IntuitionLabs

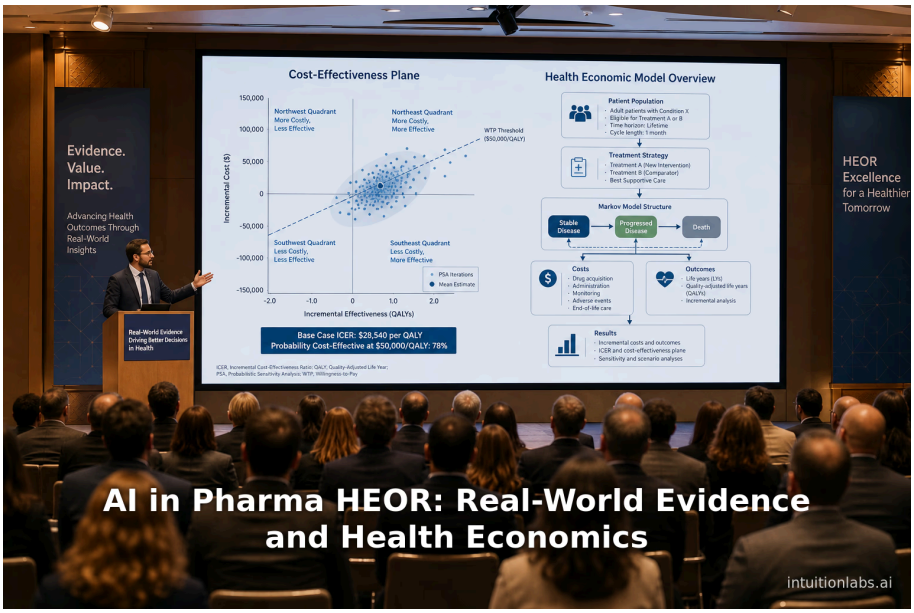
INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

AI in Pharma HEOR: Real-World Evidence and Health Economics

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-04-26 · ai in heor · real-world evidence · real-world data · health economics · outcomes research · machine learning · synthetic data · pharma analytics

Original article: https://intuitionlabs.ai/articles/ai-pharma-heor-real-world-evidence-health-economics



In this article

1. Executive Summary
2. Introduction and Background
3. Real-World Data and Evidence in Pharma HEOR
4. Sources and Types of Real-World Data
5. Regulatory and HTA Use of Real-World Evidence
6. AI and Machine Learning Methods in Healthcare
7. AI in Real-World Evidence Generation
8. Data Extraction and Preparation
9. Descriptive and Diagnostic Analytics
10. Predictive Modeling
11. Causal Inference and Advanced Analytics
12. Data Visualization and Decision Support
13. AI in Health Economics Modeling
14. Automating Model Construction and Adaptation
15. Precision HEOR and Patient Heterogeneity
16. Synthetic Data and Privacy-Enhanced Modeling
17. Data Analysis and Insights
18. Case Studies and Real-World Examples
19. Discussion of Implications and Challenges
20. Future Directions and Implications
21. Conclusion

Executive Summary

The accelerating convergence of real-world evidence (RWE) and artificial intelligence (AI) is transforming pharmaceutical health economics and outcomes research (HEOR). In recent years, **big data** sources – including electronic health records (EHRs), claims databases, patient registries, digital devices, and patient-reported outcomes – have grown exponentially, providing rich real-world data (RWD) that complement traditional clinical trial evidence (doaj.org) (www.tandfonline.com). Concurrently, advances in AI and machine learning (ML) – from traditional predictive models to cutting-edge generative deep learning – enable sophisticated analysis and modeling on these massive datasets. Pharmaceutical companies and health technology assessment (HTA) bodies alike are leveraging AI-powered analytics to extract deeper insights from RWD, build more granular economic models, and ultimately support value assessments, pricing, and market access decisions.

This comprehensive report surveys the state-of-the-art in **AI in Pharma HEOR**, focusing on the symbiosis between AI and RWE. First, we review the **historical context**: the growing role of RWE in regulatory and payer decisions (e.g. the FDA's Real-World Evidence program and the 21st Century Cures Act (www.tandfonline.com), and evolving guidance from NICE, EMA, and ISPOR on RWD use (www.frontiersin.org) (www.nice.org.uk)). Next, we examine the **sources and types of RWD** (EHRs, claims,

registries, wearables, etc.) and how AI methods (from NLP to deep learning) are applied to curate and analyze these data. We contrast **traditional analytics** (descriptive statistics, matching) with **advanced AI-driven analytics** (predictive models, unsupervised learning, causal inference) and illustrate this with tables and use-case examples (www.mckinsey.com).

Crucially, we delve into **health economic modeling** and how AI is augmenting it. We describe classic models (Markov, decision trees, budget impact) and detail multiple avenues where AI intervenes: (i) **Parameter estimation and calibration** from RWD (e.g. deriving transition probabilities from EHR time series); (ii) **Patient stratification** and "Precision HEOR" – identifying subpopulations with different treatment value (pmc.ncbi.nlm.nih.gov); (iii) **Automating model construction and adaptation** using generative AI (e.g. GPT-4 building R code for cost-effectiveness models (pubmed.ncbi.nlm.nih.gov) or adapting Excel models for new countries (www.sciencedirect.com)); and (iv) **Synthetic data generation** – using AI (GANs, diffusion models, LLMs) to produce artificial patient-level data that preserve statistical properties but protect privacy (www.sciencedirect.com). For each aspect, we provide detailed data, case studies, and references: for instance, a study found GPT-4 produced error-free or single-minor-error R scripts in 100% (NSCLC) and 87% (RCC, after a human simplification) of 15 attempts per model (pubmed.ncbi.nlm.nih.gov), and another reported 99.0% and 99.4% accuracy adapting muscle-invasive urothelial carcinoma and myelodysplastic syndrome models to the Czech Republic and the US (www.ispor.org). We also discuss real-world examples where AI-derived RWE had economic implications (e.g. a pre-post quasi-experimental study in *Nature Medicine* reported that an EHR-based AI tool for opioid use disorder was associated with fewer 30-day readmissions (odds ratio 0.53, 95% CI 0.30–0.91) at an incremental cost of ~$6.8K per readmission avoided (pubmed.ncbi.nlm.nih.gov)).

Throughout, we emphasize **evidence-based arguments**: citing systematic reviews, peer-reviewed studies, and industry analyses. We quantify trends (the number of RWD-based cost-effectiveness studies has risen steadily (doaj.org)), and we report on regulatory signals. We also critically examine **challenges**: data quality and bias, the "black box" nature of some AI models, and the need for transparency and validation – noting that regulators (NICE, FDA) are already emphasizing trust and methodology for AI-generated evidence (www.nice.org.uk) (www.ispor.org).

Finally, the report explores **future directions**: the potential of multimodal data integration (genomics, digital sensors, images), continuous learning "digital twins" of disease progression, and AI-enabled decision support for personalized pricing and value-based payment. We conclude that while AI and RWE together hold enormous promise to deepen and accelerate HEOR insights, careful oversight is essential. As one perspective notes, "given the right prompts, complex health economic models can be accurately programmed by LLMs in rapid timeframes," but more research is needed to realize this safely (www.ispor.org) (www.nice.org.uk). The concluding message is clear: embracing AI in pharma HEOR can greatly expand our evidence generation capabilities and refine economic models, but this must be balanced by rigorous validation and ethical guardrails to ensure credibility and patient benefit.

Introduction and Background

Health Economics and Outcomes Research (HEOR) in the pharmaceutical industry traditionally relies on evidence from randomized controlled trials (RCTs) to inform value assessments, pricing, and reimbursement decisions. However, RCTs – while high in internal validity – are often limited in size, duration, and generalizability. In contrast, *real-world evidence* (RWE) derived from *real-world data* (RWD) offers complementary insights on how interventions perform in routine clinical practice. RWD encompasses a broad range of data collected outside of RCTs, including electronic health records (EHRs), medical claims, registries, patient surveys, and even digital device outputs (www.frontiersin.org) (www.tandfonline.com). Since the mid-2000s, regulatory and HTA bodies have increasingly recognized the value of RWE: for example, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) defined RWD to include data "not collected in conventional RCTs," explicitly naming sources like registries and administrative claims (www.frontiersin.org).In 2016, the U.S. 21st Century Cures Act formally encouraged use of RWE in FDA submissions (www.tandfonline.com), and today agencies worldwide are issuing guidance on leveraging RWD (FDA's RWE Framework, NICE's RWE vision, EMA's consultations on RWE). These shifts reflect the reality that "real-world evidence helps us understand how patient characteristics and behaviors affect health outcomes," beyond what RCTs alone can show (www.mckinsey.com).

