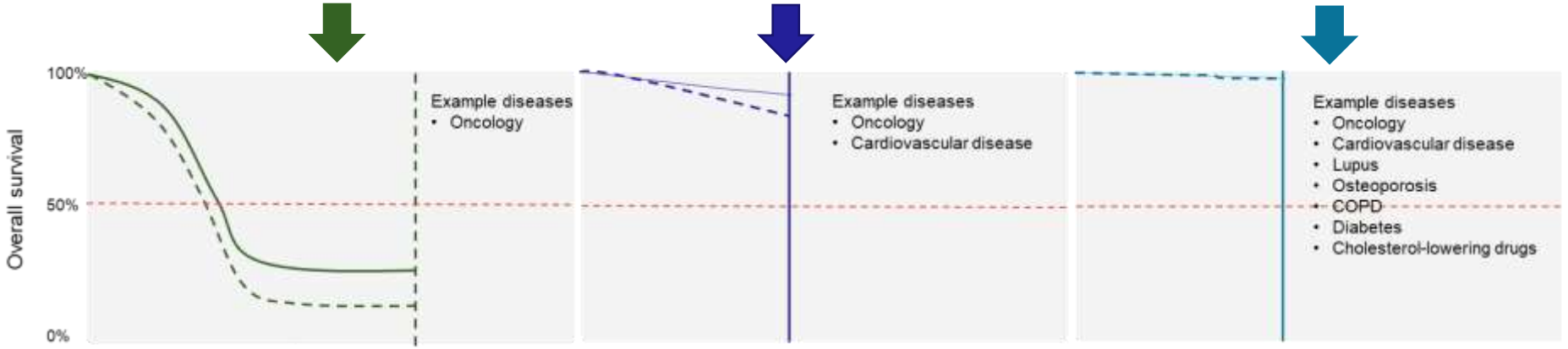
Incremental lifetime survival often is one of the biggest drivers of the incremental cost-effectiveness ratio and therefore important for HTA decision-making.

Median OS data are available at first pricing and reimbursement negotiations

Limited OS data are available at first pricing and reimbursement negotiations, and mature OS data are expected within the therapy lifecycle (before the therapy is off-patent or superseded)

Incomplete or no OS data are available at first pricing and reimbursement decisions or during the therapy lifecycle

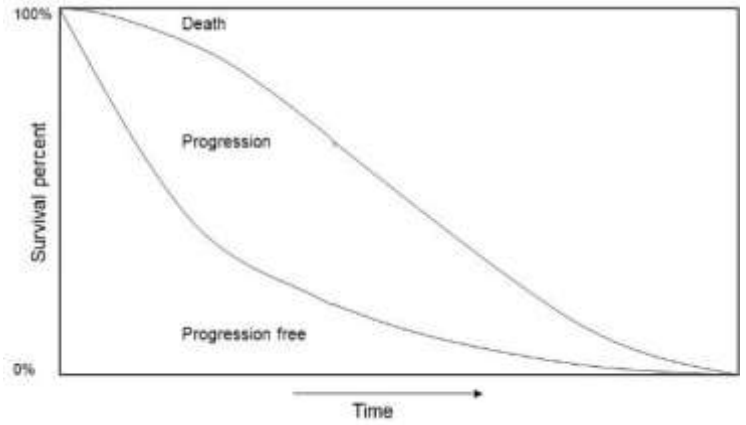
How do we develop and select clinical plausible survival extrapolations?



Source: 1. Lux MP, Ciani O, Dunlop WCN, Ferris A, Friedlander M. The Impasse on Overall Survival in Oncology Reimbursement Decision-Making: How Can We Resolve This? Cancer Manag Res. 2021 Nov 10;13:8457-8471
Abbreviation: COPD, chronic obstructive pulmonary disease

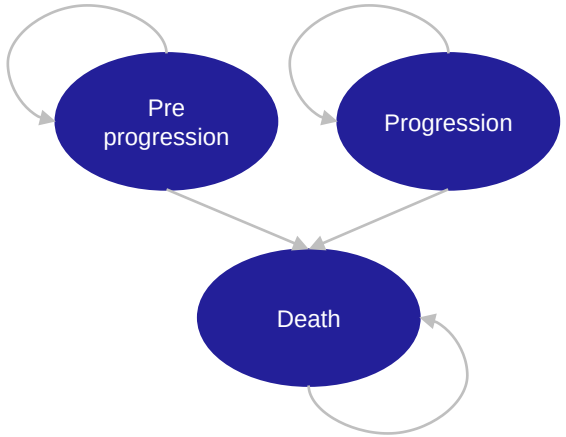
The models typically used for extrapolations

Partition survival model



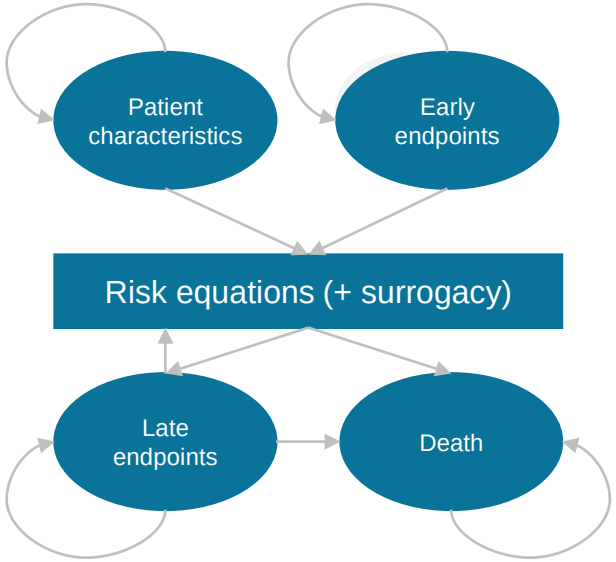
State membership is derived from mutually non-exclusive survival curves. OS is partitioned to estimate the proportion of patients in the progression-free and progression health states (many technology appraisals [TA]).

