The Place of Artificial Intelligence in HTA and HEOR

We are joined today by three panelists

Lorna Dunning

Senior Technical Adviser – Methods, NICE

Dr. Noémi Kreif

Senior Research Fellow, Centre for Health Economics, University of York

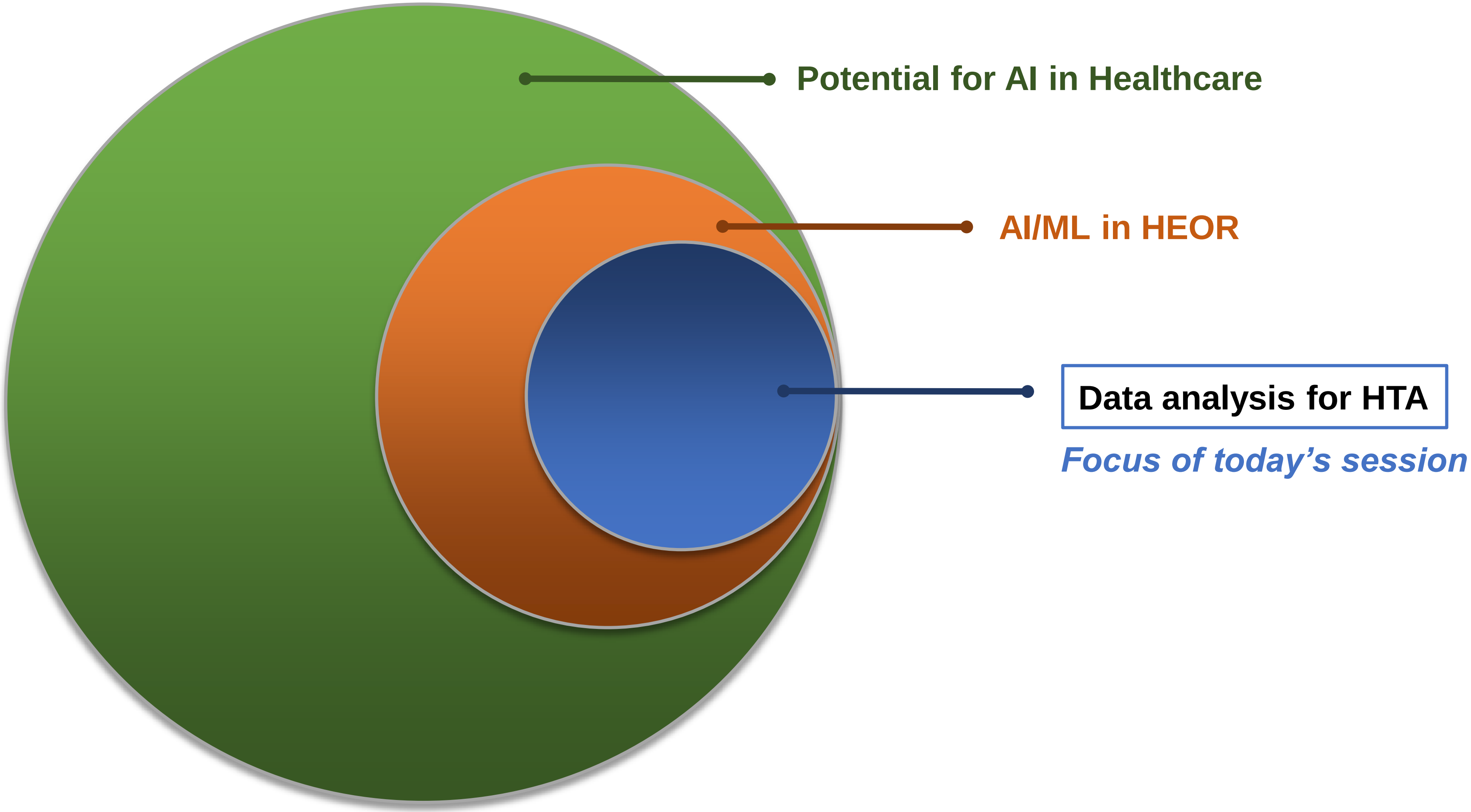
Dr. Pearl Gumbs

PhD, Value Lead & Global HEOR at Boehringer Ingelheim

Today we will discuss the use of AI/ML in HTA and HEOR

- Unlike other industries, use of AI/ML in HTA is widely considered to be in its infancy
- HTA bodies are beginning to recognize the value of AI/ML, such as in the selection of covariates in survival models, but use cases are few
- Existing concerns around the use of AI/ML in HTA include questions on appropriateness, transparency, and accessibility to non-experts
- However, AI/ML methods have proven advantages over traditional approaches in cases of multi-variable datasets often seen in healthcare
- ***How then can the benefits of these methodologies be integrated with concerns to find a measured approach?***

The focus of today's session is the use of AI/ML in HEOR for data analysis in HTA



AI/ML has potential applications across many aspects in HEOR, but as we've discussed, some will be more suitable than others

Health economics and outcomes research

Outcome research

Outcome estimation

Cohort selection

Identify treatment and control groups for a study

How?

- Classify patients based on outcome of treatment
- Cluster patients based on non-pre-specified characteristics
- Select patients based on characteristics in health electronic records

Why AI/ML?

- No need to pre-specify features for classification
- Consider non-linear/additive data

Causal inference

Identify if an intervention impacts an outcome

How?

- Identify confounders of clinical effect
- Estimate propensity score
- Estimate risk of a health outcome due to a change in a particular factor

Why AI/ML?

- Potential for better causal inference
- Potential to reduce payer uncertainty on intervention effect

Predictive modelling

Relate an outcome to inputs (i.e. patient characteristics)

How?

- Predict output based on input features
- Reduce covariates in predictive model
- Estimate time to an event

Why AI/ML?

- Potential for better predictive modelling
- Reduce overfitting, improve generalization

Economic evaluation