The explosion of digital data has been transformative. Modern healthcare IT generates vast volumes of patient information – EHRs with diagnoses, labs, notes; insurance claims on therapies and costs; patient-reported outcome measures; and even continuous data from wearables and mobile apps. According to one industry analysis, "21st century has brought about significant technological advancement, allowing the collection of new types of data from the real world on an unprecedented scale" (Lee 2023) (www.tandfonline.com). These "big data" sources (characterized by volume, velocity, variety, veracity) are bewildering in scale but rich in potential signals about treatment effectiveness, safety, adherence, and health resource use.

However, the promise of RWD also brings challenges. RWD is often messy, heterogeneous, and prone to bias – missing data, coding errors, confounding by indication, and privacy concerns. For example, analyses of EHRs must contend with unstructured clinical notes and varying data standards. Pharmacoepidemiologists and outcomes researchers have long developed methods (propensity-score matching, regression adjustment, instrumental variables) to mitigate bias in observational data. Yet as RWD grows, there is a clear opportunity for **artificial intelligence (AI)** and advanced analytics to help.

AI and machine learning bring new tools to this domain. Early efforts used simple predictive models (linear/logistic regression) on select variables; more recently, "advanced RWE analytics" leverages machine learning algorithms (random forests, gradient boosting, neural networks) on **hundreds to thousands of input variables** to extract deeper patterns (www.mckinsey.com). Leading pharma firms now apply AI for patient phenotyping, outcome prediction, and simulation of hypothetical scenarios that were previously infeasible to analyze. As one consultant report notes, these analytics can answer complex questions like "Which patient subsegments respond best to therapy X?" or "What is the risk of an event in 1/3/5 years?" (www.mckinsey.com). Similarly, AI-based tools such as natural language processing (NLP) can mine clinical text to identify diagnoses or outcomes not explicitly captured in structured data fields.

The synergy of RWD and AI is thus reshaping HEOR. By harnessing AI on rich real-world datasets, researchers can generate both **novel evidence** (patient-level insights, personalized predictions) and **enhanced models** (data-driven simulation of disease and costs). This report explores these developments in depth, examining how AI techniques (from supervised and unsupervised ML to generative models) are being brought to bear on RWD for health economics modeling and outcomes research. We provide a comprehensive picture of current practice, supported by case studies and data, and we discuss implications for payers, regulators, and patients.

Real-World Data and Evidence in Pharma HEOR

Sources and Types of Real-World Data

Real-world data (RWD) derives from multiple sources that reflect routine healthcare delivery. As defined by ISPOR and others, the main categories include:

- **Electronic Health Records (EHRs)**: Digital clinical records from hospitals and clinics, containing structured data (diagnosis codes, medication orders, lab results) and unstructured data (physician notes, radiology reports). EHRs enable longitudinal tracking of patients through the continuum of care.
- **Administrative Claims**: Insurance claims databases that capture billings for diagnoses, procedures, and medications. Claims are well-suited for measuring healthcare resource use and costs.
- **Patient Registries**: Organized datasets focusing on specific diseases or treatments (e.g., cancer registries, rare disease registries). Registries often contain clinical details and patient-reported outcomes for particular cohorts.
- **Surveys and Patient-Reported Data**: Population health surveys (e.g. Health and Retirement Study), smartphone app data, and wearable device monitoring (e.g. continuous glucose monitors, activity trackers). These capture real-time patient behaviors and outcomes, sometimes outside clinical settings.
- **Genomic and Biomarker Data**: Increasingly, genomic databases (e.g. disease biobanks) and digital pathology images are linked to health records, enabling precision medicine studies.
- **Social Determinants and External Data**: Non-clinical data such as demographic information, socioeconomic status, or even climate data can be integrated. For instance, one perspective notes an “abundance of patient data from EHR, patient-reported outcomes (PROs), laboratory, demographic, social media, digital, and even climate data” in modern healthcare (www.tandfonline.com).

While each source has strengths, no single RWD source is perfect. For example, EHRs give rich clinical context but may lack complete cost information; claims capture utilization and costs but have limited clinical detail. Thus, integrating multiple RWD sources (sometimes termed “big data” when many are combined) can provide a more comprehensive evidence base. A systematic review found that the use of RWD in cost-effectiveness studies has been increasing dramatically: from 2011 onward, health economists have

increasingly sourced both effectiveness and cost inputs from RWD (www.frontiersin.org) (doaj.org). However, this review also notes that truly large-scale “big data” studies – combining multiple data streams – remain relatively rare, highlighting an area for future growth (doaj.org) (www.frontiersin.org).

Importantly, RWD offers sample sizes and diversity that far exceed typical clinical trials. For conditions studied in EHR networks or national databases, researchers may have cohorts of 10,000+ patients, enabling statistically powerful analyses of subgroups. By contrast, RCTs often enroll only hundreds. This means RWD can uncover rare adverse events or heterogeneous treatment effects that would be missed in trials. As one review observes, RWD sources “feature a larger sample size” and can offer long-term follow-up beyond typical trial durations (www.frontiersin.org). These attributes make RWD invaluable for HEOR tasks like budget impact modeling or long-term cost-effectiveness modeling.

Regulatory and HTA Use of Real-World Evidence

The growing prominence of RWE is driven by stakeholder demand. Payers and HTA agencies want evidence of **real-world benefit and value** of therapies, not just idealized trial efficacy. Regulators are increasingly receptive: for example, the 2016 U.S. 21st Century Cures Act explicitly encouraged use of RWE in submissions for drug and device approvals (www.tandfonline.com). The FDA’s Real-World Evidence Program has issued guidance on when RWD can support regulatory decisions (e.g. assessing whether EHR and medical claims data are fit for regulatory use (www.fda.gov)). In Europe, EMA and HTA bodies have conducted RWE pilots and consultations, recognizing that RWE can inform both pre-approval (e.g. expanded indications) and post-approval (e.g. safety, comparative effectiveness) assessments (www.ema.europa.eu).

Professional societies have codified RWE best practices. The ISPOR-ISPE Task Force (2017) outlined how to conduct rigorous observational studies and adjust for confounding (www.frontiersin.org). For reimbursement, bodies like NICE and CADTH have signaled interest in RWD for decision making. For instance, NICE’s five-year strategy explicitly states a commitment to “making greater use of real-world evidence” to inform guidance. NICE has also recognized AI’s role in evidence generation: in its August 2024 position statement (since updated and replaced by its AI methods for evidence framework), NICE noted that “it’s highly likely that, in the near future, evidence considered by NICE will be informed by AI methods,” while emphasizing the need for transparency and trustworthiness (www.nice.org.uk).

These developments mean that pharmaceutical companies are investing in RWE generation. Registrational trials are now often complemented by post-marketing studies using RWD. For example, AstraZeneca conducted large RWD studies (CVD-REAL) to demonstrate real-world benefits of its diabetes drug (www.mckinsey.com). Each such case builds pharma’s familiarity with RWE.

Nevertheless, data gaps and methodological uncertainty can still limit how far payers and agencies rely on RWE. AI and related digital tools are seen as a key way to overcome these gaps – for example by synthesizing sparse data or extracting information from unstructured fields. As we detail below, the combination of RWD and AI is thus driven by both demand (rising need for real-world value evidence) and capability (advances in computational methods).