State transition model



State membership is determined by transition probabilities which determine the probability to stay or leave the health state in each time period (e.g., TA886).

Microsimulation model

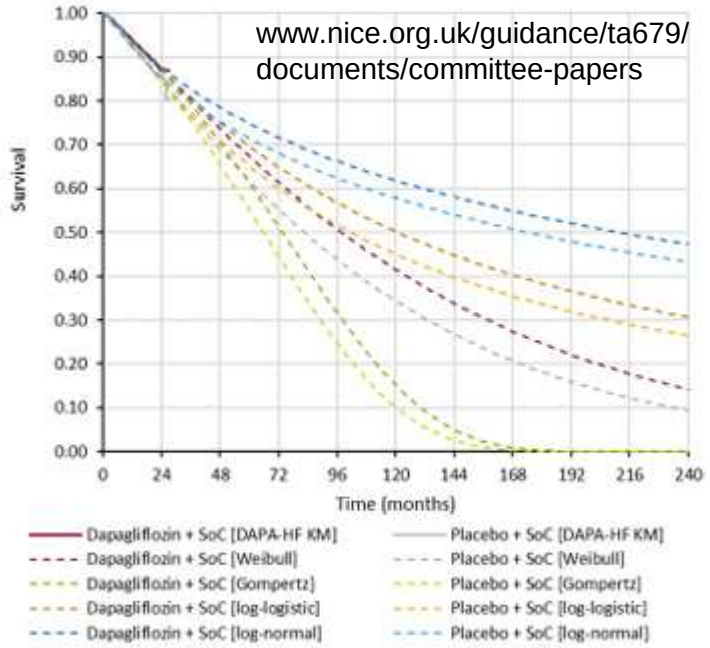


Risk of late outcomes for individual patients is based on risk equations which are informed by patient characteristics and early trial outcomes (e.g., TA418, TA397).

Not including external information in immature survival extrapolations results in structural uncertainty and clinically unrealistic extrapolations

Partition survival model

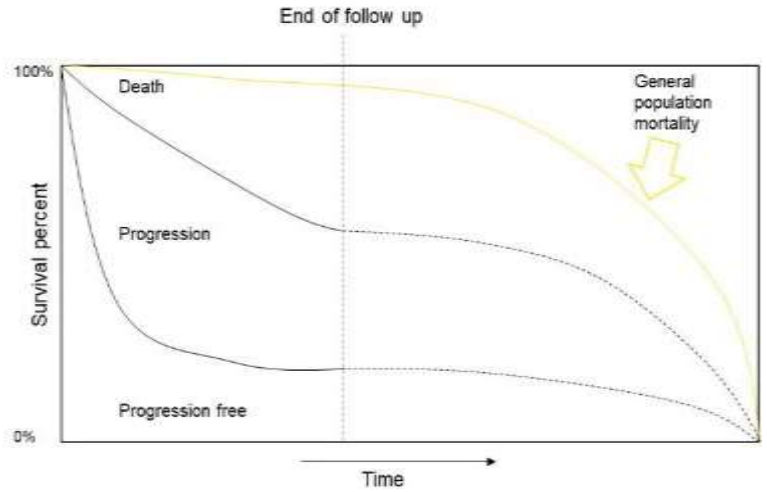
Figure 26: Alternative CV mortality survival curves (scenario analyses)



Abbreviations: CV, cardiovascular; KM, Kaplan Meier; SC, standard care.

Considering general population mortality is recommended, reduces structural uncertainty, increases clinical plausibility of extrapolations.

