lead to the same conclusions as those from the real data. ***Methods:*** A new SGD AI tool employs Conditional Tabular Generative Adversarial Networks (CTGAN), Copula GANs and Sequential Decision Trees (SDT). For the GANs, a dual-network adversarial architecture comprising a generator/ discriminator was used. The performance (e.g. accuracy quality) of each method was evaluated using hospital episodes data (HES) data (N=14,423) in a chronic kidney disease (CKD) population and RCT data (N=670) in a non-small cell lung cancer population (NSCLC). ***Results:*** The CTGAN was trained on 14,428 patient level healthcare resource use data collected between 2012 and 2015. The accuracy of the SGD was high: shape metric scores between 0.905 to 0.942 (i.e. strong similarity while maintaining privacy). These results were higher for the RCT data. In addition, for the RCT data, the reported hazard ratio (HR) of erlotinib vs BSC in the reported trial was 0.94 (95% CI: 0.81,1.10; p=0.462). Using SGD, these were: 0.93 (95% CI: 0.79,1.12; p=0.481). ***Conclusions:*** We show a simple new AI tool that can be used to import the actual data and output corresponding SGD with a high probability of achieving the same decision as the real data. Our findings confirm that SGD can closely replicate real-world healthcare data offering a practical and privacy-preserving alternative to using sensitive patient records.

MSR195
THE AI-IN-RWE TRANSPARENCY (AIRT) CHECKLIST: ESSENTIAL AND DESIRABLE STANDARDS FOR AI-ENHANCED REAL-WORLD EVIDENCE

Tushar Srivastava, MSc[1], Radha Sharma, PhD[2], Raju Gautam, PhD[1], Madhusudan Kabra, MSc[3]
[1]*ConnectHEOR, London, United Kingdom,* [2]*ConnectHEOR, Edmonton, AB, Canada,* [3]*MK Global Consulting, London, United Kingdom*

Objectives: Artificial intelligence (AI) is becoming integral to analytics that transform real-world data (RWD) into real-world evidence (RWE) for regulatory and health technology assessment (HTA) decisions. However, existing reporting checklists, originally designed for conventional statistical studies, seldom address the additional complexity introduced by machine-learning workflows. In AI-enabled RWE, crucial elements such as model training, performance validation, explainability, and bias mitigation are often inconsistently reported or omitted, limiting reproducibility and weakening decision-maker confidence. The growing regulatory focus on trustworthy AI, exemplified by the European Union's AI Act, further reinforces the need for clear documentation of AI system development, validation, and governance, particularly in high-risk domains like healthcare. ***Methods:*** This study outlines a structured, open-access checklist AI-in-RWE Transparency (AIRT) that distinguishes between items considered essential for credibility and those deemed desirable based on context. Checklist items were systematically mapped from regulatory guidance, methodological position papers, recent AI-enabled RWD publications, and stakeholder interviews spanning HTA agencies, life-science companies, and patient representatives. A multidisciplinary working group iteratively refined the checklist content and structure through facilitated discussions and written feedback, with an emphasis on practicality, transparency, and accuracy. ***Results:*** The emerging AIRT checklist captures reporting expectations across domains such as data source, algorithm development, validation processes, interpretability, transparency, adaptive model governance, and communication of uncertainty. A companion digital tool is in development to guide users through each domain, generate structured transparency summaries, and highlight potential reporting gaps. Early pilot mapping of published studies indicates that consistent use of AIRT could improve clarity and streamline both internal and external review processes. ***Conclusions:*** AIRT responds to an urgent gap in guidance for AI-enhanced RWE. By clarifying what must be reported and what should be reported, it aims to support both assessors and assessees in navigating an evolving regulatory landscape while encouraging trust, reproducibility, and informed healthcare decisions.

