IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

By Adrien Laurent, CEO at IntuitionLabs • 1/8/2026 • 45 min read

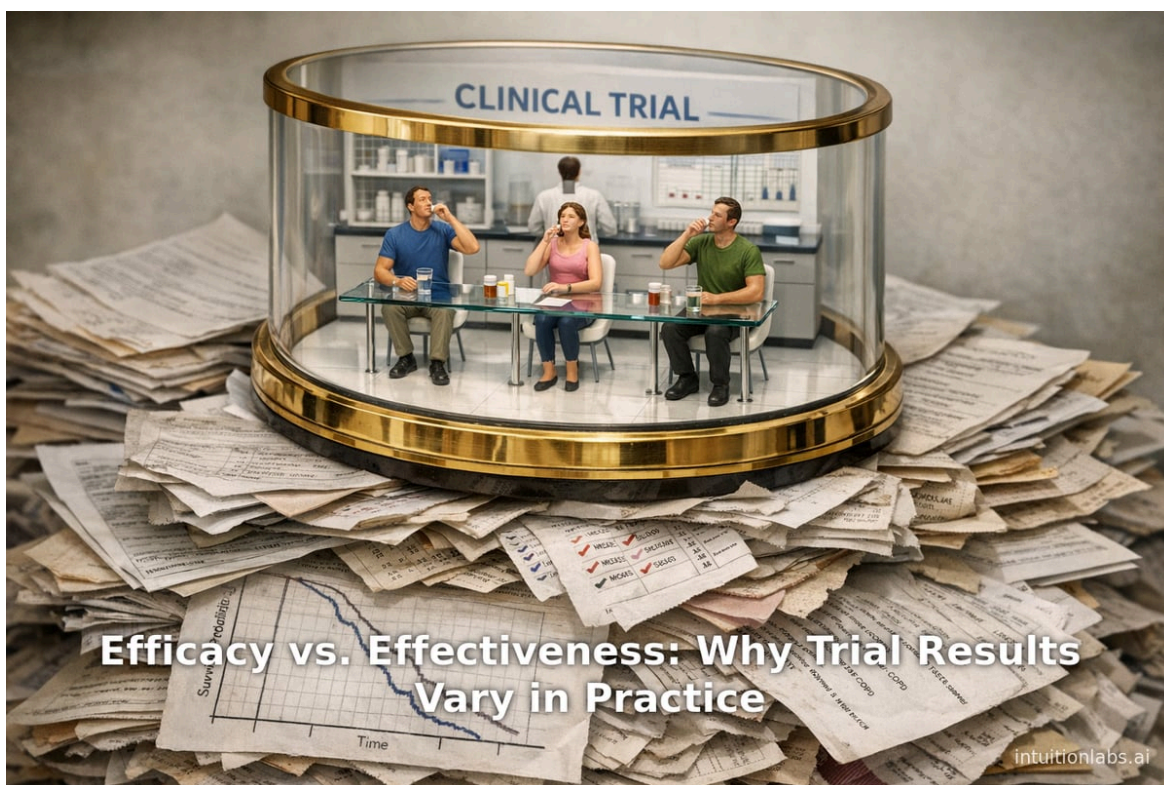efficacy vs effectiveness   heor   real-world evidence   clinical trials   rwe   external validity

health economics   rwd

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

Executive Summary

Drugs often show higher **efficacy** in controlled clinical trials than *effectiveness* in routine practice. This so-called *efficacy–effectiveness gap* arises from fundamental differences between the trial environment and real-world conditions ([1] pmc.ncbi.nlm.nih.gov) ([2] pmc.ncbi.nlm.nih.gov). In trials, rigid eligibility criteria, intensive monitoring, and strict adherence protocols maximize observed benefit, whereas in practice a broad population of patients with varying characteristics and behaviors uses treatments under heterogeneous conditions ([1] pmc.ncbi.nlm.nih.gov) ([3] pmc.ncbi.nlm.nih.gov). As a result, real-world response rates, durations of effect, and safety profiles often differ from trial observations. For example, large observational studies have repeatedly found that real-world patients experience shorter survival or treatment response than highly selected trial cohorts ([4] pmc.ncbi.nlm.nih.gov) ([5] pmc.ncbi.nlm.nih.gov). Similarly, rare adverse events or comorbid interactions that were undetected in trials have been identified post-approval through real-world data (RWD) surveillance ([6] pmc.ncbi.nlm.nih.gov) ([7] pmc.ncbi.nlm.nih.gov).