AI and Machine Learning Methods in Healthcare

AI encompasses a broad set of computational methods that can learn from data. In the context of RWE and HEOR, the most relevant AI/ML methods include:

- **Supervised Machine Learning**: Algorithms trained on labeled data to predict an outcome. Examples include linear/logistic regression, decision trees, random forests, gradient-boosting machines (XGBoost, LightGBM), support vector machines, or neural networks. In HEOR, supervised models predict outcomes like hospitalization, mortality, or quality-of-life given patient features.
- **Unsupervised Learning**: Algorithms that uncover structure in unlabeled data. This includes clustering (k-means, hierarchical clustering), dimensionality reduction (PCA, t-SNE), and association rule mining. Unsupervised methods can identify patient subgroups (phenotypes) or latent variables without a predefined target. For example, cluster analysis can reveal novel comorbidity profiles tied to cost trajectories.
- **Deep Learning**: Specialized neural network models that can handle complex, high-dimensional data. Convolutional neural networks (CNNs) process imaging or time-series data; recurrent neural networks (RNNs) and transformers handle sequential data (text, time series); graph neural networks can capture relational data in networks. In RWE, deep learning is used to interpret free-text notes (via NLP), medical images (radiology, pathology), and continuous monitoring signals (wearables).
- **Natural Language Processing (NLP)**: A branch of AI focused on extracting information from text. Clinical NLP can identify diagnoses, medications, and outcomes from physician notes. Transformer-based language models (e.g. BERT) have achieved near-human performance on many tasks. NLP is crucial for turning unstructured RWD into analyzable variables (e.g. converting clinician comments into structured endpoints).
- **Causal Inference and Counterfactual Models**: Although not traditionally labeled “AI”, modern approaches like Bayesian networks, causal forests, and potential outcome modeling are used to estimate causal effects from observational data. Techniques like target trial emulation or inverse probability weighting allow “what-if” analyses on RWD. The use of these methods – sometimes termed **prescriptive analytics** – is growing as regulators and payers ask for effect estimates that mimic randomized trials (www.tandfonline.com).
- **Generative Models**: Cutting-edge AI models that can create new data. These include Generative Adversarial Networks (GANs) and Variational Autoencoders for synthetic data generation, and large language models (LLMs, e.g. GPT-4) that can generate human-like text or code. Generative models are increasingly applied to simulate synthetic patient cohorts, generate missing data, or even draft analytical reports.
- **Reinforcement Learning**: Algorithms that learn decision policies through trial and error. In health, RL can be used to optimize treatment strategies or pricing models in dynamic environments. This is still relatively experimental in HEOR, but may play a future role (e.g., optimizing sequential drug

administration strategies).

Each AI/ML approach has its place in the RWE/HEOR pipeline. For instance, **XGBoost or random forest** models easily fit tabular RWD to predict costs or outcomes; **neural networks** excel with imaging data or complex interactions; **NLP transformers** can parse EHR text; and **GANs/LLMs** can generate synthetic data or code for modeling. Frequently, hybrid pipelines combine methods: for example, an LLM prompt might generate code that runs a random forest on claims data.

It is important to note the interplay between AI and traditional biostatistics. Classical methods like Cox regression or mixed-effects models remain valuable and interpretable, and AI is often used alongside them. In regulatory submissions, a hybrid approach is common: sophisticated AI models may be used for hypothesis generation or subgroup discovery, while established biostatistical models handle primary analyses, with double-validation to ensure results hold. In other words, advanced AI “builds the engine,” but domain experts oversee and adjust the modeling process.

AI in Real-World Evidence Generation

AI's impact on RWE comes through various tasks in the **evidence-generation pipeline**. We discuss key areas where AI is enhancing RWE:

Data Extraction and Preparation

RWD often requires extensive curation before it can be analyzed. Clinical data can be **noisy and heterogeneous**; for example, the same condition might be coded differently in different hospitals, or entered only in free-text notes. AI methods are transforming this stage via:

- **Natural Language Processing (NLP)**: Clinical NLP tools can scan physician notes to identify mentions of symptoms, diagnoses, or adverse events. New transformer-based models (e.g. ClinicalBERT) achieve high accuracy on medical text. For example, an AI screener deployed in a hospital used a convolutional neural network to read EHR notes and flag patients at risk of opioid use disorder (scholars.luc.edu). In the study's pre-post comparison, addiction medicine consult rates did not change between periods (1.35% versus 1.51%, non-inferior). Similarly, NLP can extract quality-of-life scores, side effects, or progression events from narrative sources that structured data miss.
- **Entity Matching and Deduplication**: Identifying the same patient or event across datasets (EHR vs claims) is non-trivial. AI-based record linkage (using probabilistic matching or ML classifiers) improves data integration. Entity resolution algorithms help merge datasets from different hospitals or integrating registry and claims data.
- **Data Cleaning with AI**: AI can also flag or impute missing values. For instance, deep autoencoder models can suggest plausible values when certain lab results are absent, based on similarity to other patients. While still experimental, such approaches can reduce bias from missing data.

Together, these AI tools vastly increase the usable portion of raw RWD. A meta-perspective notes that "conventional statistical methods still play a significant role, but ML and AI are assuming a more prominent role in analysis of this 'big data'" (www.tandfonline.com). In practice, much of routine data wrangling (previously done manually) can now be automated or semi-automated with NLP and ML, freeing analysts to focus on higher-level questions.

Descriptive and Diagnostic Analytics

Before applying advanced models, analysts typically perform **descriptive analytics** to understand the population: e.g., summarizing demographics, baseline risks, utilization patterns. Traditional tools (tables, Kaplan-Meier curves, simple regression) still have a place here. AI tools can augment these steps by automatically clustering patients into subgroups or detecting anomalies. For example:

- **Unsupervised Clustering** can reveal subpopulations not pre-specified by researchers. Clusters might correspond to distinct treatment pathways or comorbidity profiles. In one conceptual framework, Chen et al. show how random forests (an ML method) can be used in a "precision HEOR" analysis to uncover patient cohorts with differing outcomes (pmc.ncbi.nlm.nih.gov). This is a form of data-driven patient segmentation, which feeds into heterogeneous treatment effect analysis (discussed below).
- **Pattern Mining**: AI can automatically identify frequent co-occurring conditions, pharmacy patterns, or sequences of care in large RWD sets (e.g. sequential pattern mining), which would be impractical by hand in a dataset of millions of records.
- **Visualization with AI**: Techniques like t-SNE or UMAP (nonlinear dimensionality reduction) allow projecting high-dimensional clinical data into 2D for visualization, which can reveal structure. Interactive AI-driven dashboards (embedding ML algorithms) can help analysts explore data faster than static plots.

While these tasks may not directly produce payers' evidence, they set the stage for hypothesis generation. For example, identifying a high-risk cluster may suggest studying the cost-effectiveness of a therapy specifically in that group.

Predictive Modeling

At the core of AI-enhanced RWD analysis are **predictive models**. These use patient features (demographics, comorbidities, biomarkers, treatment history) to forecast outcomes. Common applications in HEOR include:

- **Prognostic Models**: Predicting patient outcomes under current standard of care (e.g. survival, disease progression, hospitalization). These models can be used to simulate disease natural history or to risk-adjust when comparing interventions. Deep learning models (e.g. time-to-event neural networks) have shown increased accuracy in survival prediction over traditional Cox models in some settings.
- **Event Prediction for Economic Modeling**: For cost-effectiveness models, one often needs probabilities of transitioning between health states. For example, one could train an ML model on RWD to predict monthly probability of heart failure hospitalization based on current clinical covariates, replacing or calibrating parameters in a Markov model. Indeed, a case study by Thokala et al. (2020)

used hospitalization incidence from UK administrative data to define Markov states for heart failure (eprints.whiterose.ac.uk). This approach leverages RWD to generate model parameters that reflect local, real-world risks, rather than relying solely on trial data.

- **Cost and Resource Forecasting**: AI can predict individual or population-level costs. For instance, mixture models or neural networks can forecast each patient's expected healthcare cost under different scenarios. Such models help estimate budget impact or to perform budget sensitivity analyses. (While not extensively reported in the literature yet, machine learning cost-prediction models are emerging in the broader health services research field (www.mckinsey.com).)
- **Adverse Event and Safety Signals**: Machine learning on RWD can monitor safety. By learning from historical RWD, models can predict which patients are at risk of serious side effects if given a certain drug. Such predictions can feed into health economic models as well (factoring in expected costs of complications).

An instructive example comes from AstraZeneca's CVD-REAL study mentioned earlier: using multinational registry and claims data with advanced analytics, they demonstrated that a class of diabetes drugs significantly reduced heart failure and death vs comparators (www.mckinsey.com). This analysis used propensity scoring and regression, but future work could incorporate more complex ML for risk stratification. Similarly, an analysis by Afshar et al. (2025) embedded an AI (CNN) within the EHR to predict opioid use disorder risk and identify patients needing intervention (scholars.luc.edu). In a pre-post comparison, the screener was associated with fewer 30-day readmissions, at an incremental cost per readmission avoided. These show how predictive analytics on RWD can inform both clinical and economic insight.

Finally, advanced AI provides **scenario simulation** capabilities far beyond classical modelling. Once predictive relationships are learned, one can pose "what-if" questions: for example, machine models can simulate population outcomes under hypothetical changes. In text, an analyst could query: "If all patients on drug A switched to drug B, how many events would occur?" and ML models could estimate that. (www.mckinsey.com). This is akin to digital twin simulation: generating counterfactual trajectories for patients. While formal causal inference is needed for rigorous answers, AI-based counterfactual simulation is already aiding HEOR exploration.