Partition survival model

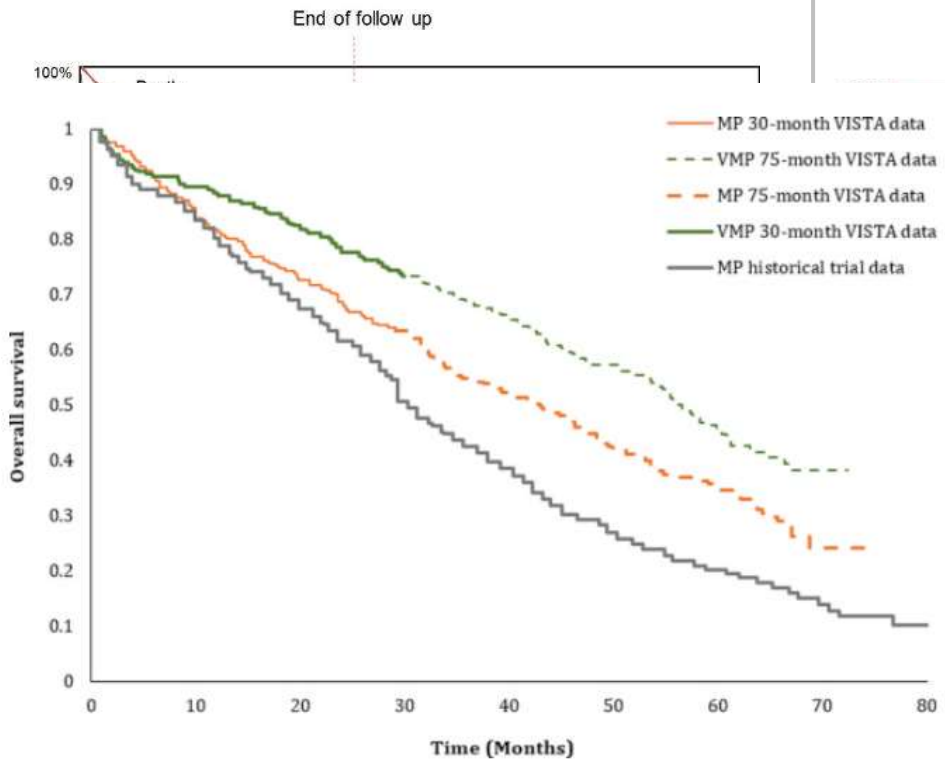


General population mortality (Relative survival)

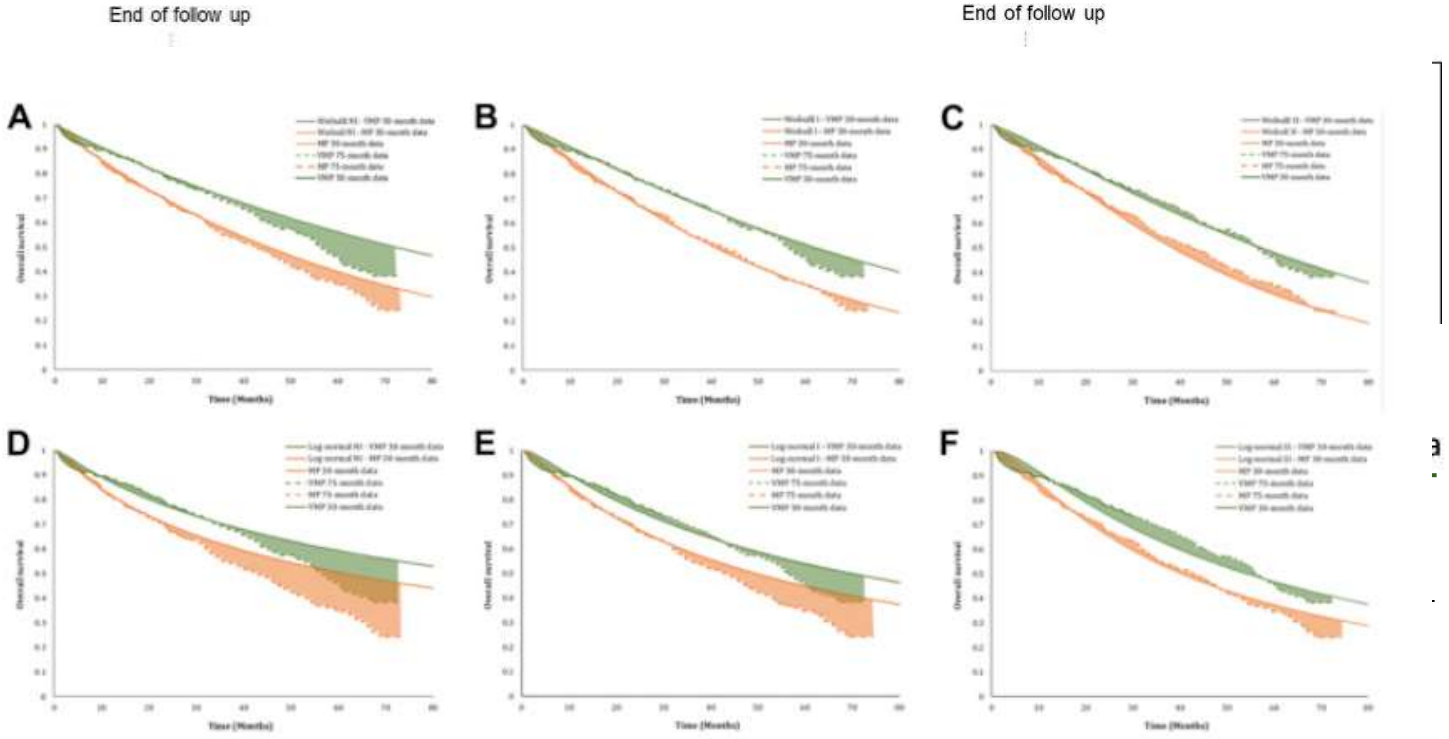
(van Oostrum, Ouwens et al. 2021; Rutherford, Lambert et al. 2020; Lee and McNamara 2023)

The modelling approach should not be based on statistical simplicity, but on biological/clinical plausibility and HTA anticipated acceptability