MSR196
THE AI-ONLY WORKFLOW: MODEL AND REPORT ADAPTATION WITHOUT HUMAN SETUP

Hanan Irfan, MSc, **Tushar Srivastava, MSc**
ConnectHEOR, London, United Kingdom

Objectives: Health economic models and reports often need to be repeatedly adapted as per the requirements of different regions and countries. Model and report adaptation has traditionally required considerable manual effort or rule-based automation through tools like VBA. These conventional approaches are deterministic, lack adaptability, and cannot be classified as true artificial intelligence. This study aims to demonstrate an agentic AI-based implementation that autonomously adapts Excel-based health economic models and corresponding Word-based reports for country-specific use without requiring any manual or rule-based pre-configuration. ***Methods:*** The implementation operates in two modules. In the first module, the user uploads a global Excel model and a list of parameters tailored to the target country. The AI system scans all sheets, identifies relevant tables and cells based on structural cues (titles, labels, table boundaries), and replaces default values with country-specific ones. It then generates a fully adapted Excel model. In the second module, the user provides a global Word report and the adapted Excel model. The AI reviews each section of the global report, automatically compares the data with Excel model, reasons over the changes, interprets them and writes them in the report section by section. The rewritten text does not merely replace data but also updates interpretation without any placeholder abstracted text. Notably, no templating, pre-defined rules, or VBA scripting is used, demonstrating true generative and autonomous functionality. ***Results:*** The tool successfully adapted models and reports across multiple country settings without any time and labour-intensive manual tagging or rule-based configuration. It consistently identified and replaced model parameters and report content with contextual accuracy. ***Conclusions:*** This tool exemplifies the distinction between traditional automation and AI in HEOR workflows. As AI adoption grows, clarity in distinguishing true AI capabilities from automation is vital for credibility and innovation in health technology assessment.

MSR197
THE EFFECTS OF POOLING TREATMENT EFFECTS TARGETING TREATMENT POLICY AND HYPOTHETICAL ESTIMANDS WITH RANK-PRESERVING STRUCTURAL FAILURE TIME MODEL IN ONCOLOGY AGGREGATE-LEVEL META-ANALYSES

Rebecca Metcalfe, MA, PhD[1], Shomoita Alam, PhD[2], Antonio Remiro Azócar, PhD[3], Richard Yan, MSc[4], Jay JH. Park, PhD[1]
[1]*Core Clinical Sciences, Vancouver, BC, Canada,* [2]*Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada,* [3]*Novo Nordisk, Madrid, Spain,* [4]*Department of Statistical and Actuarial Science, Simon Fraser University, Vancouver, BC, Canada*

Objectives: The implications of the estimands framework, which emphasizes the importance of post-randomization (intercurrent) events and their analytical strategies in randomized clinical trials (RCTs), have been under-explored for meta-analysis. For RCTs in oncology, rank-preserving structural failure time modelling (RPSFTM) is often recommended to adjust for the intercurrent event of treatment switching, since failure to account for treatment switching can produce misleading results for overall survival (OS). Using simulations, we examined the bias and coverage that can be caused by combining trial evidence that estimates different target estimands in a meta-analysis of RCTs. ***Methods:*** We simulated OS data for eight RCTs that allowed patients in the control group to switch to the intervention treatment after disease progression. We estimated treatment effects that ignored treatment switching (treatment policy estimand) and another that accounted for switching with RPSFTM (hypothetical estimand). These results were pooled via aggregate-level meta-analyses with varying the proportions of treatment policy and hypothetical effect estimates. ***Results:*** On average, meta-analyses pooling only hypothetical estimates derived via RPSFTM produced larger treatment effects than the meta-analyses that pooled only the treatment policy estimates. This was consistently observed across all scenarios with different randomization ratios and treatment switching rates. Bias and coverage were directly influenced by the concordance between the estimands pooled and the meta-analytic target, with greater discordance resulting in more bias and poorer coverage. ***Conclusions:*** We found that combining treatment-policy and RPSFTM hypothetical estimates yields pooled effects that correspond to neither estimand, potentially leading to misleading conclusions even with random-effects models. Applying the estimands framework and ensuring that trial-level estimands are aligned with meta-analysis should improve the relevance and validity of synthesized evidence.