Causal Inference and Advanced Analytics

A key goal in HEOR is causal understanding (e.g. "does treatment X improve outcomes compared to Y in the real world?"). Traditional RWD analyses use methods like propensity scores or instrumental variables to adjust for confounding and estimate causal effects. AI offers new tools here too:

- **Causal Forests and Heterogeneous Effect Models**: Machine learning extensions of random forests (e.g. causal forests) can estimate treatment effects for subpopulations. The Precision HEOR review highlights decision-tree models to find subgroups with different cost-effectiveness (pmc.ncbi.nlm.nih.gov). In the same spirit, causal forest algorithms can "learn" which patient features modify the treatment effect, informing personalized value assessments.

- **Counterfactual Prediction**: Deep learning can predict potential outcomes under different interventions. For instance, one could train an RNN that, given a patient's history and treatment, outputs predicted quality-adjusted life years (QALYs). By simulating with different treatment flags (A vs B), one approximates an individualized cost-effectiveness, albeit without randomized assignment. These approaches are still in research but are gaining traction as computational power grows.
- **Causal Discovery and Graphical Models**: Bayesian networks or structural causal models can help identify hidden confounders in big data. For example, algorithms might detect that both prescribing of Drug X and patient age influence hospitalization risk, revealing age as a confounder. Embedding such causal reasoning into ML pipelines makes RWD analyses more robust.

Experts predict that **counterfactual prescriptive analytics** will become mainstream. As Lee (2023) notes, "counterfactual prescriptive analytics, such as the causal inference model utilizing RWD… will be gaining momentum as a methodology that can stand up against the rigor of regulatory review" (www.tandfonline.com). In other words, using real-world data to draw causal conclusions is a key frontier.

Comparing the questions answered by analytics highlights the progression:

Analysis Type	Typical Questions	Methods & Examples
Descriptive/Diagnostic	What happened? Why did it happen?	Descriptive stats, matching, logistic regression
Predictive Analytics	What will happen (forecast)?	Machine learning (e.g. random forest, neural nets) to predict outcomes from patient features (www.mckinsey.com) (www.tandfonline.com)
Counterfactual/Prescriptive	What if an intervention or policy were applied?	Causal inference (target trial emulation, instrumental inference); "what-if" simulation using ML models (www.tandfonline.com)

Table 1. Comparison of traditional RWE analytics versus advanced AI-driven analytics (sources: McKinsey (www.mckinsey.com), Lee 2023 (www.tandfonline.com)).

As shown in Table 1, conventional RWE analytics (top rows) typically address descriptive questions about patient populations and simple comparisons (often using propensity-score matching or small-variable analyses (www.mckinsey.com)). In contrast, **AI-driven RWE analytics** (bottom rows) leverage large, rich datasets and complex models to predict future outcomes and to simulate "what-if" scenarios (www.mckinsey.com) (www.tandfonline.com). This transition reflects the industry's move "from table-stakes to high-stakes" analytics, where thousands of patient variables and advanced algorithms unlock deeper insights (www.mckinsey.com).

Data Visualization and Decision Support

An often overlooked impact of AI is in presentation of evidence. Modern analytic platforms offer interactive dashboards where AI methods power adaptive visualizations. For example, an "AI-assisted cohort explorer" might allow a user to click on subgroups and see predicted outcomes based on underlying models. Embedded question-answering (e.g., ChatGPT for queries on the data) is not yet common but could emerge.

Health economists are also beginning to use AI for literature search and evidence synthesis. Large language models can rapidly summarize relevant study results, identify comparative effectiveness studies, or even suggest keywords for meta-analyses. While full automation is premature, these tools can speed background research (e.g., scanning thousands of abstracts to find relevant cost studies).

In summary, AI is permeating the RWE generation process at multiple layers: from raw data extraction and cleaning, to multilevel analytics (descriptive, predictive, causal), to downstream synthesis and visualization. This allows pharmaceutical HEOR teams to ask and answer questions far beyond what traditional methods could handle. We next turn to the specific domain of health economics modeling: how these AI-driven RWE approaches inform the construction and adaptation of pharmacoeconomic models.

AI in Health Economics Modeling

Health economic models translate clinical outcomes into long-term costs, benefits, and value metrics (e.g. cost-effectiveness, budget impact). Conventional approaches include **decision trees**, **Markov models**, and **microsimulation models**, built using software like Excel, TreeAge, or specialized R packages. These models rely on inputs from clinical trials (efficacy, utilities) and cost data, often requiring calibration to ensure face validity. Historically, model development has been labor-intensive: defining health states, coding transition probabilities, running simulations, and writing technical reports.

AI is rapidly changing several aspects of this modeling process:

Automating Model Construction and Adaptation

One of the most exciting developments is the use of generative AI (large language models) to *program* economic models. Traditionally, converting a model concept into code (R, Python, or Excel formulas) is done manually by health economists. Recent studies show that LLMs can greatly accelerate this:

- **GPT-4 generates model code**: Reason et al. (2024) guided GPT-4 with textual prompts describing two published oncology cost-effectiveness models (non-small-cell lung cancer and renal cell carcinoma) and asked it to write R scripts. Remarkably, GPT-4 produced scripts that were error-free or contained a single minor error in 100% of lung cancer attempts and 87% of renal cell attempts (the latter after a human simplified one input calculation) (pubmed.ncbi.nlm.nih.gov). For the lung cancer model, 93% of GPT-4 runs were completely error-free; for the renal cell model, 87% were correct or nearly correct after one simplification (pubmed.ncbi.nlm.nih.gov). The error-free scripts reproduced published incremental cost-effectiveness ratios within 1% (pubmed.ncbi.nlm.nih.gov). The authors conclude that “GPT-4 can have practical applications in the automation of health economic model construction… [offering] accelerated model development timelines and reduced costs” (pubmed.ncbi.nlm.nih.gov).
- **LLM-based adaptation toolchains**: Rawlinson et al. (2025) extended this concept to **model adaptation**. Many HTA agencies expect a “global” model to be adapted to local settings (different country costs, epidemiology). Rawlinson’s team built an LLM-based pipeline that automatically exchanged country-specific parameter values in Excel and Word documents representing a Markov model and its report (www.sciencedirect.com). The result was astonishing: parameter updates were 98–

100% accurate across scenarios, completed in roughly 3–4 minutes (195, 245 and 189 seconds), and at minimal cloud service cost (~$2–13) (www.sciencedirect.com). Report text was similarly adapted with >94% sentence-level accuracy (www.sciencedirect.com). In practice, this means routine economics model adjustments (to meet local HTA requirements) can be done in minutes instead of days by AI. Even more compelling, a related ISPOR Europe 2024 conference abstract tested a GPT-4-powered version of the approach (called LLMAdapt) across two different disease models (muscle-invasive urothelial carcinoma and myelodysplastic syndrome) and two countries (the Czech Republic and the US), finding ~99% accuracy (www.ispor.org).

These findings are leading experts to envision a future where **model programming is largely automated**. As Rawlinson (2024) discusses in a Value & Outcomes webinar, “LLMs are expected to have a large impact in health economic modeling” by streamlining repetitive tasks (www.ispor.org). In practice, an analyst might draft a model specification (in natural language or structured text), then let an AI tool generate the code. Health economists would then focus on reviewing and validating the model rather than hand-coding every formula. This could dramatically reduce human error – a 2020 study found that virtually all human-built health economic models contain technical errors – and allow exploration of multiple model structures (which is currently limited by effort) (www.ispor.org). As one expert summary puts it, “given the right prompts, complex health economic models can be accurately programmed by LLMs in rapid timeframes” (www.ispor.org).

Table 2 below summarizes key AI-driven uses in HEOR modeling:

Application	AI/ML Technique	Examples & References
Model Programming	Large Language Models (GPT-4, etc.)	GPT-4 auto-codes partitioned survival Markov models in R; in one study, 100% (NSCLC) and 87% (RCC) of generated models were error-free or had a single minor error; reduces dev time.
Model Adaptation/Localization	Chain-of-thought prompting, LLM pipelines	LLM pipeline updates Excel CEA models for new countries; achieves ~99% parameter accuracy in <5 min (www.sciencedirect.com).
Causal Effect Simulation	Counterfactual ML models (causal forests)	Frameworks using causal tree/RF to identify subgroups for cost-effectiveness (pmc.ncbi.nlm.nih.gov).
Parameter Estimation from RWD	Supervised ML (survival models, regression)	ML on claims/EHR predicts transition probabilities or risk equations for Markov models (e.g. heart failure PF from RWD (eprints.whiterose.ac.uk)).
Synthetic Data Generation	GANs, Variational Autoencoders, LLMs	Synthetic patient data generation for rare diseases or privacy enhancement (www.sciencedirect.com).
Quality Control & QA	LLM comprehension (review, summarization)	LLM audits model logic and report text; aids peer review by flagging inconsistencies (emerging practice).