Partition survival model



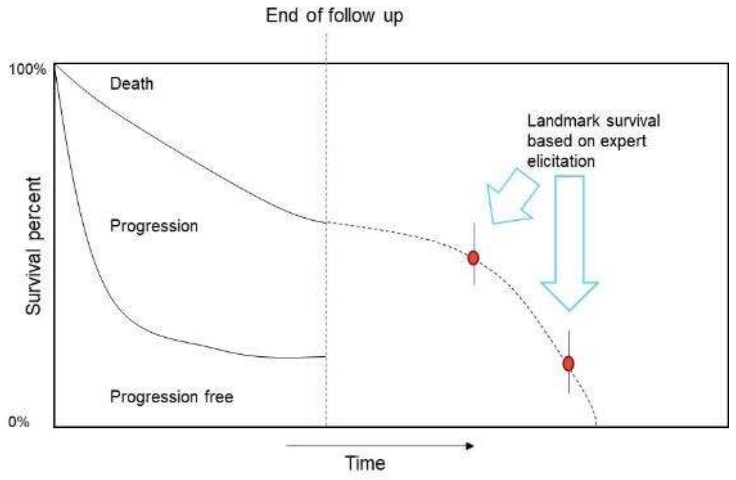
Partition survival model



Partition survival model

The modelling approach should not be based on statistical simplicity, but on biological/clinical plausibility and HTA anticipated acceptability

Partition survival model



Expert elicitation

(Bojke, Soares et al. 2022; Grigore, Peters et al. 2016; Willigers, Ouwens et al. 2023)

Partition survival model

Guyot et al conclude:

Long-term extrapolation using parametric models based on RCT data alone is highly unreliable and these models are unlikely to be consistent with external data. External data can be integrated with RCT data using spline models to enable long-term extrapolation. Conditional survival data could be used for many cancers and general population su
conditions. The use of external data sho
history and treatment mechanisms.

Chaudhary *et al* conclude:

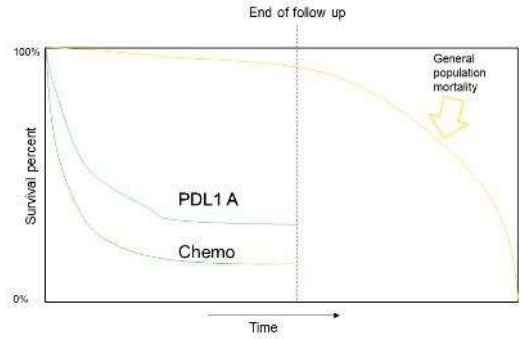
Our work demonstrates the po
advanced parametric models t
sources, such as B-MPES and MCMs, to allow for accurate evaluation of treatment clinical and cost-effectiveness from trial data with limited follow-up.

Methods needed on pooling RWE and SoC data (e.g, Dynamic borrowing)

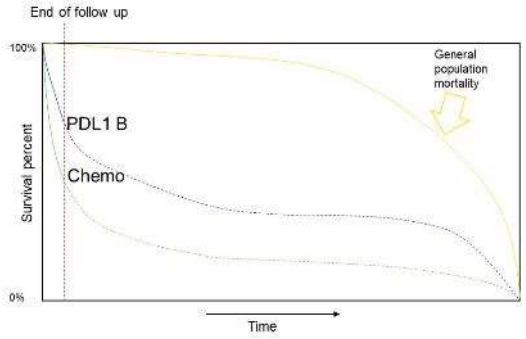
Immature treatment effects may also be informed by external data by assuming a class effect

Partition survival model

Historical trail



Pivotal trail

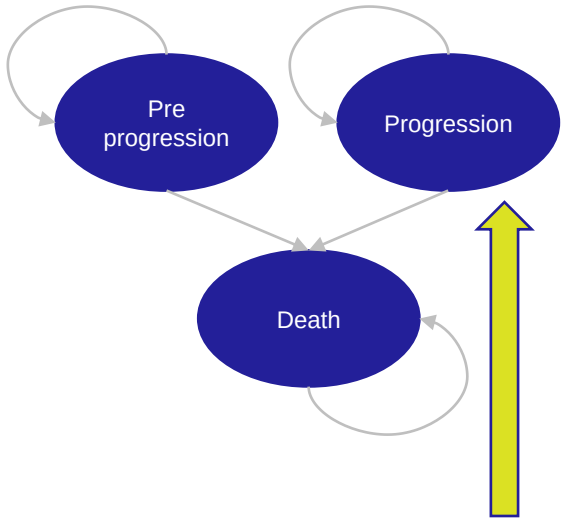


Class effects

Daly et al., 2021, Heeg, Verhoek et al. 2023; Rutherford, Lambert et al. 2020

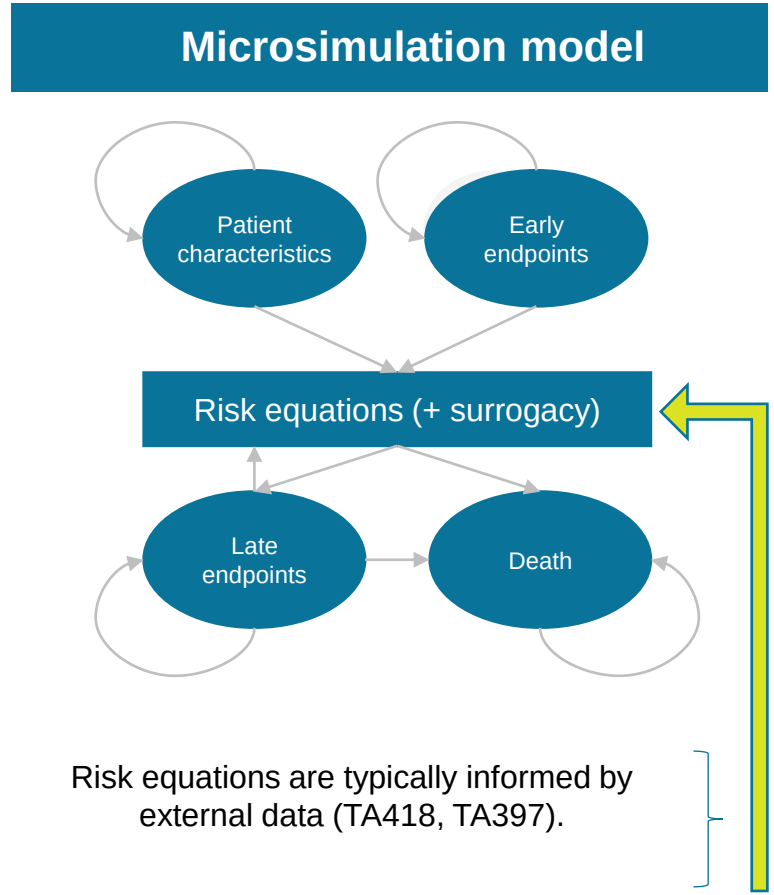
External data has been used and accepted by HTA to inform survival extrapolations in any model structure.

