HTA364

Can Artificial Intelligence and Machine Learning Be Used to Demonstrate the Value of a Technology for HTA Decision-Making?

Szawara P[1], Zlateva J[2], Kotseva F[2], Stoyanov P[2], Guerra I[3], Halmos T[3], Wagner P[4], van Engen A[5]

[1]IQVIA, Warsaw, Poland, [2]IQVIA, Sofia, Bulgaria, [3]IQVIA, London, UK, [4]IQVIA, Frankfurt, Germany, [5]IQVIA, Amsterdam, Netherlands

Introduction and objectives

Artificial intelligence (AI) and machine learning (ML) methods have potential applications in Health Economics and Outcomes Research (HEOR) and Health Technology Assessment (HTA). However, the use of AI/ML is still very new and not much research has been conducted on its use in HTA. This study aimed to examine the current use of AI/ML in HTA and how this might evolve in the future.

Methods

- We conducted the search using IQVIA Market Access Insights (previously known as HTA Accelerator) and Linguamatics platforms, with the relevant keywords (machine learning, artificial intelligence, neural network, deep learning, supervised learning, principal component analysis and LASSO) to identify HTA reports that mentioned AI/ML methods
- The keywords were also translated to German, French, Italian and Spanish
- The analysis focused on HTAs from seven countries (Australia, Canada, France, Germany, Italy, Spain and United Kingdom) and EUnetHTA published between 2012 and 2023
- For the identified HTAs, we assessed whether using the AI/ML methods had an impact on the final HTA decision
- Additionally, HTA policies and methodology guidelines were reviewed to see if they include any recommendations on using AI/ML in HTA

Results

- The analysis identified 11 HTA reports (4 from the UK, 6 from Canada, 1 from Italy) where AI/ML methods were mentioned (Figure 1)
- NICE HTAs mentioned AI/ML methods used for selecting covariates to either adjust for imbalances in the baseline characteristics in subgroups or to improve the economic model fit, for determining the probability of cure in the economic model and to predict exacerbations based on patients' self-reported symptoms (Figure 2)
- Six CADTH HTAs mentioned principal component analysis used to validate patient-reported outcomes (PRO) instruments (Figure 2)
- PRHDM (Lombardian HTA body) evaluated the ultrasound AI-based software for the identification of late myocardial damage and compared it to diagnosis based on standard cardiac magnetic resonance (Figure 2)
- In only two of the identified HTAs, the use of AI/ML had a positive impact on the HTA outcome. For Kyprolis in multiple myeloma, using the preferred LASSO method to estimate the overall survival (OS) in the model in NICE HTA, contributed to changing the preliminary negative recommendation to positive. For ultrasound AI-based software for identification of myocardial damage in oncology patients, PRHDM recognized the advantages of AI-based technology (Figure 2)
- In one of the identified HTAs, a cure-mixture model using ML methods was criticized by NICE as it was not validated for use in small sample sizes (Figure 2)
- In the review of HTA policies and methodology guidelines, we found France, Italy, Germany and EUnetHTA mentioning the use of AI/ML for medical devices (France), in clinical trials (Italy, France) or in conducting systematic literature reviews of clinical evidence (Germany, EUnetHTA) (Figure 1)
- In Canada, horizon scanning reports with an overview of clinical applications of AI/ML have been published. One of the reports, mentioned the application of AI/ML in health research and drug discovery and development, e.g. to discover new biomarkers to improve diagnosing of mental disorders, to identify mutations causing disease and help predict the effects of treatments, to generate hypotheses for future research and to automate data extraction and searches in systematic reviews and HTAs

Conclusions

- **To date, the usage of AI/ML in HTA remains very limited and does not seem to have much impact on HTA outcomes**
- **The acceptance and impact of AI/ML may be expected to gradually increase as the HTA agencies begin to consider these methods in their methodology guidelines**

Legend: ⚠ Positive impact on HTA decision; ⚠ Negative impact on HTA decision; ⚠ No impact on HTA decision

Limitations: The analysis only considered publicly available information. Some HTA bodies might not have clearly specify what methods were used in the HTA submission.