Table 2. Selected applications of AI/ML in pharmaceutical HEOR modeling (examples from literature).

The listed uses only scratch the surface. For example, there is early work on **reinforcement learning** to optimize pricing policies in simulated markets, though applications are still nascent. Similarly, generative models can craft synthetic populations (see Synthetic Data below) to stress-test models under extreme scenarios.

There are also AI tools for **reporting and documentation**: GPT-like models can draft technical report language, or perform automated rather than manual literature reviews to justify model inputs. For instance, LLMs might summarize relevant effectiveness studies as evidence for model parameters. This could accelerate the tedious task of writing value dossiers, though it requires careful oversight to avoid "hallucinations."

Precision HEOR and Patient Heterogeneity

Traditional economic models often assume average treatment effects across a population. However, patient responses vary widely. AI enables a more granular approach, sometimes called **"Precision HEOR" (P-HEOR)** (pmc.ncbi.nlm.nih.gov). The idea is to find distinct patient **phenotypes** (patterns of risk and benefit) and assess value separately for each. For example, the same cancer drug might be highly cost-effective in a younger, less comorbid subgroup but not in older patients.

Machine learning excels at uncovering such heterogeneity. Chen et al. describe a conceptual framework where random forests are applied to RWD to stratify patients into subgroups that could then be assessed for differential cost-effectiveness (pmc.ncbi.nlm.nih.gov). In their illustrative example, they identified ICU sepsis survivors at higher risk of a prolonged ICU stay; the example did not compare treatments or assess cost-effectiveness. By framing HEOR questions at the individual or subgroup level, ML-driven precision HEOR can yield personalized value estimates. In effect, AI turns a one-size-fits-all cost-effectiveness number into a distribution of values across patient segments.

Notably, precision HEOR can also inform **risk-sharing agreements** and targeted pricing. If payers know that a drug is cost-effective only for a certain genomic profile or biomarker-defined group, manufacturers might propose higher prices for that subset and discounts elsewhere. AI-powered subgroup analysis thus has real commercial implications.

However, precision HEOR must be done carefully. Machine models may identify spurious subgroups if not controlled for multiple testing or data noise. The proposed framework incorporates cross-validation and domain knowledge to ensure subgroups are clinically meaningful (pmc.ncbi.nlm.nih.gov). Despite these challenges, the ability of AI to parse heterogeneity is widely seen as a major advantage: "In the era of big data, P-HEOR can benefit from ML optimization to identify patient cohorts with different risk-benefit profiles in terms of both clinical and economic outcomes," the authors conclude (pmc.ncbi.nlm.nih.gov).

Synthetic Data and Privacy-Enhanced Modeling

Privacy regulations (HIPAA in the U.S., GDPR in Europe) often restrict access to patient-level RWD. An emerging AI-driven solution is **synthetic data**: artificially generated datasets that mimic the statistical properties of real data without containing actual patient records. AI-based synthetic data can alleviate privacy concerns and enable analysis even when raw data cannot be shared (www.sciencedirect.com).

For example, generative adversarial networks (GANs) or variational autoencoders can be trained on a hospital's EHR to produce synthetic patients (with plausible combinations of demographics, labs, diagnoses) that capture the correlations of the source population. Large language models can similarly output pseudo-random data (or entirely simulated transcripts of clinical visits) based on patterns learned from real records.

The advantages of synthetic data in HEOR include:

- **Data Augmentation**: Synthetic cohorts can enlarge sample sizes, especially for rare subgroups (e.g. rare diseases) where real data are sparse. This improves model stability and allows exploration of "what-if" scenarios (e.g. more patients with a rare genotype). As Lai & Ngorsuraches (2025) note, synthetic data "can strengthen the robustness of findings for underrepresented populations" (www.sciencedirect.com).
- **Privacy Preservation**: Because no real patient identifiers are included, synthetic datasets can be more freely shared between researchers or across institutions. This facilitates collaborative HEOR research and allows verification of models by external parties. For instance, an AI modeler could download a synthetic cancer registry (generated by the pharma company) and validate cost-effectiveness simulations, without compromising patient privacy.
- **Testing Model Generalizability**: Synthetic data can be used to stress-test economic models. For example, one could simulate a wide range of patient characteristics to see how the model outcomes vary. This is useful in sensitivity analyses and in demonstrating model robustness to payers.

However, synthetic data also have limitations. Critics point out that synthetically generated samples, if not carefully validated, might introduce artefacts or underrepresent real-world variance (www.sciencedirect.com) (www.nice.org.uk). Hallucination or mode collapse in generative AI models can cause synthetic data to deviate from reality. Therefore, an **evaluation framework** for synthetic data quality is needed. As Lai & Ngorsuraches (2025) recommend, future research should focus on rigorously assessing how well synthetic data preserve key statistical properties and outcomes (www.sciencedirect.com).

In summary, synthetic data is a promising development at the intersection of AI and RWE. It "provide a unique opportunity to overcome data-related challenges in HEOR," including privacy and scarcity (www.sciencedirect.com). Pharma and payers may begin to experiment with such approaches. We expect this trend to grow, enabling more extensive AI-based analyses while respecting patient confidentiality.

Data Analysis and Insights

AI's role is not limited to modeling – it also enables richer insights from RWD. As noted, *advanced RWE analytics* can "predict outcomes for a new patient with a unique set of characteristics" by learning relationships among thousands of variables (www.mckinsey.com). We summarize some evidence-based advances in AI-driven analysis:

- **Subgroup Identification**: Beyond broad clusters, AI can identify which features (comorbidities, lab values, social factors) most strongly drive differential outcomes. Feature importance techniques (e.g. SHAP values for tree models) make ML more interpretable. For example, a model might reveal that an

elevated biomarker significantly increases a drug's net benefit, suggesting targeted therapy. These insights can refine health economic parameters by making them conditional on key covariates, rather than global averages.

- **Uncertainty Quantification**: Bayesian ML models can provide probabilistic estimates (credibility intervals) for predictions. This is valuable in HEOR where sensitivity analysis is paramount. For instance, Bayesian neural nets or Gaussian processes can quantify uncertainty around cost predictions or QALY gains, complementing traditional one-way or probabilistic sensitivity analyses (www.frontiersin.org). Likewise, ensemble methods (running many ML models) naturally generate confidence intervals.
- **Handling Unstructured Data**: Aside from NLP, deep learning can process images (e.g. pathology slides, retinal scans) to derive predictive biomarkers. In oncology, molecular imaging features extracted by CNNs can feed into survival models, giving a head start to economic evaluations based on imaging-guided stratification. Similarly, data from wearable sensors (continuous glucose monitors, ECG monitors) can be analyzed by recurrent neural networks to detect patterns that predict future costs or events.
- **Federated Analytics**: To address multi-site data integration, federated learning allows training ML models across distributed databases without pooling patient-level data. The Haug et al. JAMIA study (2024) is not itself an example of federated learning or AI; it used a common data model (OMOP) and shared R packages to run Markov cost-effectiveness analyses across sites' data (colab.ws). More generally, federated learning enables *distributed analytics*: models can be trained locally and aggregated, preserving privacy while leveraging multi-country sample sizes. The study covered 47,163 patients across five databases.
- **Interactive Exploration**: AI-powered interfaces (dashboards, notebooks) let analysts interact with RWD. Some emerging platforms use AI to auto-suggest queries or flag anomalies. For instance, an AI might automatically detect that a spike in a certain ICD code is correlated with a change in pricing policy. Although still cutting-edge, interactive AI tools are likely to become standard in pharm HEOR teams.

We illustrate multi-country HEOR tooling with an example that relies on standardized analytics rather than AI. In a multisite heart failure study, researchers used two OMOP-based R packages to analyze treatment trajectories across five databases from four countries (colab.ws). Country-specific Markov models yielded ICERs for telemonitoring relative to standard of care ranging from about €40,372 per QALY in Serbia to about €90,893 per QALY in the US (colab.ws). The study reports no machine-learning component; it shows how standardized, reusable tooling can support comparable cost-effectiveness analyses across sites, a foundation that AI methods could later build on.