State transition model



The progression state or states can be informed by external data (e.g., TA461, TA761)

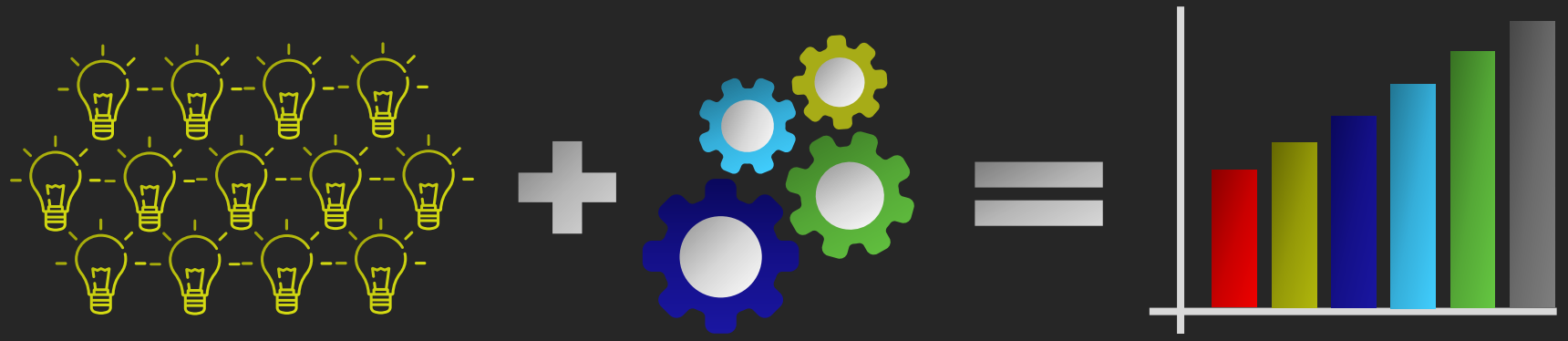
External data has been used and accepted by HTA to inform survival extrapolations in any model structure.



Conclusions

Complexity is relative, but "If you can't explain it simply, you don't understand it well enough"

- We have been relying on methods incorporating external data for decades in any type of model structure
- There is HTA guidance, there are methods, given the right data these methods likely improve clinical plausibility of extrapolations and reduce structural uncertainty
- Combining RWE with RCT raises some complex issues, which ignored may cause clinical implausible survival extrapolations or may not be feasible. We need "complex" methods, e.g.
    - Ensuring mature RWE data is generalizable to current clinical practice. (e.g. Left censoring or tx switch methods)
    - Ensuring we pool RCT or RWE appropriately (e.g. Bayesian dynamic borrowing).
- In the meantime, prepare, tackle these and other "complex" issues early, sometimes with complex methods and prepare a clinical plausible base case, and verify during early HTA consultations
- We need consistent international HTA guidance's to decomplexify the needed methods



Cytel

Cytel

Questions?

University of Exeter

KISS!

Additional complexity may not improve our ability to make good decisions

Dawn Lee (PenTAG)

Common use cases for external data in survival analysis

Do we need complex methods to make the most of external data?

1. Provide a comparator for single arm trials
2. Improve generalisability
3. Make up for immature data

Do we need complex methods to make the most of external data?