Health economics and outcomes research (HEOR) explicitly accounts for these variances by integrating real-world evidence (RWE) into economic models, comparative-effectiveness analyses, and policy decision-making. HEOR practitioners use diverse data sources – including electronic health records, insurance claims, registries, and patient-reported outcomes – to **adjust** model inputs for actual patient characteristics, adherence patterns, and resource use. For example, cost–effectiveness models often incorporate RWE-derived estimates of long-term event rates, comorbidity prevalence, and healthcare costs to "calibrate" projections to routine settings (apices.es) ([8] pmc.ncbi.nlm.nih.gov). Value frameworks and payers increasingly demand RWE to validate or refine the benefits measured in trials, especially for heterogeneous patient subgroups (apices.es) ([9] pmc.ncbi.nlm.nih.gov). This approach can dramatically affect value assessments: HEOR analyses using real-world inputs have shown that certain therapies with seemingly favorable trial results may be less cost-effective in broader populations, and conversely that some interventions retain high value due to unanticipated real-world benefits ([4] pmc.ncbi.nlm.nih.gov) ([10] pmc.ncbi.nlm.nih.gov).

This report explores the historical context, underlying causes, and implications of trial–real-world discrepancies. We review multiple perspectives – including clinical, regulatory, payer, and statistical – and detail how HEOR bridges the gap. We examine theoretical concepts (like internal vs external validity and the efficiency-effectiveness gap), present data and case studies across disease areas, and discuss methodological tools (pragmatic trials, advanced statistical adjustment, digital health metrics) that HEOR uses to reconcile differences. The report concludes with future directions: as regulators and payers formalize RWE frameworks and as AI-enabled analytics mature, HEOR will play an even more central role in generating robust evidence that reflects how drugs actually perform outside trials.

Introduction and Background

Randomized controlled trials (RCTs) have long been considered the **gold standard** for demonstrating drug efficacy and safety during development ([11] pmc.ncbi.nlm.nih.gov) ([12] pmc.ncbi.nlm.nih.gov). By randomizing patients and enforcing tightly-controlled protocols (fixed dosing, stringent follow-up, etc.), RCTs **minimize bias** and ensure that observed effects can be attributed to the intervention ([12] pmc.ncbi.nlm.nih.gov) ([2] pmc.ncbi.nlm.nih.gov). However, these very design features – while maximizing internal validity – limit **external validity**. Strict eligibility criteria and extra monitoring produce an artificial environment: trial patients are often younger, healthier, and more adherent than typical patients, and their care is delivered by dedicated research teams. As a result, trial outcomes (statistical efficacy) may not generalize to usual practice (actual effectiveness) ([2] pmc.ncbi.nlm.nih.gov) ([6] pmc.ncbi.nlm.nih.gov). Indeed, as one clinician noted, applying trial

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

results "is often not straightforward" – one must ask whether the trial population matches real patients and whether the totality of evidence (outside the trial) supports the findings ([13] pmc.ncbi.nlm.nih.gov).

The notion of an "efficacy–effectiveness gap" has gained growing attention in recent decades ([14] pubmed.ncbi.nlm.nih.gov) ([6] pmc.ncbi.nlm.nih.gov). Historically, decision-makers (regulatory agencies, payers, guideline bodies) have relied almost exclusively on trial evidence when approving drugs or writing guidelines ([11] pmc.ncbi.nlm.nih.gov) ([12] pmc.ncbi.nlm.nih.gov). Yet as healthcare systems evolved, stakeholders began observing that many therapies do not achieve their full trial-proven benefits in practice. For example, an intervention that reduced hospitalization by 30% in an RCT might show only a 15% reduction in ordinary care. Repeatedly, retrospective studies and registries have uncovered discrepancies: trials failing to predict rare issues, or new populations responding differently than expected. Such findings cast doubt on the assumption that trial efficacy implies broad clinical effectiveness ([12] pmc.ncbi.nlm.nih.gov) ([6] pmc.ncbi.nlm.nih.gov).

This realization has led to a paradigm shift: regulators, payers, and researchers now seek **Real-World Evidence (RWE)** to complement trials. In the US, the 21st Century Cures Act of 2016 authorized FDA to consider RWE (from electronic health records, registries, etc.) for supplemental approvals, and in 2018 the FDA launched an RWE Program with rigorous guidance ([15] pmc.ncbi.nlm.nih.gov) ([16] pmc.ncbi.nlm.nih.gov).The European Medicines Agency (EMA) and other authorities have pursued similar visions ([17] pmc.ncbi.nlm.nih.gov) (www.nice.org.uk). Health economics and outcomes research was an early adopter of RWE: HTA bodies like NICE now explicitly promote using real-world data to "reduce uncertainties and improve guidance" beyond what RCTs provide (www.nice.org.uk). The ISPOR community has framed this as not just an optional add-on, but a non-negotiable element for value assessment in modern healthcare (apices.es) ([14] pubmed.ncbi.nlm.nih.gov).