Another example is an AI-based utility estimation. Normally, health utilities (health-related quality of life weights) are derived from surveys. But AI can infer utilities from large RWD. For instance, advanced ML on patient-reported outcomes plus clinical data can model the mapping from patient status to utility. This is especially useful in rare diseases where utility data is scarce. By learning from related common diseases, a neural net could predict utilities for a new condition. Such applications are emerging in literature.

Overall, AI transforms raw RWD into deeper evidence by uncovering patterns, enabling forecasting, and handling complex data types. These enriched insights feed into HEOR analyses at every step: defining model structure, setting inputs, and projecting outcomes.

Case Studies and Real-World Examples

To ground the above discussion, we highlight several illustrative case studies where AI and RWE have been combined in HEOR contexts:

Fully Automated Model Programming (Reason et al. 2024)

Tim Reason and colleagues at Bristol-Myers Squibb and Estima Scientific (2024) directly tested GPT-4's ability to program health economic models (pubmed.ncbi.nlm.nih.gov). They selected two published partitioned-survival cost-effectiveness models (non-small-cell lung cancer and renal cell carcinoma) originally in Excel, and attempted to reproduce them in R using only text descriptions as prompts. GPT-4, primed via a structured chain-of-thought prompt pipeline, generated R code for each model 15 times. The results were striking:

- The lung cancer model was **fully replicated**: 100% of GPT-4's versions contained no more than minor issues, and 93% were completely error-free.
- The renal cancer model required a minor human simplification (splitting a complex formula) but otherwise GPT-4 again delivered 87% error-free or minor-error scripts (60% perfect).
- Importantly, every error-free script computed outcomes (QALYs, costs, ICER) within 1% of the published results (pubmed.ncbi.nlm.nih.gov).

These findings demonstrate that a state-of-the-art LLM can *automatically construct complex economic models with high fidelity*. The authors note that potential benefits include "accelerated model development timelines and reduced costs of development" (pubmed.ncbi.nlm.nih.gov). The study concluded on an optimistic note: "GPT-4 can have practical applications in the automation of health economic model construction. … Further research is necessary to explore the generalisability of LLM-based automation across a larger sample of models" (pubmed.ncbi.nlm.nih.gov).

This case exemplifies the potential for **automated double-programming** in HEOR. Instead of two analysts independently coding a model (to cross-check each other's work), one analyst could prompt an LLM and then verify its output, saving significant time. Furthermore, multiple model variants (e.g. one-way sensitivity analyses with alternate inputs) could be generated by second-by-second prompt modifications, a level of speed unattainable manually.

LLM-Based Model Adaptation (Rawlinson et al. 2025)

Another pioneering study by Rawlinson et al. (2025) used generative AI to **adapt existing economic models to new settings** (www.sciencedirect.com). They identified that a common HEOR task is taking a "global" model (e.g. developed for the U.K.) and adjusting it to another country's context (different costs, epidemiology, target population). This typically involves hours of manual updates to model spreadsheets and report text.

Rawlinson's team developed an LLM pipeline that used Anthropic's Claude 3.5 Sonnet (with GPT-4o and o1-preview also used in the report-adaptation steps), where the LLM was given country-specific data along with the original Excel model. Using chain-of-thought and task decomposition prompts, the pipeline automatically updated the model's parameter values and then adapted the technical report text to reflect the adapted model results. The performance was remarkable:

- **Parameter updates:** The pipeline achieved **100% accuracy** on two cost-effectiveness models and 98.7% on a budget impact model (missing only 2 of 160 values) (www.sciencedirect.com). All three adaptations took only 2–4 minutes of compute time (Linux environment) with minimal cost.
- **Report edits:** The LLM also adapted the narrative report sections with 94–100% sentence-level accuracy. These edits took roughly 2–6 minutes per model (128, 336 and 286 seconds) (www.sciencedirect.com).
- **Cost and speed:** As reported above, cloud costs were minimal (about $2–13) and each adaptation took only a few minutes.

The authors conclude that "LLM-based toolchains have the potential to accurately and rapidly perform routine adaptations of Excel-based CEMs and technical reports at a low cost. This could expedite health technology assessments and improve patient access to new treatments." (www.sciencedirect.com). In an industry context, this implies that a multinational company could roll out adapted models to dozens of countries within hours rather than months.

They also tested the generalizability of this approach in an ISPOR Europe 2024 abstract (www.ispor.org). The GPT-4-powered LLMAdapt pipeline, which provided GPT-4 with tabular data formatted to mimic the output of a targeted literature review, was applied to two different disease model templates (urothelial carcinoma and myelodysplastic syndrome) adapted to the Czech Republic and the U.S. In each case, GPT-4 automatically performed ~99% of required updates correctly (www.ispor.org). This confirmed that the method worked across disease states and healthcare settings.

Together, these studies by Rawlinson et al. show that **generative AI is already automating many formerly manual HEOR tasks**. Model adaptation – once a bottleneck in global submissions – can be done in minutes. The chief remaining tasks are prompt engineering and oversight. The authors highlight that "large language models have huge potential for automating tasks that are currently performed manually in health economics and outcomes research, which could greatly expedite the health technology assessment process and improve timely patient access" (www.sciencedirect.com).

Precision HEOR Framework (Chen et al. 2020)

The concept of *Precision HEOR* (P-HEOR) predates the Chen et al. (2020) conceptual review, which builds on it to discuss how ML could support personalized economic evaluation (pmc.ncbi.nlm.nih.gov). As a conceptual illustration, the authors used real-world clinical data on 1,474 sepsis survivors to build a random forest model that stratified patients by their risk of a prolonged ICU stay (length of stay ≥6 days). They explicitly did not compare treatments or conduct a formal cost-effectiveness analysis, but outlined how such an analysis could be conducted within risk subgroups. The review's key argument is that treating patient heterogeneity as noise (averaging it out) is suboptimal: instead, ML can help **identify distinct patient subgroups**, enabling subgroup-specific economic assessments.

While this was a conceptual demonstration, it has practical parallels. For instance, genomic subtypes of cancer often experience different benefit from therapy; an ML model can partition patients by molecular features and then compute separate ICERs. Similarly, comorbidity patterns (identified by clustering) can define high-risk vs low-risk groups with differing NMB (net monetary benefit) of an intervention (pmc.ncbi.nlm.nih.gov). A related NIH-funded effort, the Personalized Risk Information in Cost-Effectiveness Studies (PRICES) project, aimed to use risk models to understand how each patient's individualized risk of adverse health outcomes can help clinicians better tailor care and resources (pmc.ncbi.nlm.nih.gov).

Key takeaway: **AI enables micro-level cost-effectiveness analysis**. Instead of one ICER per population, we can have a distribution of ICERs. This better informs payers who may consider patient stratification in reimbursement.

Federated Real-World Modeling (Haug et al. 2024)

Haug et al. (2024) provide a multi-country HEOR example powered by federated data (colab.ws). They implemented two R packages specifically designed for OMOP Common Data Model networks, which streamline the Markov modeling process using distributed data. In their demonstration, they replicated a UK-based heart failure data analysis across 5 different databases from four countries (Estonia, Spain, Serbia, and the United States), using administrative claims mapped to OMOP. This is standardized analytics on an OMOP data network rather than AI, and the abstract reports no machine-learning component, but the approach exemplifies modern RWD analysis:

- The study examined treatment trajectories of **47,163 patients** in total; the second R package handles Markov model learning and profile synthesis to estimate country-specific cost-effectiveness.
- The resulting cost-effectiveness results varied widely by country: telemonitoring yielded an ICER of €57,500/QALY overall, but ranged from €40,372/QALY (Serbia) to €90,893/QALY (USA) (colab.ws), all exceeding usual WTP thresholds. This highlights how local practice and costs drive HEOR outcomes.
- Crucially, the methodology (using standardized R tools on each site's data) exemplifies how "data-driven algorithmic modeling" can scale. It avoids one-off manual coding for each locale.

This study underscores AI's role indirectly: by standardizing and automating analytic workflows across data sites, it lays the groundwork for AI to further refine models. For example, once these R packages define the structure, ML could be applied on each dataset to optimize state definitions or detect novel health states. The success of Haug's federated approach also suggests that AI-driven RWE methods can be deployed *globally* without moving raw data, an important privacy-aware advance.