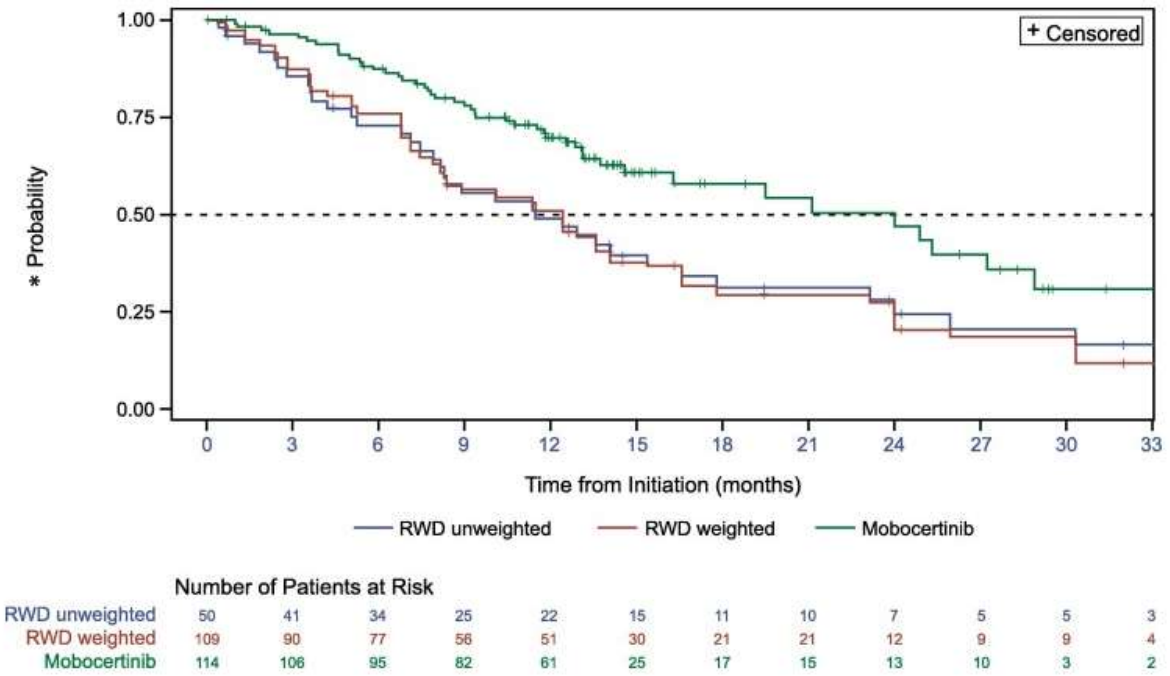
1. **Single arm trial comparator**

- Residual confounding always a concern no matter complexity of company methods (TA894, TA855, TA872, TA559)
- Comparisons of SACT vs trial data as part of CDF re-reviews have systematically shown differences that matching methods could not adjust for (e.g. TA897, TA780)

- Adjustment methods often don't make a massive difference – Committee's therefore often default to naïve comparison with acknowledgement of unquantified amount of potential bias

- Why not reverse the question and ask what level of bias may meaningfully affect results (simple threshold analysis) and how plausible that is?

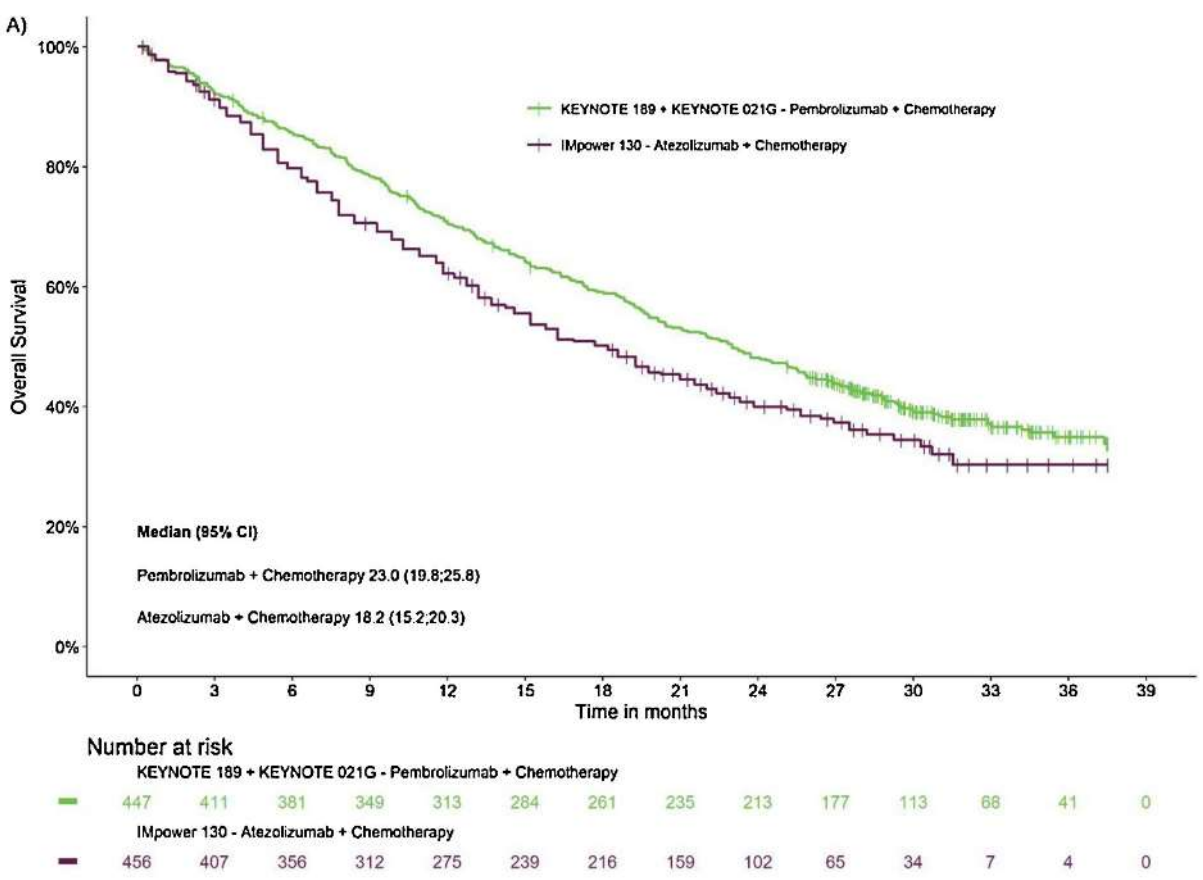
Are we really capturing the differences between datasets with statistical adjustment methods available?