Examine factors that impact cost and quality of healthcare

How?

- Support cost-effectiveness analysis by predicting costs

Why AI/ML?

- Potential to reduce uncertainty in cost-effectiveness analysis
- Similar potential gains to predictive modelling due to overlap in the objective of prediction

Knowledge synthesis

Synthesize or extract information contained in text

How?

- Global value dossier generation
- Q&A Chatbot to query large text documents
- Auto-summarization of market research reports

Why AI/ML?

- Higher efficiency
- Broader and quicker access to information
- Potential for better insights from information

Key vocabulary for today's discussion

Machine Learning

Mathematical techniques that give computers the ability to learn without being explicitly programmed in the task by a human

Black box

A system or device whose internal workings are hidden, or not readily understood by the human user

Explainability

A concept similar to "black box" focused on the understanding/interpretation of model outputs essential in healthcare decision making

Validation

A standard process by which a new scientific method/model is tested and proven to be effective against a typical benchmark

The Place of Artificial Intelligence in HTA and HEOR

Lorna Dunning
Senior Technical Adviser – Methods
Centre for Health Technology Evaluation
NICE

NICE National Institute for Health and Care Excellence

About NICE

NICE helps practitioners and commissioners get the best care to patients, fast, while ensuring value for the taxpayer.

We do this by:

- Producing useful and useable guidance for health and care practitioners.
- Focusing on what matters most by prioritising topics that are most important to the health and care system or address an unmet need.
- Providing rigorous, independent assessment of complex evidence for new health technologies.
- Encouraging the uptake of best practice to improve outcomes for everyone.

The health and care system is changing

We're seeing:

- health service pressures
- shared decision making
- growth in innovation
- vast amounts of data.

NICE is transforming too

NICE's core purpose remains the same: to help practitioners and commissioners get the best care to people fast, while ensuring value for the taxpayer.

But as the NHS transforms to meet future challenges, we need to play our part too.