Despite this shift, challenges remain. High-quality RWE studies require careful design to mitigate bias, and not every difference between trial and practice is easy to quantify. The balance between learning from trials and justify decisions under uncertainty is delicate. The emergent field of HEOR explicitly addresses these issues by blending methodological rigour with practical modeling. In cost-effectiveness models and budget impact analyses, HEOR teams routinely 'calibrate' trial effect estimates to real-world settings using RWE inputs (apices.es) ([8] pmc.ncbi.nlm.nih.gov). In clinical pathways analysis, HEOR disaggregates outcomes by sub-populations (age, race, comorbidity) found in RWD (apices.es). By doing so, HEOR produces an evidence base that reflects the *actual* benefits, harms, and costs of interventions in the healthcare system.

This report proceeds as follows: we first dissect *why* and *how* trial results diverge from real-world performance. We then review empirical data and case studies illustrating the gap across different contexts. Next, we examine HEOR methods – modeling and analytic approaches – that explicitly incorporate or adjust for real-world variance. We discuss the implications for decision-making (regulatory, coverage, and clinical) and look ahead to future developments in RWE-driven HEOR (including digital data and AI assistance). Throughout, we provide extensive references to peer-reviewed literature and expert sources to substantiate key points.

Key Differences: Controlled Trials vs Real-World Practice

Internal vs External Validity

RCTs are designed to maximize **internal validity** – ensuring that differences in outcomes are attributable to the treatment, not confounders. To achieve this, RCTs enforce **strict eligibility criteria** (e.g. excluding elderly, patients with comorbidities or concomitant medications) ([1] pmc.ncbi.nlm.nih.gov) ([2] pmc.ncbi.nlm.nih.gov). Patients are monitored intensively, take medications under supervision, and follow a rigid protocol without the

IntuitionLabs IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

complexities of routine care ([2] pmc.ncbi.nlm.nih.gov). As a result, the study population tends to be highly selected. By contrast, **real-world practice** involves a broader patient base: individuals of all ages, with multiple chronic illnesses, varied socio-economic backgrounds, and different levels of health literacy participate in treatment ([3] pmc.ncbi.nlm.nih.gov) ([18] pmc.ncbi.nlm.nih.gov). In practice, physicians may treat patients whom no trial would have enrolled. For example, one population study found that *"up to 70% of real-world patients would have been excluded from landmark clinical trials based on their age, comorbidities, cytopenias, or organ function."* ([18] pmc.ncbi.nlm.nih.gov).

Likewise, RCTs tightly control the **treatment protocol**. Dosages, administration schedules, and even background care are standardized and enforced. Patients in trials are often required to adhere to medication regimens under frequent supervision, and investigators follow pre-specified dose modifications if adverse effects appear. In routine care, by contrast, medications are prescribed more flexibly: doctors adjust doses based on patient tolerance or cost, pharmacies dispense refill medications which patients may skip, and there is no centralized monitoring of adherence unless specifically studied. Thus, **adherence rates** tend to be much higher in trials than in the community. Patients in trials also tend to be more motivated (having consented to research), which can inflate treatment effects. For instance, in oncology RCTs, practically every participant remains on therapy until progression, whereas in practice a significant fraction of patients discontinue or deviate from the ideal schedule. A recent study of multiple myeloma noted that while trial patients followed *"strict protocols that require close patient monitoring and prespecified dose reductions,"* real-world patients **"may have lower adherence or may have received lower doses of the same regimens"** ([19] pmc.ncbi.nlm.nih.gov). Such differences can naturally lead to smaller observed benefits in practice.

Furthermore, trials typically impose **short follow-up times** (1–3 years) because of cost and logistics, which may be insufficient to capture long-term outcomes or late-emerging effects. In contrast, real-world observational studies can follow patients for much longer, sometimes over decades, thereby revealing effects on chronic outcomes that trials miss. For example, a drug whose trial was 6 months cannot teach us its impact on 5-year mortality; only real-world data can fill that gap. Conversely, some short-term benefits seen in trials (e.g. rapid symptom relief) may be less enduring in broader use.