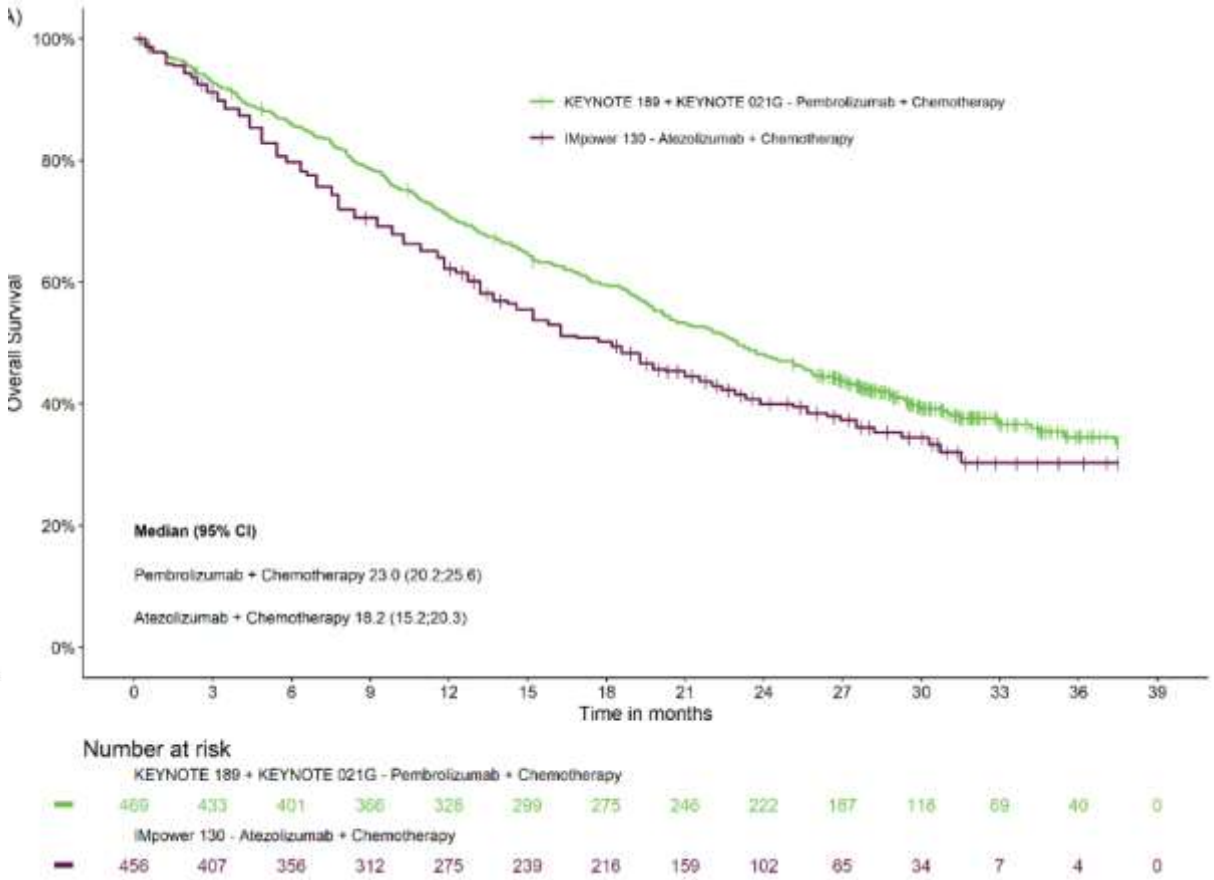


Comparative effectiveness of mobocertinib and standard of care in patients with NSCLC with *EGFR* exon 20 insertion mutations: An indirect comparison. Lung Cancer. Volume 179, 107186, May 2023

	Mobocertinib	**Unweighted RWD**	**Weighted RWD**
No. of patients	114	50	109
Prior treatment %			
EGFR TKI	25.4%	3.0%	Not inc.
IO	43.0%	4.0%	
>= 2 lines	58.8%	4.0%	
Time since initial diagnosis (months)			
Mean (SD)	23.8 (27.92)	17.2 (20.29)	20.9 (34.70)
ECOG			
0 or 1	100%	58%	Not inc.
Missing	0%	42%	
History of smoking, n (%)			
Yes	28.9%	42.0%	30 (27.2)
No	71.1%	58.0%	79 (72.8)

Can you spot the difference?





Pembrolizumab+chemotherapy versus atezolizumab+chemotherapy+/−bevacizumab for the first-line treatment of non-squamous NSCLC: A matching-adjusted indirect comparison. Lung Cancer Volume 155, P175-182, May 2021

Do we need complex methods to make the most of external data?

University of Exeter

2. Improve generalisability?

- Use external data for baseline risk
- Don't necessarily need a complex method here
- Standard extrapolation methods for RWE with relative effects applied via NMA may suffice

"Specifically, the committee thought that **using randomised data to estimate absolute event rates runs the risk of results that do not reflect NHS practice**. It also thought that using observational data to estimate relative effects runs the risk of biased treatment effects because of unadjusted confounding variables. The committee noted that NICE's technical support document 13 makes this distinction, advocating registry data to estimate absolute baseline event rates and randomised evidence to quantify relative differences. **The committee concluded that it still preferred using the real-world evidence to estimate survival for people having cabazitaxel and the network meta-analysis to estimate the relative treatment effect** of cabazitaxel compared with lutetium-177" NICE ID3840

Do we need complex methods to make the most of external data?