NICE

Possibilities with AI…

- Efficiencies and improvements for internal processes at NICE, reducing errors, and improving critical thinking skills.
- NICE has used NLP and machine learning to streamline and automate various processes, including;
  - Objectively-derived search strategies
  - Priority screening
  - RCT classifier
  - Build Your Own Classifier (in EPPI Reviewer 4)
  - NORMA (NICE ONS Recommendation Matching Algorithm) created to speed up our guideline surveillance work.

NICE

10

Assessing and using AI in health technology assessment

Limited use of AI or ML in Health Technology Assessment (HTA) so far, but it is increasingly being explored.

Trial design or cohort selection → Systematic evidence reviews → RWE and data processing → Predictive analysis or end point analysis → Optimising health economic models → Composition of value assessment dossier

- Systematic evidence reviews: Direct: use of NLP in SLR update for NICE submission (ongoing TA)
- RWE and data processing: Indirect. E.g., cohort selection from EHR using ML and Flatiron Health database (Birnbaum et al. 2019) could be used in TA
- Predictive analysis or end point analysis: Direct: ML model used to minimise bias in covariate selection (TA657)

Challenges around the use of AI in HTA

- Ensuring the rigour, transparency, explainability, and reproducibility of AI/ML:
  - concerns regarding how AI systems arrive at their decisions or recommendations
  - challenging to identify and address potential biases or errors
- AI systems can introduce or reflect bias and discrimination:
  - in patterns of health discrimination within RWE datasets
  - in data representativeness
  - in human choices made during the design, development, and deployment of models.
- Possibilities of ‘hallucinations’ or ‘confabulation’ where AI systems can sometimes generate inaccurate outputs

Addressing these problems requires proactive measures and ongoing evaluation and development.

Assessing and using AI in health technology assessment

Currently, there is no specific guidance in the use or reporting of use of AI in submissions to NICE we are guided by:

1. PALISADE checklist: ISPOR's ML Task Force
2. NICE internal AI principles
3. RWE framework
4. Methods manual

NICE

Guiding use:

The PALISADE Checklist—key considerations for evaluating the transparency of ML

Developed by NICE staff as part of membership of ISPOR's ML Task Force

Element	Definition
Purpose	Is the purpose of the algorithm clearly stated at the outset? Is the implementation of the algorithm in a healthcare setting fair and ethical?
Appropriateness	Is there a clear justification that the algorithm is acceptable in the context within which it is being applied?
Limitations	Have the strengths and limitations, in the context of the purpose, been identified? This should cover both the algorithm and any data used.
Implementation	Consideration of access, implementation, and resource issues when implemented in healthcare settings.
Sensitivity and specificity	For classification algorithms, has the model performance and accuracy (specificity and sensitivity) been appropriately evaluated?
Algorithm characteristics	Has the ML mechanism been clearly characterized and described? Is there sufficient transparency for the results to be reproducible?
Data characteristics	Is the selection of data sets justified and are the key characteristics known? This should extend to training sets, test sets and validation sets.
Explainability	Are the outputs of the algorithm clearly understandable by both the healthcare professional and the patient?

NICE

Guiding use:

NICE Real World Evidence framework – AI/ML content

NICE's RWE framework

- Describes how RWE supports guidance and best-practice principles for conducting and reporting RWE

Current content on AI/ML

"Where human abstraction or artificial intelligence tools are used to construct variables from unstructured data, ***the methods and processes used should be clearly described and their validity documented****."*

"The selection of covariates may use advanced computational approaches such as machine learning to identify a sufficient set of covariates, for example, where the number of potential covariates is very large (Ali et al. 2019, Tazare et al. 2022). ***The use of these methods should be clearly justified and their consistency with causal assumptions examined.*** *Choosing covariates based on statistical significance should be avoided"*

NICE

Guiding use:

Programme methods and process manuals

Centre for Guidelines Manual, section 6.1

"A percentage (at least 10%, but possibly more depending on the review question) should be screened independently by 2 reviewers (that is, titles and abstracts should be double-screened). …

Priority screening refers to any technique that uses a machine learning algorithm to enhance the efficiency of screening. Usually, this involves taking information on previously included or excluded papers, and using this to order the unscreened papers from those most likely to be included to those least likely. This can be used to identify a higher proportion of relevant papers earlier in the screening process, or to set a cut-off for manual screening, beyond which it is unlikely that additional relevant studies will be identified…."