MSR199
THE HERC DATABASE OF MAPPING STUDIES: A 2025 UPDATE OF A LIVE SYSTEMATIC REVIEW OF STUDIES PREDICTING EQ-5D FROM OTHER PATIENT-REPORTED OUTCOME MEASURES

Kristian Mallon, BSc, MPhil, MSc[1], Rositsa Koleva-Kolarova, BSc, MSc, PhD[2], Richeal Maria Burns, MSc PhD[1], Yaling Yang, PhD[3], Helen Dakin, MSc DPhil[2]
[1]*Atlantic Technological University, Sligo, Ireland,* [2]*Health Economics Research Centre, University of Oxford, Oxford, United Kingdom,* [3]*University of Oxford, Oxford, United Kingdom*

Objectives: The Health Economics Research Centre (HERC) Database of Mapping Studies was established in 2013. It comprises a live systematic review of studies developing mapping algorithms predicting EQ-5D. Mapping (or cross-walking) non-preference-based measures to preference-based measures allows retrospective estimation of utility scores, supporting the use of existing clinical data in cost-utility analyses. ***Methods:*** Studies were included in the database if they conducted statistical mapping to predict EQ-5D utilities or responses from any source instrument and reported the estimated algorithms in sufficient detail to allow other researchers to use them to predict EQ-5D in other studies. The original searches conducted in December 2012 were updated nine times between July 2013 and February 2025. The database records key characteristics of each study: author, year, sample size, source and target instruments, disease, model type, and validation methods. ***Results:*** As of 2025, the HERC Database includes over 279 mapping studies across a wide range of clinical areas, including cancer, infectious diseases, cardiovascular disease, digestive system diseases, central nervous system disorders, and musculoskeletal disorders. Around 17 studies were published in 2023, 13 in 2024, and 2 in 2025. The database

includes studies mapping from 113 different source instruments. Ordinary least squares (OLS) regression remains the most frequently applied model (81%), though more recent studies have adopted advanced techniques such as beta regression, mixture models, and machine learning to potentially improve prediction accuracy. ***Conclusions:*** The HERC Database of Mapping Studies provides a convenient and timely resource for identifying mapping studies. As interest in real-world data increases, the database remains an important resource for economic evaluations and health technology assessment. Ongoing updates and expansion of the database, and particularly the inclusion of validation performance and newer modelling approaches will enhance its applicability and relevance for future economic evaluations.

MSR200
THE NATURAL HISTORY OF PANCREATIC CANCER AND THE EFFECTIVENESS OF SCREENING A MODEL-BASED ANALYSIS

Robert Wittram, MSc, Christian Brettschneider, PhD, Hans-Helmut Koenig, MD
Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Objectives: Early detection of pancreatic cancer is crucial for reducing mortality, as it enables patients to access curative treatments before metastasis. Blood-based tests could improve screening, which will be modelled in this study. The aim is to assess average life-years-gained per individual detected compared to no screening in a hypothetical population. ***Methods:*** A microsimulation model simulates the progression of pancreatic cancer in 100,000 individuals at 1.9% lifetime risk. The full natural history from low-grade precancerous lesions to metastatic pancreatic cancer in male and female individuals can be assessed. Cumulative proportions from the literature are used to calculate pancreatic cancer-related survival. Individuals are screened four times starting at age 55 in 1-year intervals. Stage-specific test parameters and full screening adherence were assumed. The comparator was pancreatic cancer diagnosis due to e.g., symptoms. ***Results:*** Screening resulted in an average life-year gain of 1.6 years per true-positively screened individual. The 5-year survival rate was 29% vs. 5% when true-positively screened versus not screened. Overall, 44 individuals had a true-positive test result to receive subsequent cancer treatment. Among those patients, 36 received an early diagnosis at stage 1 or 2. Sensitivity analysis was conducted to test the impact of the screening start age on the outcome. Starting screening later than 55 increased the average life-year gain up until age 67, before decreasing this gain. A specificity of 90% or 95% would lead to 30,000 or 17,500 false-positive results. ***Conclusions:*** Pancreatic cancer screening in the normal population would result in an unacceptable number of false-positive and few true-positive patients. This model will investigate which threshold risk, test parameters, and optimal screening protocols, could enable early detection under certain circumstances. Evaluating the cost-effectiveness of blood-based testing for pancreatic cancer will be the ultimate purpose of the model.