Finally, the **outcomes measured** often differ. RCTs frequently rely on **surrogate or physiological endpoints** (tumor shrinkage, blood pressure change, biomarker levels) to detect effects under controlled conditions. These do not always translate directly to meaningful patient outcomes. In routine practice, the true concerns are patient-centric endpoints (hospitalizations, quality-of-life, disability, productivity), which may or may not correlate tightly with trial measures. For instance, the "view 1" and "view 2" trial of an ophthalmic drug showed somewhat different numerical results, illustrating that even under similar trial protocols, variability exists ([20] pmc.ncbi.nlm.nih.gov). In practice, outcomes like vision-related quality of life or ability to work might be more relevant.

Collectively, these contrasts between trial design and real conditions – summarized in Table 1 – explain much of the performance gap. RCTs maximize precision and causality at the expense of generalizability, whereas everyday care features all the complexity of a heterogeneous health system.

Feature	Clinical Trial (High Efficacy Setting)	Real-World Practice (Effectiveness Setting)
Population	Highly selective (strict inclusion/exclusion; comorbidities often excluded) ([1] pmc.ncbi.nlm.nih.gov) ([2] pmc.ncbi.nlm.nih.gov)	Broad and heterogeneous (all ages, multiple comorbidities, varied social/demographic factors) ([3] pmc.ncbi.nlm.nih.gov) ([21] pmc.ncbi.nlm.nih.gov)
Patient Adherence	Monitored and enforced (patients closely followed, high adherence)	Variable (determined by patient behavior, cost, side-effects; often lower)
Treatment Protocol	Fixed dosing schedule and adjustments per protocol ([2] pmc.ncbi.nlm.nih.gov)	Flexible dosing (clinician-driven adjustments, early discontinuation possible) ([19] pmc.ncbi.nlm.nih.gov)

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

Feature	Clinical Trial (High Efficacy Setting)	Real-World Practice (Effectiveness Setting)
Follow-Up Intensity	Frequent, structured visits and tests to detect endpoints ([22] pmc.ncbi.nlm.nih.gov)	Irregular follow-up per standard care; outcomes captured in routine records
Outcome Measures	Clinical/surrogate endpoints defined in protocol (e.g. tumor response)	Pragmatic outcomes (e.g. hospitalization, survival, QoL) reflecting real patient impact
Adverse Events	Limited detection (small sample, short duration; rare events often missed) ([23] pmc.ncbi.nlm.nih.gov) ([7] pmc.ncbi.nlm.nih.gov)	Better detection (large, diverse population over time; signals rare/late AEs) ([7] pmc.ncbi.nlm.nih.gov)
Bias Control	Randomized, controlled (minimizes known and unknown confounders) ([24] pmc.ncbi.nlm.nih.gov)	Observational (susceptible to confounding; requires statistical adjustment)

Table 1. Key contrasts between randomized-controlled trial conditions and real-world clinical practice. Trial results often reflect ideal circumstances, while actual patient care is subject to heterogeneity and practical constraints.

Statistical Explanations: Efficiency vs Effectiveness

Economists and outcomes researchers often frame this difference as an **efficacy-effectiveness (EE) gap** ([14] pubmed.ncbi.nlm.nih.gov). Under this paradigm, efficacy is what a drug does under ideal (trial) conditions, while effectiveness is what it does under routine usage. Multiple factors contribute to the EE gap ([25] pubmed.ncbi.nlm.nih.gov): healthcare system characteristics (access, provider behavior), data and measurement methods, and the interaction of drug biology with context. In the words of an ISPOR working group, bridging this gap requires understanding *"how a drug's biological effect interacts with contextual factors"* in practice ([26] pubmed.ncbi.nlm.nih.gov).

From a methodological standpoint, **replicability** and **reproducibility** play a role ([20] pmc.ncbi.nlm.nih.gov). Replicability (getting the same result under identical conditions) is difficult even between two trials – patient samples will always slightly differ. But reproducibility (observing consistent effects under varied conditions) is what matters clinically ([27] pmc.ncbi.nlm.nih.gov). RCTs often sacrifice reproducibility for experimental control. If an RCT's result is highly idiosyncratic to its narrow setting, it might not generalize. Practical experience shows this: even identical trials (like the VIEW1/VIEW2 ophthalmology trials) can yield different numeric outcomes ([20] pmc.ncbi.nlm.nih.gov). Thus clinicians rightly ask whether trial findings can be reproduced on average in their broader practice.