Centre for Health Technology Evaluation Manual, section 3.4.5

"More than 1 reviewer should assess all records retrieved by the search strategy to increase the validity of the decision. Clearly report the procedure for resolving disagreements between reviewers."

NICE transformation
Ongoing work to consider alignment within NICE

NICE

Ongoing work within NICE on assessing and using AI in real world and trial data

- Updates to the RWE framework:
    - Aim to include best practice principles for diagnostic and predictive studies, where appropriate will add further considerations for the use of AI and machine learning in these kinds of studies.
    - Potential for development of sections covering use of AI/ML for unstructured data, federated data networks, and causal analysis in the future.
- HTx project:
    - Using AI methods for analysis of real-world clinical and economic data and comparing them to traditional statistical methods
- The ISPOR ML Methods Emerging Good Practices Task Force:
    - NICE staff continue to support developing guidance for HEOR and decision makers on the use of ML methods, specifically for analysing observational data
- Information Services and Centre for Guidelines continue exploring the use of using NLP in improving/automating searches

Next steps and alignment to NICE's principles

1. Highlight areas of high or low risk of use of AI in submissions to NICE
    - Human oversight and augmentation, not replacement
2. Developing standards, validation (model performance) and evidence requirements for use in NICE submission
    - Transparent and explainable AI algorithms and models
3. Exploring where and how expertise in AI can contribute to NICE appraisals, either via committee membership and/or supporting evidence critiques
    - Identify and minimise biases, ensuring data is representative

The Place of Artificial Intelligence in HTA and HEOR

The “academic perspective”

Noémi Kreif

Senior Research Fellow
Centre for Health Economics
University of York

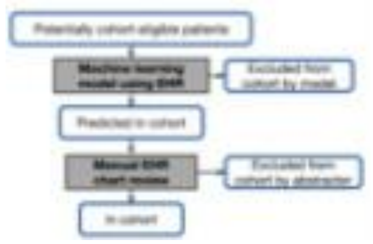
How can academics contribute?

HEOR methodologists (biostatistics, epidemiology, economics, health economics, decision science) all started embracing machine learning

Map out relevant areas where ML can contribute

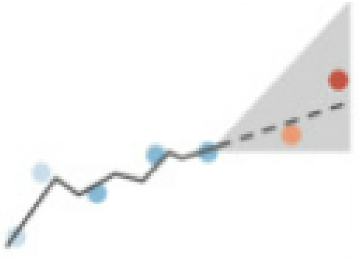
The ISPOR ML taskforce made a good start, but more details needed

Cohort Selection

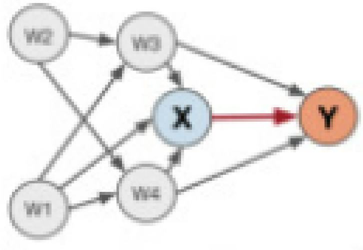


Feature Selection

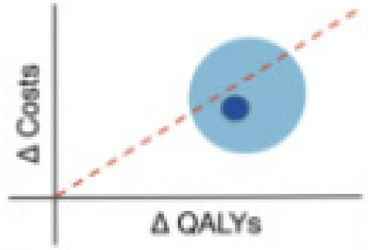
Predictive Analytics



Causal Inference



Economic Evaluation



- Survival analysis?

- Prognostic scores?
- HRQoL mapping?

- Analysis of RWD?
- Subgroups in RCTs?