MSR201
THE ROLE OF QUALITATIVE RESEARCH IN RANDOMIZED CONTROLLED TRIALS INVOLVING DRUGS AND MEDICAL DEVICES: A SYSTEMATIC MAPPING REVIEW

Yuan-Tao Huang, MSc, LLM[1], Benjamin Gregory, MSc[2], Sharon Greenwood, PhD[3], Evi Germeni, PhD[1]
[1]*Health Economics and Health Technology Assessment (HEHTA), School of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom,* [2]*Social Medicine, Leeds Institute of Health Sciences, School of Medicine, University of Leeds, Leeds, United Kingdom,* [3]*Public Health, School of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom*

Objectives: The benefits of using qualitative research with randomised controlled trials (RCTs) are increasingly recognised. With both fields undergoing significant changes, this review documented the value of qualitative research when conducted with RCTs involving drugs and medical devices. ***Methods:*** We systematically searched five electronic databases (MEDLINE, EMBASE, PsycINFO, CINAHL, ASSIA), as well as the NIHR Journals Library, to identify original, peer-reviewed qualitative research published between 1 January 2019 and 31 December 2020, and conducted in the context of a specific RCT involving a drug or a medical device. Examples illustrating how qualitative research contributed to the RCTs were identified through line-by-line reading and categorised into an existing framework that was gradually refined. This led to the development of a list of questions that qualitative methods can address in the context of trials, and to deductions about the roles that qualitative inquiries can play. ***Results:*** We included 208 qualitative studies, of which 22.1% (46/208) explicitly articulated their value to the trial. A total of 780 examples demonstrating how qualitative methods had enhanced the trial endeavour were identified and categorised into 27 subcategories. The most frequently identified subcategory focused on assessing the 'feasibility and acceptability of interventions in practice' (19.5%, 152/780). Also, 60 research questions were developed that reflect the nuanced complexity and contextual diversity encountered when conducting trials and implementing interventions in the real world. ***Conclusions:*** While qualitative methods have conventionally been associated with complex interventions, this review highlights their expanding relevance to trials involving drugs and devices. The proposed list of questions may serve as a reference for those exploring how qualitative methods could be applied across different stages of trials. In addition to its established roles, future trials could further benefit from expanding the use of qualitative methods to deepen the understanding of outcomes, strengthen measurement validity, and capture experiences of health conditions.

MSR202
THE STATISTICAL ABYSS: REAL-WORLD EVIDENCE FOR HEALTH TECHNOLOGY ASSESSMENT

Martin Scott, BSc, MSc, **Kerry Mueller, B.Sc. M.Sc**
Numerus, Reutlingen, Germany