In observational studies (and HEOR analyses), the goal is to assess effectiveness in the actual health system. This requires careful handling of confounding biases that RCTs avoid by design. While many comparative observational analyses report **similar relative effects** as RCTs ([5] pmc.ncbi.nlm.nih.gov), some show significant discrepancies. A major systematic review found that about 80% of comparisons had no statistically significant difference between RCTs and observational studies, but in roughly 20% of comparisons the results varied substantially, even showing opposite directions ([5] pmc.ncbi.nlm.nih.gov). This inconsistency can arise from two sources: true population differences (e.g. trial excluded sicker patients who in reality fare worse) and from bias/methodological flaws in the observational work ([5] pmc.ncbi.nlm.nih.gov). Distinguishing these requires advanced HEOR techniques (propensity score methods, censoring adjustments, etc.) and often meta-research comparing emulated trials to real ones ([28] pmc.ncbi.nlm.nih.gov) ([5] pmc.ncbi.nlm.nih.gov). A recent study of extensive-stage small-cell lung cancer is illustrative: by comparing matched cohorts, researchers showed that outcome *discrepancies depended critically on both patient selection and operational differences (e.g., censoring rules, assessment frequency) in trials versus real care* ([29] pmc.ncbi.nlm.nih.gov). In short, statistical differences in trial operation add further *operational bias* on top of simple patient differences.

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

In summary, trials are statistically optimized to answer *efficacy* questions, whereas real-world studies – if well-designed – provide *effectiveness* answers, capturing broader variability. Understanding when an apparent gap reflects a methodological artefact or true difference in outcomes is a core challenge. The HEOR field uses specialized frameworks (e.g. the PRECIS criteria for trial pragmatism ([30] pmc.ncbi.nlm.nih.gov), or "target trial emulation" techniques) to quantify how each evidence source contributes to overall understanding. We return below to how HEOR models synthesize these sources to produce robust decision inputs.

Evidence of the Efficacy-Effectiveness Gap: Data and Case Studies

A wealth of empirical studies document the trial-to-practice gap. We highlight representative findings from different fields, showing how real-world outcomes have diverged from RCT predictions. (Table 2 at the end of this section summarizes selected examples.) Wherever possible we quote actual data from the literature.

Oncology: Cancer therapies provide perhaps the most dramatic illustrations. For example, a recent population-based study in multiple myeloma (MM) directly quantified outcomes in patients receiving standard regimens in routine care versus those in the registrational RCTs ([4] pmc.ncbi.nlm.nih.gov). The results were striking: real-world transplant-ineligible MM patients had a *51% higher risk of progression or death* compared to trial patients on the same lenalidomide-based therapy (pooled hazard ratio 1.51) ([4] pmc.ncbi.nlm.nih.gov). Six out of seven standard regimens showed significantly shorter median progression-free survival (PFS) in the real world – often 7–18 months shorter than trials – and six of seven showed significantly worse overall survival ([4] pmc.ncbi.nlm.nih.gov) ([31] pmc.ncbi.nlm.nih.gov). Key differences explain this gap: the real-world cohort was older, more heavily pre-treated, and had more comorbidities than trial patients (70% of real patients would have been excluded from those RCTs by common eligibility criteria) ([18] pmc.ncbi.nlm.nih.gov). Moreover, almost two-thirds of real patients had prior therapy and many had only short "bridging" treatment in early disease, whereas trials typically enrolled clean (no prior chemo) patients with better prognoses. The study emphasizes that *"stringent RCT inclusion criteria"* exclude exactly those sicker patients who, in practice, weaken overall outcomes ([18] pmc.ncbi.nlm.nih.gov). The same analysis also noted that close trial monitoring and strict dose titration probably yielded higher adherence than real care, further boosting trial results ([19] pmc.ncbi.nlm.nih.gov). This seminal work – one of the largest head-to-head RCT vs RWD comparisons – underscores how real outcomes can be substantially worse when therapy is delivered in broader practice.

Similarly, in colorectal cancer, comparative meta-analyses have been mixed but instructive. One systematic review comparing RCTs and real-world cohorts treated with modern regimens (bevacizumab, cetuximab, XELOX) found that, on average, the overall survival and response rates were *not significantly different* between settings ([32] pmc.ncbi.nlm.nih.gov). However, subtle shifts appeared: trial patients were slightly more likely to be male and much less likely to have poor performance status (ECOG ≥2) ([32] pmc.ncbi.nlm.nih.gov). Importantly, in one subgroup (cetuximab), the RCTs reported about 4 months longer median OS, which the authors linked to differences in treatment line (trial patients generally got the drug earlier in their disease) ([33] pmc.ncbi.nlm.nih.gov). Likewise, in multiple myeloma, another population study found real-world patients treated for newly diagnosed disease fared better than trial expectations (possibly due to earlier treatment), whereas relapsed patients fared much worse than RCT norms ([34] pmc.ncbi.nlm.nih.gov). These examples show that gaps can vary by context: sometimes real-world practice may improve on trial results (for example, if uptake of supportive measures is better) or worsen them (especially in second-line or refractory settings).