In practice: workshops with academics, industry and decision makers -> publications

Some ML methods need development to be used for HEOR

More mature

- ML for predictive modelling (e.g. development of risk scores)
- ML for feature selection
- Causal (ML) to estimate average treatment effects from RWD

Less mature

- (Causal) ML for subgroups identification
- (Causal) ML for individualized treatment recommendations
- ML for extrapolating survival
- ML for mapping HRQoL measures

In practice - Scoping reviews, systematic reviews of available methods

Research can make ML methods better understood

Demonstrate value of existing tools for HTA

Assess relative advantages of ML (versus non-ML) methods for HEOR tasks

Create ML tools with improved abilities for HEOR tasks (e.g. uncertainty)

- Method illustrations
- Simulation studies
- Tutorials
- Original methods research

In practice - Designated funding streams by national, international funders & decision makers

Guideline development

- methods guidelines
- reporting guidelines
- decision support doc/s

In practice:

- **Workshops:** academics (including ML experts), industry and decision makers
- **Designated funding streams**

- New guidelines for AI/ML in HEOR can **build on related frameworks** (e.g. regulatory, AI as medical device, prognostic modelling etc.)
- **Guidance can mandate key processes** (e.g. independence of training/test data)
- **Emphasis on transparent reporting**, use of non-proprietary software

Protocol for the development of an artificial intelligence extension to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022

Claire Hawksworth, Jamie Elvidge, Saskia Knies, Antal Zemplenyi, Zsuzsanna Petykó, Pekka Siirtola, Gunjan Chandra, Divya Srivastava, Alastair Denniston, Anastasia Chalkidou, Julien Delaye, Petros Nousios, Manuel Gomes, Tuba Saygin Avsar, Junfeng Wang, Stavros Petrou, Dalia Dawoud

doi: https://doi.org/10.1101/2023.05.31.23290788

Advisory roles

- Academics can support **decision makers** as part of external assessment groups
  - ✓Capacity building in ML necessary

- Collaboration with **industry** can increase relevance of research
  - ✓Independence of research needs to be maintained
  - ✓Use of synthetic data could be explored

Final thoughts

ML *can* increase transparency in data analysis

E.g. versus ad-hoc model specification

Simpler is not always better

More flexible methods can lead to less biased estimates and better decisions

HTA community embraced complex methods before (e.g. treatment switching)

Usual good practices of data analysis should apply

State assumptions of method

Report choices

Conduct sensitivity analysis

The Place of Artificial Intelligence in HTA and HEOR

The “industry perspective”

Dr. Pearl Gumbs

Value Lead & Global HEOR at Boehringer Ingelheim

Our industry faces significant challenges; AI/ML has potential to help

- Financial pressure on health systems has increased, leading to **greater scrutiny of product value** claims and **expanded use of new tools to manage access**
- There is an evolution towards **higher global evidence thresholds**, representing a significant commercial challenge
- Policy changes, such as the US Inflation Reduction Act, the GKV Stabilization Act in Germany, and the EU HTA Regulation, are accelerating the **emergence of increasingly global evidence thresholds**
- The shift in the evidence environment has already begun, and it is increasingly critical to **demonstrate meaningful, patient-relevant benefit over the standard of care**

However, the number of published HTAs that mention machine learning is currently very low

<10

English-language publications only

Using machine learning in HTA is in the earliest days, and positive impact is not always guaranteed (1/2)

USE CASE #1 – MULTIPLE MYELOMA

1. In the original submission, a range of **clinician-identified covariates** (i.e., age) were used in the clinical model to estimate survival

2. However, the HTA agency found that the choice of these covariates was **unclear and lacked sufficient justification**

3. In response to this critique, the company presented a range of methods to select for covariates, **including an AI/ML method**