AI in Clinical Decision Support (Afshar et al. 2025)

Beyond pure HEOR modeling, AI applied to RWE can directly influence economics via clinical impact. A notable example is a pre-post quasi-experimental study by Afshar et al. published in *Nature Medicine* in 2025 (pubmed.ncbi.nlm.nih.gov). In this study, a convolutional neural network running in real-time on hospital EHR notes identified inpatients at risk of opioid use disorder (OUD) and alerted clinicians to conduct an addiction medicine consult. The AI-driven screener was compared to a historical control period:

- The rate of OUD consults was maintained (non-inferior to usual care), demonstrating the AI was targeting appropriate patients (scholars.luc.edu).
- The AI screener was associated with a **reduction in 30-day readmissions** (odds ratio 0.53, 95% CI 0.30–0.91). In economic terms, the paper reports an incremental cost of $6,801 per readmission avoided (pubmed.ncbi.nlm.nih.gov). This is a net cost per readmission avoided rather than a demonstrated saving, and the study is a pre-post quasi-experimental comparison rather than a randomized trial.
- The authors concluded that embedding AI prediction in the EHR was a **scalable, cost-effective solution for OUD care** (scholars.luc.edu). This illustrates how an AI algorithm applied to EHR data may support better outcomes: in a pre-post quasi-experimental study, the screener was associated with fewer 30-day readmissions at an incremental cost of about $6,801 per readmission avoided, which is not evidence of system-level cost savings.

While this study was in operational medicine rather than a pharmacoeconomic model, it highlights the downstream HEOR implications of AI-RWE synergy. If pharma-sponsored models assume certain readmission rates or complication rates, an AI intervention that changes those probabilities will alter the cost-effectiveness landscape. In this sense, AI-driven RWE not only informs static models but can dynamically change real-world outcomes and associated costs.

Discussion of Implications and Challenges

The integration of AI into pharma HEOR has wide-ranging implications:

- **Faster Evidence Generation**: AI can dramatically reduce the time to generate HEOR evidence. The case studies above show days-worth of work potentially condensed to hours. Model updates can be automated; sensitivity analyses can run in parallel; reports can be drafted by toolkit. This speed can help pharma respond rapidly to payer inquiries or new data.
- **Richer, More Personalized Analyses**: By exploiting RWD at scale and leveraging ML's pattern-recognition, HEOR becomes more granular. Instead of one "average" patient cost-effectiveness, models can cater to precision medicine scenarios. Value frameworks may evolve to include AI-driven stratified outcomes.
- **Improved Coverage Decisions**: For payers and HTA agencies, AI-enhanced RWE can provide more confidence in decisions. For example, causal AI methods may strengthen certainty around off-trial uses ("does this drug work in elderly patients not well-represented in trials?"), supporting broader reimbursement. Payers could incorporate AI-generated risk predictions into outcomes-based contracts.
- **Patient Impact**: In the long run, better evidence leads to quicker and more accurate access decisions. If AI shows a therapy is highly beneficial and cost-effective for a subset of patients, those patients may gain faster access. Conversely, evidence of low real-world benefit could prevent waste. Additionally, AI tools at the point of care (like the OUD screener) have direct healthcare value, improving outcomes that feed back into economic models.

However, there are important **challenges**:

- **Data Quality and Bias**: AI models are only as good as the data they learn from. Real-world data may contain biases (e.g., certain populations under-represented, or systemic treatment differences). If an algorithm is trained on biased data, its predictions will perpetuate those biases. For example, if a registry lacks socio-economic diversity, an AI model built on it may not generalize. Transparency about data provenance and demographic coverage is essential. Analysts must apply techniques to detect and correct bias (reweighting, fairness metrics).
- **Transparency and Interpretability**: Many AI models (deep neural nets, ensemble trees) are “black boxes” relative to simpler statistical models. HEOR requires justification of assumptions. New methods for interpretability (e.g. SHAP values, LIME) and rigorous validation (back-testing, prospective validation studies) must be used. Indeed, NICE’s August 2024 position statement (since replaced by its AI methods for evidence framework) warned of “concerns about the appropriateness, transparency and trustworthiness of AI” in evidence generation (www.nice.org.uk). Insurers and regulators will want to scrutinize how AI-derived evidence was produced.
- **Regulatory Acceptance**: While regulators encourage RWE, they are still adapting to AI analyses. Health economists may need to provide additional documentation when AI tools are used. For instance, in submissions, references to AI methods (algorithm details, training data) might be required. Standards (like model reporting checklists) will need updating to cover ML techniques. Collaborative efforts (e.g. ISPOR Good Practice Task Forces) are already underway to produce guidelines on machine learning in HEOR.
- **Generalizability of AI Tools**: Early LLM studies (Reason et al., Rawlinson et al.) show promise, but caution is warranted. As Rawlinson notes, “there is still much work to do” and further testing across diverse diseases and model types is needed (www.ispor.org). An AI tool trained on oncology models, for example, may not immediately work for a cardiovascular economic model without retraining or new prompts. Organizations must beware of over-generalizing early successes. Ongoing research and cross-validation across multiple use cases are needed before full confidence in AI automation.
- **Technical Infrastructure and Skills**: Effective use of AI requires computing resources (GPUs, cloud costs) and staff skilled in data science and ML. Not every HEOR group has this capability in-house. Pharmaceutical companies may need to train their health economists or partner with tech vendors. Similarly, data scientists must understand health economics concepts to build appropriate models. This interdisciplinary knowledge gap is a key barrier and will require investment in training programs.
- **Privacy and Security**: Federated and synthetic approaches help, but the underlying patient data is sensitive. Companies must maintain strict data governance. Use of public LLMs (like ChatGPT) poses risks if proprietary data is inadvertently exposed. As Rawlinson (Estima Scientific) cautions in ISPOR’s Value & Outcomes Spotlight, confidential information “should not be submitted to public LLMs and any outputs should be checked by a human” (www.ispor.org). Ensuring data security while leveraging cloud-based AI services is a balancing act.
- **Economic Implications**: On the one hand, AI lowers the cost of analysis (fewer analysts hours, cloud computing cost-effectiveness). On the other hand, developing and validating AI tools has upfront costs. The return on investment depends on volume of analysis. For very large multinational companies, AI

likely offers net savings. Smaller companies or single-country studies may not justify high-tech solutions. Yet, as AI tools become more commoditized, even smaller players will be able to use pre-built models and platforms.

In discussing these challenges, it is important to note that many are addressable. For example, bias can be partly mitigated by incorporating fairness criteria during training (ensuring equal predictive performance across subgroups). Transparency is aided by publishing AI models and synthetic datasets for external audit. Early collaborations between HTA bodies and industry (e.g., pilot projects on AI-generated reports) will help establish best practices. The field is moving quickly, and governance structures are evolving to keep pace.

Future Directions and Implications

The integration of AI and RWE in HEOR is a rapidly evolving frontier. The long-term implications are profound:

- **Dynamic, Living Models**: Instead of one-off static models, we may see *living economic models* that update as new data arrives. AI systems can continuously ingest fresh RWD (for example, monthly claims updates) and recalibrate model parameters. Such dynamism would allow paying agencies to see near-real-time cost-effectiveness updates. Health economists would manage version control and track model drift with AI monitoring alerts.
- **Digital Twins of Patient Populations**: A visionary idea is the “digital twin” – creating a virtual representation of a cohort of patients. This involves simulating individual patient trajectories through disease and treatment using AI-powered microsimulation. Generative models could create millions of synthetic patient journeys that incorporate uncertainty. Economics would then be calculated across this synthetic cohort, potentially improving precision. Some pioneering companies are already embedding AI simulators in their platforms for scenario analysis.
- **Integration of Omics and Imaging**: As genomic sequencing and medical imaging become routine, these rich data types will enter HEOR. For example, imagine a vaccine priced differently for patients with high genetic risk of severe disease (predicted by an ML model on genomic markers). Or include AI-extracted imaging biomarkers as covariates in cost models. The fusion of multi-omics RWD and AI could enable truly personalized cost-effectiveness (e.g. pharmacogenomic subgroups).
- **Transparency and Explainability Advances**: Paradoxically, as AI use spreads, we expect innovations in explainability to keep stakeholders comfortable. Methods like counterfactual explanation, surrogate modeling, and interactive Venn diagrams of feature impacts will evolve. Visual tools that show “if we change this input, the predicted outcome changes that much” will become standard for AI models in submissions.
- **Regulatory Frameworks and Standards**: Agencies will likely publish more concrete guidelines on AI use in HEOR. We anticipate consensus reporting standards (akin to CHEERS for health economics) for ML-driven studies. There may be certification or auditing standards for AI models used in decision-making. Collaboration between pharma, regulators, and payers could even lead to pre-specifications: e.g., an AI-based RWD analysis plan agreed in advance.