Cardiology and General Medicine: In chronic diseases, many examples have emerged. For instance, analyses of antihypertensive and heart-failure therapies often find smaller mortality reductions in observational cohorts than in trials ([5] pmc.ncbi.nlm.nih.gov). A notable case is rofecoxib (Vioxx), a COX-2 inhibitor. Trials of rofecoxib in arthritis excluded patients with cardiovascular disease, and no signal for heart attacks appeared in those

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

reports ([35] pmc.ncbi.nlm.nih.gov). Once rofecoxib was used widely, massive RWD analyses (case–control and cohort studies) revealed a **reproducible increase in myocardial infarction risk** (studies showed roughly 2–3 times higher MI rates) for patients on rofecoxib ([35] pmc.ncbi.nlm.nih.gov). This real-world finding ultimately prompted the drug's withdrawal despite its absence in trial data. Another high-profile case is intensive glucose-lowering in diabetes: large RCTs had demonstrated that strict glycemic targets reduced some microvascular outcomes, but pragmatic registries later showed that very tight control (especially in older patients) could increase overall mortality (likely due to hypoglycemia and comorbidities) ([28] pmc.ncbi.nlm.nih.gov) ([5] pmc.ncbi.nlm.nih.gov). In stroke prevention, observational comparisons (e.g., in AFib patients on warfarin vs NOACs) often yield effect sizes in line with trials, but subclasses with poor adherence or socio-economic barriers clearly do worse outside trials ([36] pmc.ncbi.nlm.nih.gov).

Rare Adverse Events and Pharmacovigilance: Both [2] and [30] emphasize that rare or delayed safety issues are often *hidden* in pre-approval trials but appear post-marketing. For example, immune checkpoint inhibitors for cancer looked very safe in initial trials (which involved small, selected populations), but real-world registries uncovered a broad spectrum of rare immune-related adverse events ([37] pmc.ncbi.nlm.nih.gov). Similarly, SGLT-2 inhibitors (diabetes drugs) showed striking reductions in heart failure hospitalizations in trials, but retrospective RWE soon linked them to unusual infections like Fournier's gangrene. A large pharmacovigilance study confirmed a small but statistically significant rise in this serious adverse event in SGLT-2 patients ([38] pmc.ncbi.nlm.nih.gov). The FDA issued warnings based on this RWD finding – a risk never seen under trial conditions because the sample size and follow-up were insufficient ([38] pmc.ncbi.nlm.nih.gov). In practice, HEOR analyses would incorporate these findings by adding the cost and utility decrement of such AEs into long-term models.

Value and Cost Implications: Differences in effectiveness translate directly into cost-effectiveness. On one hand, reduced real-world benefit can make an otherwise promising drug less cost-effective: a therapy that extends life by 12 months in a trial might only add 6 months in practice, halving its cost-per-QALY. On the other hand, RWE can reveal additional benefits or cost offsets. For example, [2] describes an RWD-based cost-effectiveness analysis of insulin therapy in Spain, finding that a combination insulin regimen yielded fewer complications and substantial cost savings compared to standard treatment ([10] pmc.ncbi.nlm.nih.gov). Another RWE study found that using generic antiretrovirals (lamivudine, abacavir, efavirenz) in HIV saved about 25% on drug costs ([10] pmc.ncbi.nlm.nih.gov), a finding not obvious from clinical trial data. Notably, health-economic analyses driven by real-world data have demonstrated huge system-wide savings from biosimilars: switching to biosimilar rituximab or trastuzumab is projected to save tens of millions of euros per year in many countries ([10] pmc.ncbi.nlm.nih.gov). These economic impacts can reinforce or alter formulary decisions.

Meta-Analyses and Systematic Reviews: Beyond individual examples, systematic reviews have attempted to quantify the gap. Bagshaw *et al.* (2021) reviewed 30 meta-analyses and found that about 80% of the time the pooled relative effects from observational studies matched those from RCTs (no significant difference) ([5] pmc.ncbi.nlm.nih.gov). However, in ~20% of cases there were large discrepancies (risk ratios differing by 1.43-fold or more) and occasionally even opposite outcomes ([5] pmc.ncbi.nlm.nih.gov). The authors conclude that when RCTs and RWD disagree, the cause must be investigated – whether it is due to differences in patient mix or simply biases in the observational research. Their review also notes that regulators have been quick to use RWE for safety (rare AEs) and in rare diseases or for long-term outcomes, but remain cautious about RWE for labeling changes unless trials are infeasible.