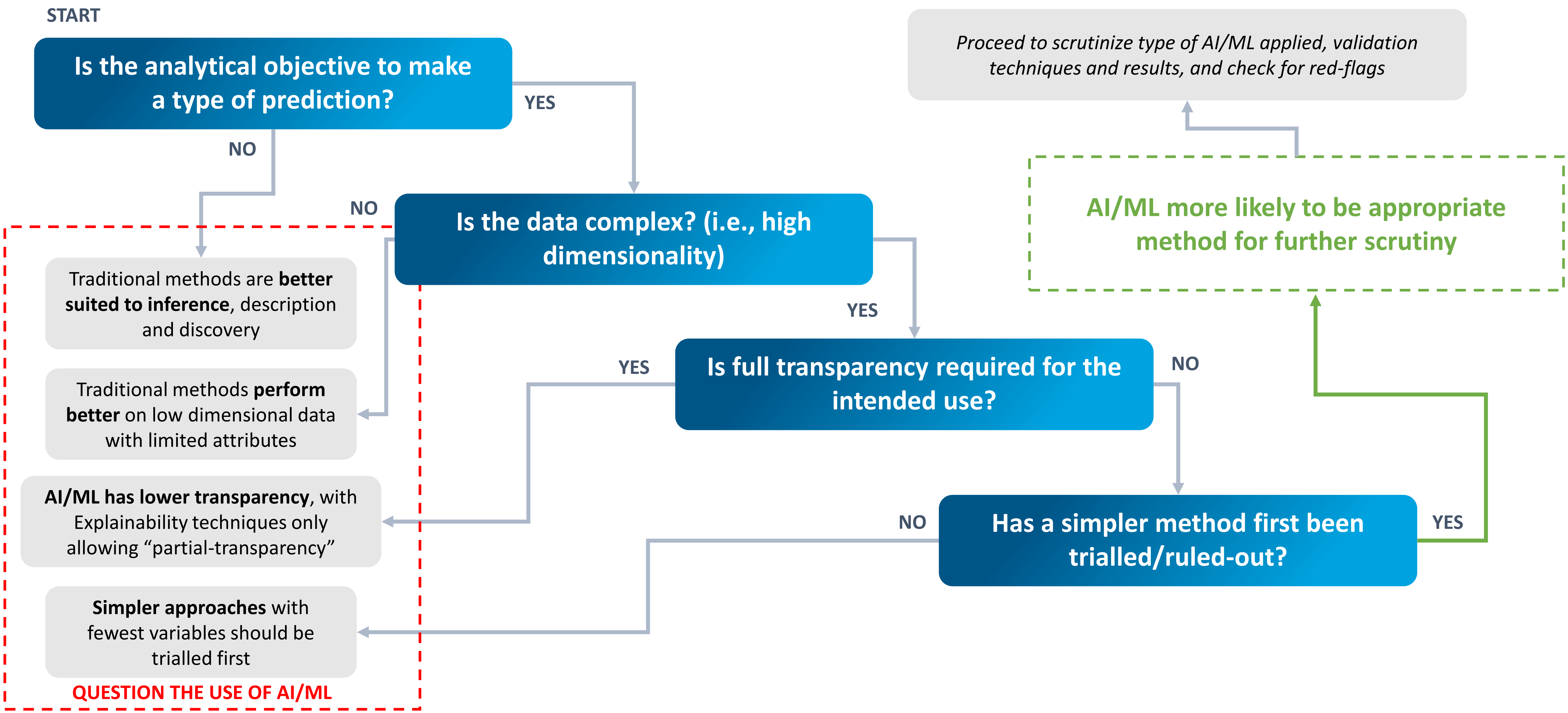
4. After review of the submitted methods for covariate selection, **the agency preferred the AI/ML-based LASSO method** for this task

Using machine learning in HTA is in the earliest days, and positive impact is not always guaranteed (2/2)