3. Immature data

- Complex methods don't make up for lack of data
- Presenting a large number of analyses can result in confusion and delays – particularly when not accompanied with intuitive visualisation
- Where data is immature there is unlikely to be a best curve
- Clinical plausibility is critical:
  - Absolute survival; and
  - Relative effects

"Using data from COU-AA-302 to estimate the effect of a second newer androgen receptor inhibitor and to adjust for survival **causes uncertainty**"

NICE TA740 – apalutamide, this took 3 Committee meetings to resolve

"The EAG considered that the company's choice of survival curves for modelling treatment effectiveness was **not transparent**. It had **concerns with the company's choice of a complex parametric survival curve** to model OS (spline knot 1) instead of a standard parametric model. The company explained that it had selected its curves using clinical expert opinion. Because the data was immature, using a visual and statistical fit of the parametric curves alone was insufficient to select the most appropriate curves. So expert opinion was needed to inform plausible survival at longer time horizons. **The EAG was not convinced by the company's reason and highlighted that in some cases the company's curves were not close to the estimates suggested by its clinical experts**."

NICE TA911

EAGs have very little time to review company submissions

Same for Committees

EAG receives CS

Day 14
EAG submit clarification queries

Day 21
Clarification TC

Day 28
Clarification response

Day 60
EAG report

Day 64
Mark-up and FAC

Day 71
EAG response

~ Day 150
ACM1

Abbreviations: ACM, appraisal committee meeting; CS, company submission; EAG external assessment group; FAC, factual inaccuracy check; TC, telephone conference

Incorporating External Evidence into Extrapolations: A Debate on the Trade-Offs Between Complexity and Simplicity

An industry consultant view

14 November 2023

Ash Bullement (abullement@deltahat.com)

Delta Hat

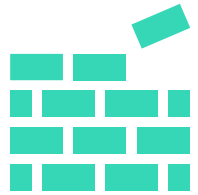
Disclosures

- I am an employee of Delta Hat: a health economic and outcomes research consultancy based in the UK
- The views expressed in this presentation are my own, and do not necessarily reflect those of Delta Hat or any other institution Delta Hat has worked with (past or present)
- Outside of my consultancy role, I am a part-time PhD student at SCHARR (University of Sheffield), and I also collaborate with PenTAG on assessments as part of an External Assessment Group (EAG) – my opinions might not reflect a 'typical' industry view!

Delta Hat

Are we doing this already? Kind of…

- Incorporating external evidence is in keeping with the ethos of HTA
  - HTA is about bringing together **all relevant evidence** to inform decision making - why should this not apply to survival analysis?



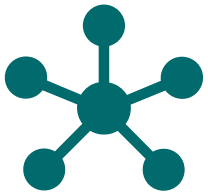
- Arguably we already use external evidence for informing survival estimates, but not usually explicitly as part of the model fitting process
  - Instead, we may comment on how our projections align with published literature, which is 'easier' than formally integrating external evidence into extrapolations

- We also usually account for background mortality within our projections
  - This is most often done simply to address implausible long-term hazards, and is usually a *post-hoc* adjustment

Key: HTA, health technology assessment.

Complexity goes beyond survival extrapolation

- Standardisation of survival extrapolation has improved over time
  - Guidance development (e.g., the TSD series), has played an important role in setting the standard and a benchmark for further development



- Generating robust survival extrapolations is just one part of the workstream required for developing an HTA submission
  - How are we identifying external evidence? Systematic review? Google?
  - What if we have multiple sources? Do we pick ‘the best one’? Should we combine them?
  - How does our survival extrapolation method tie in with our ITC approach?

- HTA timelines are often tight but for good reason – to avoid, wherever possible, delays to patient access
  - HTAs often run in parallel for a range of products, across indications, for multiple countries, and so there is high demand for biostatistics support
  - Planning is key, but hindsight is 20/20!

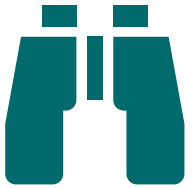
Key: HTA, health technology assessment; ITC, indirect treatment comparison; TSD, technical support document.

Delta Hat

Practical considerations

- Getting access to relevant data in a suitable format can be challenging
  - External data may not be ‘owned’ by the submitting company, and so we may need to *approximate* data – does this add to uncertainty, or does it not matter?

- If ‘standard’ methods produce implausible estimates, then we have no choice but to look to other methods which are usually more complex
  - After all, these complex methods have been developed to address limitations of simple approaches – but how do we define ‘implausible’?



- Even if we think our method is the very best approach, decision makers will likely want to see alternative approaches
  - Are we then ‘wasting our time’ trying to seek out an optimal extrapolation, or should we focus on finding a plausible range of estimates?

Addressing the elephant in the room…

- Fundamentally, trial populations tend to be fitter than ‘real-world’ populations
  - For HTA, we are interested in **real-world** outcomes – should we *really* be using absolute survival estimates from a trial population?
  - Maybe we should focus our efforts on estimating the treatment effect and applying this to a baseline survival estimate using real-world data?
  - … but what do we do when we have rare diseases, single-arm studies, disagreements about the suitability of external data for decision-making?
  - … also, what about the treatment effect? Can we generalize trial-derived effect measures to a real-world population? What about time-dependency?

Key: HTA, health technology assessment.

Delta Hat

Thank you!

Contact: abullement@deltahat.com

Delta Hat