Collectively, the evidence makes it clear that "theory and practice often diverge" – RCT efficacy is not a guarantee of real-world effectiveness. The extent of the gap varies by setting (e.g., oncology vs chronic disease, new therapy vs existing drugs), but its existence is well-documented ([32] pmc.ncbi.nlm.nih.gov) ([4] pmc.ncbi.nlm.nih.gov). The next section examines **how HEOR uses this evidence** to produce meaningful valuations and predictions.

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

HEOR Approaches for Incorporating Real-World Variance

HEOR spans a range of methods – epidemiologic studies, modeling, comparative effectiveness research – all aimed at answering "how well does this intervention work and at what value in actual practice?" Here we detail key HEOR strategies that account for trial–real-world differences.

Use of Real-World Data in Modeling

One of the most direct ways HEOR bridges the gap is by **infusing models with RWE**. Cost-effectiveness and budget-impact models often start with base-case inputs from RCTs (efficacy, trial-based adverse events) but then adjust for real-world conditions. This includes:

- **Baseline Risk and Disease Progression:** RCT populations often have lower baseline event rates (due to healthier patients or use of background therapies). To correct this, models use observational data to set background risk. For example, ICER analyses commonly draw on registry or claims data for *disease progression rates* – e.g. rate of relapse, hospitalization, or clinical failure – rather than relying solely on trial-based natural history ([8] pmc.ncbi.nlm.nih.gov). Indeed, Lee *et al.* (2021) found that **28.7% of real-world evidence (RWE) inputs** in ICER models concerned disease progression, compared to only ~1.5% informing treatment effectiveness ([8] pmc.ncbi.nlm.nih.gov). This ensures the model reflects how patients actually worsen over time without or with standard therapy, rather than the better-than-average trial experience.

- **Treatment Patterns and Adherence:** Real-world databases reveal treatment persistence and switching patterns. A therapy might show 90% adherence in a trial, but claim data might indicate only 60% of patients persist at one year in practice. Models incorporate these rates to reduce effective exposure in the "real-world" scenario. For example, if an oral drug has a 20% discontinuation rate in claims studies, a model may simulate that proportion of patients ceasing treatment early, affecting long-term outcomes and costs.

- **Utilities and Health-Related Quality of Life:** Trials sometimes include quality-of-life instruments under ideal conditions. HEOR can refine these by using patient surveys and registries to see how actual quality-of-life compares. Patients in routine care often report lower baseline utility (due to comorbidities) and smaller gains than trial participants. Incorporating RWD utilities leads to more conservative QALY gains.

- **Resource Utilization and Costs:** RCTs usually do not capture full costs of care (they often cover only the drug and protocol-defined procedures). RWE provides real cost data – hospitalizations, outpatient visits, co-medications – under usual practice. This is crucial: for instance, if real-world hospitalization rates for a disease are higher than expected, the economic value of a therapy that prevents those hospitalizations is correspondingly higher. Conversely, RWE may reveal higher background use of other expensive services, raising the overall healthcare costs in the model. Apices notes that HEOR teams *"draw on RWE for inputs such as resource use, event rates, and utilities; local calibration of global models to reflect how care is really delivered"* (apices.es). This approach aligns the model with the specific healthcare setting (country, payer, etc.).

- **Long-Term Extrapolations:** When trial follow-up is limited, HEOR often turns to registries and longitudinal cohorts to project long-term outcomes. For example, a pivotal trial may report 2-year survival curves, which alone are insufficient for lifetime modeling. HEOR analysts may tap into national cancer registries to estimate 5-year survival or cure rates. This real-world anchor prevents models from overprojecting the durability of benefit seen in short trials.

A recent systematic review of cost-effectiveness analyses using RWD confirms this trend: RWE is *"recognized as a valuable source of data for market access and reimbursement, and as a complement to clinical trial evidence for treatment pathways, resource use, long-term natural history, and effectiveness."* ([39] pubmed.ncbi.nlm.nih.gov). However, the review also cautioned that RWE brings challenges (confounding, data gaps, quality issues), and that guidance on properly integrating RWE into models is still evolving ([40] pubmed.ncbi.nlm.nih.gov). Indeed, many modeling submissions still default to trial efficacy for the core treatment

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

effect – as [28] found, ICER models *used RWE to inform drug-specific effectiveness only 1.5% of the time* ([8] pmc.ncbi.nlm.nih.gov) – but this situation is changing.