Abbreviations: 2L+: Second and further line of treatment; AI: Artificial intelligence; BR: Bendamustine + rituximab; CADTH: Canadian Agency for Drugs and Technologies in Health; COPD: Chronic obstructive pulmonary disease; DLBCL: Diffuse large B-cell lymphoma; ERG: Evidence Review Group; EUnetHTA: European Network for Health Technology Assessment; G-BA: Gemeinsamer Bundesausschuss; HEOR: Health Economics and Outcomes Research; HTA: Health Technology Assessment; ICER: Incremental Cost-Effectiveness Ratio; IQWiG: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; ML: Machine Learning; MRI: Magnetic Resonance Imaging; NICE: The National Institute for Health and Care Excellence; OS: Overall survival; PRHDM: Programma Regionale HA Dispositivi Medici; PRO: Patient reported outcomes; RCT: Randomized controlled trial; Rd / d: Lenalidomide + dexamethasone or dexamethasone alone

Reference: IQVIA Market Access Insights; (1) NICE (2022) 'NICE health technology evaluations: the manual' (2) G-BA (2021) 'Entscheidung über die Gewährung der sekundären Datennutzung – Institut für Ökonometrie und Statistik der Universität zu Köln' (3) IQWiG (2023) 'Allgemeine Methoden. Version 7.0' (4) CADTH Emerging Tech (5) AIFA (2021) 'Guide to the submission of a request for authorisation of a Clinical Trial involving the use of Artificial Intelligence (AI) or Machine Learning (ML) systems' (6) HAS (2021) 'Methodological guide. Real-world studies for the assessment of medical products and medical devices' (7) HAS (2019) 'Guide to the specific features of clinical evaluation of CMD in view of its application for reimbursement' (8) HAS (2020) 'A new tool to evaluate medical devices using AI – Press Release' (9) HAS (2019) 'Assessing medical devices embedding AI' (10) HAS (2020) 'LPPR: Dossier submission to the medical device and health technology evaluation committee' (11) EUnetHTA (2019) 'Methodological guidelines: Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness'

Figure 1. AI/ML in HTA records: Prisma search diagram

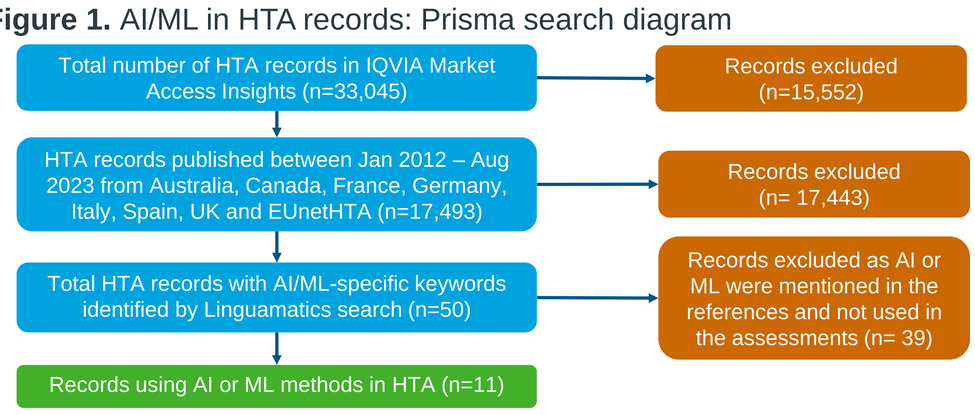

Table 1. HTA methodology guidelines mentioning AI/ML

	• For HTA, NICE details preferred methods, and any deviation must be **clearly justified**[1] • **It is not enough** that a deviant method has been previously accepted by NICE in HTA
	• In 2021, the G-BA and IQWiG granted permission for the **use of real-world perinatal data for AI/ML-analysis**[2] • In the General Methods guideline, IQWiG mentions that **validated classifiers based on ML** (e.g. RCT classifiers) can be used in **search strategies** for systematic literature reviews of clinical evidence. ML approaches (e.g., prioritization, application of classifiers) can also be used to support **study selection and prioritization**[3]
	• CADTH has published **multiple horizon scanning reports covering AI/ML**[4]
	• In 2021, AIFA published official guidelines for the **submission of clinical trials involving the use of AI/ML methods**[5]
	• HAS suggests that AI/ML **may be used in clinical trials** in the future, but recognizes that there are **risks and limitations to be considered**[6] • It has released multiple guides that consider **AI/ML for medical devices**[7-10]
	• EUnetHTA accepts **ML-based RCT classifiers for limiting search results** in the search strategies for systematic literature reviews on clinical effectiveness[11]