Objectives: To highlight the methodological challenges and strategic considerations in using real-world evidence (RWE) for health technology assessment (HTA) in the European Union, particularly under the Joint Clinical Assessment (JCA) framework. ***Methods:*** This conceptual analysis draws on regulatory guidance from the European Medicines Agency (EMA) and the EU HTA Coordination Group. It synthesizes key challenges in non-interventional study design, including selection bias, attrition, and confounding. Advanced statistical techniques such as propensity score adjustment, double robust methods, and marginal structural modelling are discussed, alongside the importance of early planning and interdisciplinary collaboration. ***Results:*** Non-randomised evidence presents a high risk of bias in estimating treatment effects. The EMA and JCA guidelines emphasise the need for rigorous design and analysis to mitigate these biases. Real-world data (RWD) often lacks completeness and consistency, complicating causal inference. Sensitivity analyses, including Quantitative Bias Assessment (QBA), are essential to validate findings. Strategic planning—such as developing RWE protocols and conducting analyses prior to PICO scope finalisation—is critical for successful dossier submission. ***Conclusions:*** While randomised controlled trials remain the gold standard, the increasing use of RWE in regulatory and HTA contexts demands robust statistical approaches and careful planning. Statisticians play a pivotal role in guiding clinical and market access teams to ensure internal validity and scientific rigor. Given the implications for oncology and ATMPs under the EU HTA JCA since January 2025, proper planning and methodologically sound RWE strategies are more important than ever.

MSR203
TRADITIONAL VS. GENERATIVE AI: A RAPID SYSTEMATIC REVIEW ASSESSING ACCURACY AND EFFICIENCY OF AI IN TITLE/ABSTRACT SCREENING

Emily Hardy, MBiol, Amelia Peddle, MSc, Judith Peatman, MSc, Janine Ross, MSc, Shona Lang, PhD
Petauri Evidence, Bicester, United Kingdom

Objectives: The aim of this systematic literature review (SLR) was to assess the accuracy and efficiency of artificial intelligence (AI) tools for SLR title/abstract screening. ***Methods:*** Electronic database searches were conducted in Embase® from inception to March 2025, and supplemented with desktop research. Search results were uploaded to Laser AI for screening and data extraction; two independent reviewers performed screening, with data extraction performed by one reviewer and checked by a second. AI platforms were categorised as commercially available tools (i.e. machine learning and natural language processing models) or generative large language models (LLMs). Prototype or in-house algorithms/models were excluded. Quantitative data for comparative accuracy and efficiency were extracted in addition to a qualitative summary of factors influencing rates. ***Results:*** The SLR identified 51 studies investigating commercially available tools, the most common being Distiller (n=14), Abstrackr (n=9), and ASReview (n=9), plus 45 studies investigating LLMs, primarily OpenAI's GPT models (n=41). Accuracy (tools, n=43; LLMs, n=45) was reported more frequently than efficiency (tools, n=37; LLMs, n=15), however AI use cases plus outcome definitions were highly heterogenous across publications. Time and cost savings were clear, however the impact on accuracy was less consistent. There were notable differences in practical methodology for integrating commercially available tools versus LLMs into SLR workflows (i.e. variable human role), and in factors influencing outcomes of interest. However, consistent factors modulating accuracy and efficiency potential included review methodology, research question complexity, and review size. ***Conclusions:*** Integration of AI technology into SLR workflows offers clear efficiency savings versus conventional methodology, however conclusions regarding comparative accuracy are less clear. Researchers considering the use of AI should identify similar use cases and manage expectations based on potential modulating factors. Clear guidelines on AI methodology across health economics and outcomes research are essential to validate and quantify both accuracy and efficiency.

MSR204
TRANSPARENT, TRACEABLE, AND REPRODUCIBLE EVIDENCE GENERATION FOR HTA USING GENERATIVE AI AND RETRIEVAL-AUGMENTED GENERATION (RAG)

Rajdeep Kaur, PhD, Vedant Soni, B.Tech, Ankita Sood, MPH, PharmD, Mrinal Mayank, B.Tech, Barinder Singh, RPh, Shubhram Pandey, MSc
Pharmacoevidence Pvt. Ltd., Mohali, India

Objectives: Health technology assessment (HTA) bodies such as the National Institute for Health and Care Excellence (NICE) and the Canadian Drug Agency (CDA) acknowledge the growing potential of artificial intelligence (AI) to enhance the decision-making process. This study aimed to explore the application of RAG, a generative AI approach, to enable transparent, traceable, and reproducible evidence