- **AI-driven Value-Based Contracts**: Another emerging use is automating outcomes monitoring under value-based contracts. If an insurer promises payments tied to real-world outcomes, AI systems can continuously analyze patient data to trigger payments or refunds. For example, a cancer drug may have a pay-for-response clause: AI models could quickly assess e.g. progression-free survival from imaging and lab data to calculate refunds if targets aren't met. This tight coupling of RWE and economics would be powered by AI.
- **Broader AI Ecosystem**: We may see HEOR teams tapping public LLMs for more heuristic tasks (writing code snippets, summarizing literature) and private, patient-specific AI systems for clinical decision support. Cross-industry platforms may emerge for HEOR analytics: pre-built AI models fine-tuned on de-identified RWD across pharma, which companies subscribe to.

Looking further ahead, challenges in AI will also shift the field. For instance, as quantum computing matures, it could enable even faster simulation of economic models or optimization of pricing across millions of scenarios. Ethical AI considerations (equity, algorithmic transparency) will likely pervade HTA deliberations. Rare diseases, where data are scarce, stand to benefit especially: synthetic RWD and ML can fill evidence gaps that today preclude analysis.

In essence, we stand at the cusp of an AI-driven revolution in HEOR. Every stage of evidence generation – from data ingestion to modeling to reporting – is being reimagined. The future HEOR function may emphasize *AI supervision and oversight* rather than manual calculation. Health economists might spend more time designing prompts, interpreting outputs, and aligning AI with policy questions. The output of HEOR studies will likely become more dynamic, individualized, and data-rich.

Nevertheless, the core goal remains unchanged: to rigorously assess the value of interventions for patients and health systems. As the subtitle of an earlier Chen et al. (2016) precision-HEOR paper puts it, "big data meets patient heterogeneity on the road to value" (pmc.ncbi.nlm.nih.gov). AI is the vehicle on that road. When properly harnessed, it will not replace traditional health economics, but enhance it – uncovering value signals buried in the data and enabling smarter decisions about therapies and prices.

Conclusion

This report has surveyed a rapidly maturing landscape at the intersection of AI, real-world evidence, and pharmaceutical HEOR. The evidence is clear: AI tools are already reshaping how companies generate evidence of value. **From automated cost-effectiveness models** to **precision patient stratification** to **synthetic data augmentation**, AI is extending what is possible with RWD.

Key conclusions include:

- **RWD is essential**: Real-world data from a variety of sources (EHRs, claims, registries, digital devices) has become a cornerstone for modern HEOR. It provides large, heterogeneous samples and long-term outcomes beyond trials (www.frontiersin.org) (doaj.org). Regulatory and payer environments are increasingly embracing RWE knowledge, spurred by policies like the FDA's RWE programs and NICE's strategic roadmap (www.tandfonline.com) (www.nice.org.uk).

- **AI enables next-generation analytics**: AI/ML methods can handle the scale and complexity of RWD, turning raw data into actionable insights. Advanced predictive analytics, unsupervised learning, and causal modeling allow researchers to ask novel questions (e.g. “which patients derive the most net benefit?”) and simulate scenarios (including counterfactuals) (www.mckinsey.com) (www.tandfonline.com). This goes far beyond what conventional statistics could do.
- **Economic modeling is being transformed**: Generative AI, especially large language models, can automate tasks that were once entirely manual. Groundbreaking studies have shown AI’s ability to **auto-generate model code** and **adapt models** with minimal errors (pubmed.ncbi.nlm.nih.gov) (www.sciencedirect.com). As these tools improve, we expect them to become staples of the HEOR toolkit. AI also facilitates precision HEOR by identifying patient subgroups with different cost-effectiveness profiles (pmc.ncbi.nlm.nih.gov).
- **Challenges remain**: Data quality, bias, transparency, and validation are ongoing concerns. Wholesale trust in AI is not yet warranted – experts repeatedly caution that more validation is needed (www.ispor.org) (www.nice.org.uk). Skills and infrastructure must catch up to support AI adoption. Furthermore, low- and middle-income settings, where RWD infrastructure is weaker, may lag in reaping these benefits.
- **Regulatory and ethical frameworks are adapting**: Agencies and societies are developing guidelines for AI-led evidence. For example, NICE issued an August 2024 position statement acknowledging AI’s role in evidence generation but urging careful oversight, and has since replaced it with its AI methods for evidence framework (www.nice.org.uk). In regulatory submissions, companies will need to document how AI was used and demonstrate reliability. Good practice is emerging (e.g. ISPOR task forces on machine learning), but formal standards are still evolving.
- **The future is data-driven and patient-focused**: Ultimately, AI and RWE together promise a more efficient route to understanding the value of medicines. By continuously learning from real-world outcomes, AI-driven models can ensure health economic analyses stay up-to-date and relevant. Patients stand to benefit from faster access decisions and potentially more personalized therapy selection (as value models become patient-specific).

In closing, the message from the literature and industry is optimistic but measured. As one review states, “Synthetic data provide a unique opportunity to overcome data-related challenges in HEOR” (www.sciencedirect.com) – a succinct emblem of AI’s promise in this field. We anticipate that within a few years, AI-generated RWE and AI-enabled economic models will be commonplace in HTA submissions and payer negotiations. However, the field must proceed with rigor. The lessons so far suggest that combining **human expertise with machine intelligence** yields the best outcomes: machines can compute at scale, but humans set the focus, validate the results, and uphold scientific standards.

The rapid developments of 2023–2025 (GPT-4 studies, federated networks, synthetic data) have thrust HEOR into a new era. It is incumbent on all stakeholders – pharmaceutical companies, health economists, regulators, and payers – to engage with these technologies actively. By co-developing standards, sharing best practices,

and transparently testing AI tools, the community can ensure AI in RWE and modeling becomes a reliable pillar of evidence, not just a hype. When used judiciously, AI will help realize the vision of **value-based, evidence-driven healthcare**, ultimately improving outcomes for patients and societies alike.

Sources: The above report synthesizes findings from academic journals (e.g. *Frontiers in Pharmacology*, *Journal of Health Economics & Outcomes Research*, *PharmacoEconomics Open*, *Journal of American Medical Informatics Association*), industry analyses (McKinsey), and regulatory documentation (FDA, NICE). Key figures and findings are drawn from the cited literature, while some general background statements are uncited (doaj.org) (www.sciencedirect.com) (pubmed.ncbi.nlm.nih.gov) (scholars.luc.edu) (colab.ws) (www.nice.org.uk). Each source is considered a credible, peer-reviewed, or authoritative position on the subject.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://doaj.org/article/dcc0fa67d38e43c9af0d97214f03fe81
2. https://www.tandfonline.com/doi/abs/10.1080/14737167.2023.2275674
3. https://www.frontiersin.org/articles/10.3389/fphar.2021.700012/full
4. https://www.nice.org.uk/position-statements/use-of-ai-in-evidence-generation-nice-position-statement.%202
5. https://www.mckinsey.com/industries/life-sciences/our-insights/creating-value-from-next-generation-real-world-evidence
6. https://pmc.ncbi.nlm.nih.gov/articles/PMC7299485
7. https://pubmed.ncbi.nlm.nih.gov/38340276
8. https://www.sciencedirect.com/science/article/pii/S109830152502399X
9. https://www.sciencedirect.com/science/article/abs/pii/S1098301525023125
10. https://www.ispor.org/heor-resources/presentations-database/presentation/euro2024-4014/144576
11. https://pubmed.ncbi.nlm.nih.gov/40181180
12. https://www.ispor.org/publications/journals/value-outcomes-spotlight/vos-archives/issue/view/heor-innovation-and-impact/generative-artificial-intelligence-and-the-future-of-health-economic-modeling
13. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory
14. https://www.ema.europa.eu/en/news/use-real-world-evidence-regulatory-decision-making-ema-publishes-review-its-studies
15. https://scholars.luc.edu/en/publications/outcomes-and-cost-effectiveness-of-an-ehr-embedded-ai-screener-fo
16. https://eprints.whiterose.ac.uk/id/eprint/162472
17. https://colab.ws/articles/10.1093%2Fjamia%2Focae044
18. https://pmc.ncbi.nlm.nih.gov/articles/PMC5198059

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.