Figure 2. Case study analysis of AI/ML use in HTA, critique and impact

⚠ Kyprolis + Rd / d for 2L+ multiple myeloma (2017)

Use of AI/ML techniques:

- Post-hoc subgroup analyses were performed by the manufacturer to estimate overall survival (OS) in the model, using a Cox proportional hazards model to adjust for imbalances in the baseline characteristics
- The ERG criticized the unclear choice of covariates and preferred AI/ML-based alternative submitted in response by the manufacturer (**LASSO method**)

Critique & outcome:

- The preliminary recommendation was negative, mainly due to uncertain OS estimations. Using the preferred LASSO method to estimate the OS in the model, contributed to changing the preliminary negative recommendation to positive

⚠ Polivy + BR for relapsed / refractory DLBCL (2020)

Use of AI/ML techniques:

- A cure-mixture model was used, classifying patients into long-term vs non-long-term survival clusters, by applying standard ML methods (e.g. **clustering using the expectation–maximisation algorithm**)

Critique & outcome:

- The ERG criticized the algorithm as it was designed for big datasets and may not be suitable for determining the cure probabilities from a dataset of 80 patients
- NICE preferred standard parametric modelling as the cure-mixture model was implausible due to the highly uncertain estimated cure rate. This increased the ICER and higher discount was needed to reach cost-effectiveness

⚠ Ultrasound AI-based software for identification of myocardial damage in oncology patients (2019)

Use of AI/ML techniques:

- PRHDM (Lombardian HTA body) compared the new AI-based software for ultrasound imaging with standard cardiac MRI
- Using fully programmable parallel processors allowed to optimally analyzing a greater number of patients in less time

Critique & outcome:

- PRHDM noted growing evidence that new technologies can be used successfully for the assessment of cardiac function
- However, extensive experience prior to routine clinical application and larger studies are needed
- PRHDM did not issue a formal reimbursement recommendation

⚠ myCOPD app for self-management of COPD (2022)

Use of AI/ML techniques:

- In addition to clinical trials, the manufacturer submitted an ongoing study (Chmiel 2020) that used self-reported data on symptoms from myCOPD to predict exacerbation events using **heuristic and ML models**
- Results showed that models showed moderate ability in predicting exacerbations

Critique & outcome:

- NICE noted that evidence on rates of exacerbations was inconclusive
- The Chmiel 2020 study was not mentioned in the Final Guidance and it did not have a clear impact on the HTA outcome
- NICE recommended that further good quality evidence is needed to address uncertainties about myCOPD's clinical benefits

⚠ Skyrizi for moderate to severe Crohn's disease (2023)

Use of AI/ML techniques:

- The submitted economic model included four Markov matrices estimated using ordered probit models. The ERG asked for the justification for choosing the ordered probit approach over alternatives
- The manufacturer considered that selecting additional variables with **a hypothetical use of ML** would make the model even more complex, with no improvement in accuracy

Critique & outcome:

- The ERG retained its critique on the ordered probit models but its preferred approach also did not include ML methods
- No impact as ML was not used neither by the manufacturer nor by ERG

⚠ Saxenda, Beovu, Taltz, Iluven, Kymriah, Rexulti (2017-2021)

Use of AI/ML techniques:

- **Principal component analysis** was used to validate the PRO instruments used

Critique & outcome:

- Validation of PROs was described in the appendices to the clinical reports and was not commented on by CADTH
- The used methods for PROs validation did not have an impact on the final recommendations

Poster presented at ISPOR Europe 2023, 12-15 November 2023, Copenhagen, Denmark

For further information please contact: IQVIA | Market Access Insights | peter.wagner@iqvia.com