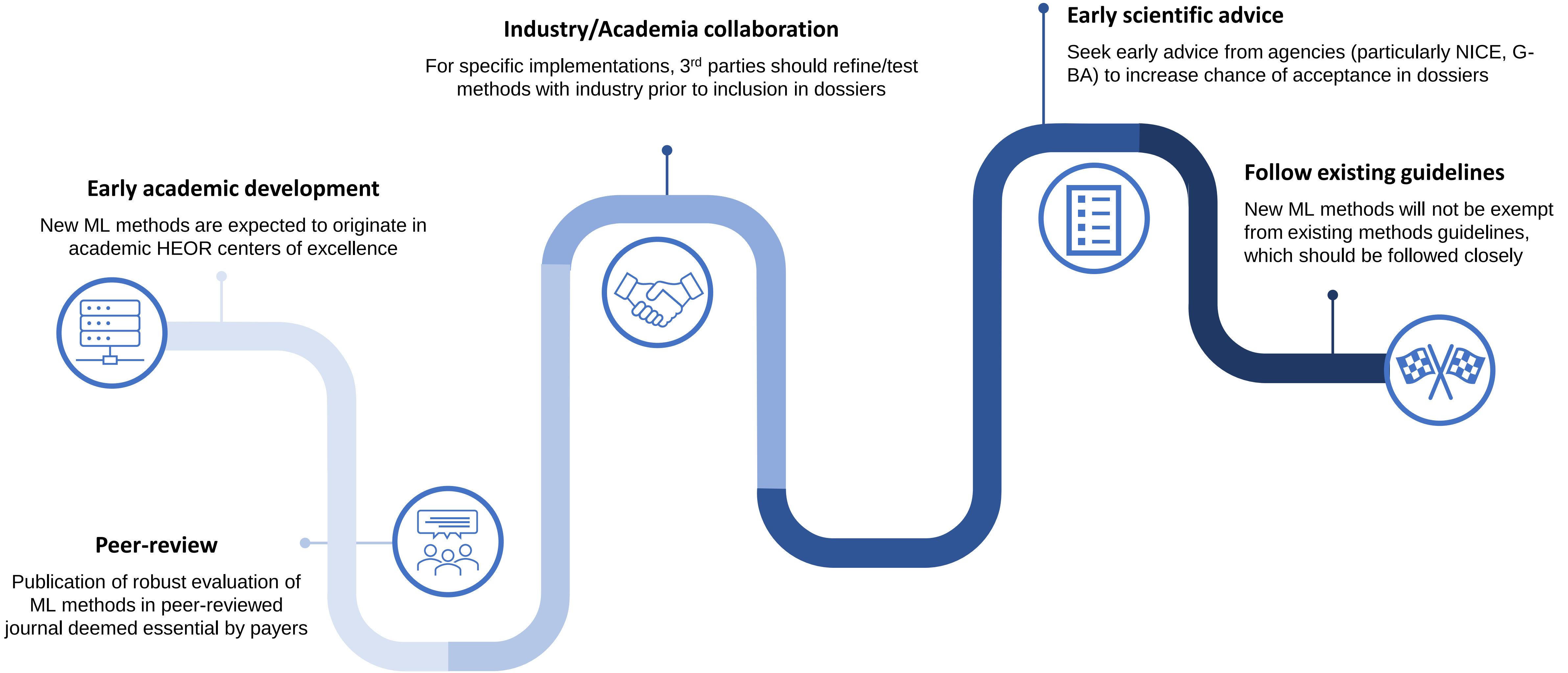
USE CASE #2 – B-CELL LYMPHOMA

1. The company used a cure-mixture model for the survival analysis assuming that a proportion of patients became cured
2. Instead of using a standard cure-mixture modelling software package, **the company developed its own code** which included expectation-maximization clustering, an AIML algorithm
3. The **HTA agency expressed concerns on whether the clustering algorithm was appropriate** given the limited number of patients and the choice of prognostic factors
4. The agency thought that the cure assumption lacked evidence support and concluded that the company's model was **not suitable for decision making**

Use cases should be evaluated case-by-base, but there are simple rules we can all follow



Once a use case is identified, a simple roadmap can help ensure multi-stakeholder buy-in



We hope today’s session has helped answer some key questions for ML in HEOR; and we will now open for Q&A

What is the value that ML provides to payers and HTA agencies for decision making?

How can industry foster development and implementation of new ML methods in HEOR?

What are some of the key risks / drawbacks of ML methods and how can they be overcome?

How can we identify use cases when ML should or <u>should not</u> be implemented?