Observational Comparative Effectiveness and Confounding Adjustment

Beyond parameter inputs, HEOR often directly analyzes RWD to estimate comparative effectiveness under real conditions. These observational studies (cohort, registry analyses) attempt to emulate the clinical question of RCTs but in routine practice. The goal is to quantify how well a drug works versus alternatives in "patients like mine." Such studies are pivotal when trials are absent or incomplete.

Comparative effectiveness RWD studies typically use methods like propensity scores, matching, or instrumental variables to adjust for baseline differences ([5] pmc.ncbi.nlm.nih.gov) ([41] pmc.ncbi.nlm.nih.gov). For example, an RWD analysis comparing two anticoagulants may match patients on stroke risk factors so that the treatment and control groups are comparable. These methods aim to mimic randomization but are imperfect; HEOR models account for residual uncertainty by sensitivity analysis. When done properly, these studies can reveal heterogeneity of treatment effect: for instance, [2] cites a real-world study showing that DPP-4 inhibitor linagliptin produced greater hemoglobin-A1c reductions in elderly, Black, and higher-eGFR subgroups than in the overall trial population ([42] pmc.ncbi.nlm.nih.gov). Identifying such subgroup differences helps HEOR to personalize value estimates and guide targeted use (e.g., payers may only cover the drug for groups where benefit was confirmed in RWE).

Importantly, HEOR does not take all observational findings at face value. Variability in RWD study design can produce conflicting results. The comparative review by Hong *et al.* highlights this: roughly 80% of pooled RWE effect estimates agreed with RCTs, but many varied widely, sometimes even reversing sign ([5] pmc.ncbi.nlm.nih.gov). HEOR experts thus critically assess whether discrepancies are plausible. If an observational result contradicts a well-done trial in a biologically implausible way, modellers examine potential biases. Many agencies now require that RWE studies used in submissions follow methodological reporting standards (e.g. the GRACE checklist or STaRT-RWE guidelines) to ensure high quality. When strong RWE for or against a therapy emerges, HEOR can incorporate it by using it to adjust model assumptions or by presenting alternative scenarios (e.g. a base case from trials and a scenario using real-world efficacy).

Pragmatic Trials and Hybrid Studies

A complementary approach is to conduct **pragmatic randomized trials** that sit between explanatory RCTs and pure observational studies. Pragmatic trials relax many of the RCT constraints: they enroll broader patients, allow flexible dosing, and typically measure outcomes using administrative data. The goal is to directly generate high-quality effectiveness data. As Capili and Anastasi (2025) note, pragmatic trials *"are designed to test the effectiveness of interventions in real-world settings"*, providing data that are "directly applicable to a broad patient population" ([43] pmc.ncbi.nlm.nih.gov). In practice, such trials (often cluster-randomized or registry-based) ask whether the drug works under usual care, rather than under ideal conditions.

HEOR embraces pragmatic trials because their results require less extrapolation. For example, if a pragmatic trial shows 5-year survival on a therapy in a typical clinic environment, that endpoint can be plugged more confidently into economic models and budget impact analyses. Moreover, pragmatic trials often include cost or utilization endpoints from the start, explicitly linking efficacy to economic outcomes. While they are still relatively rare outside academia, there is growing interest in pragmatics to inform policy (e.g. in infectious disease or chronic disease management). In many jurisdictions, payers now view pragmatic trial results as highly

relevant evidence. If resources allow, HEOR teams may even sponsor pragmatic extensions or registries post-approval to gather such data prospectively.

Subgroup and Heterogeneity Analyses

Another HEOR strategy is to **identify and incorporate subgroup variation**. Real-world evidence has made it clear that average trial results often hide large differences across patient types. For instance, [25] emphasizes that pragmatic trials or RWD studies are better equipped to measure outcomes in older, frailer patients or those with multiple comorbidities ([3] pmc.ncbi.nlm.nih.gov) ([44] pmc.ncbi.nlm.nih.gov). Using RWD, HEOR analysts can stratify patients by age, biomarker status, baseline severity, etc., and measure comparative effectiveness in each stratum. This can be translated into the model via multiple "cohorts" or health states. For example, if RWE shows a treatment works well only in biomarker-positive patients, the model may apply efficacy only to that subgroup and give lower utility values or none to others. Such heterogeneity analyses better align model predictions with real practice, where decisions are often made on a patient-by-patient basis.

Real-World Safety and Benefit–Risk Assessment

HEOR also systematically incorporates the real-world safety profile. Costs of adverse events are included based on RWE (e.g. healthcare utilization after severe side effects). Furthermore, the benefit-risk balance can shift after launch. A drug that in trials had a 1% risk of a complication might have 5% in practice; this drastically changes both patient outcomes and cost incentives. HEOR studies will often re-evaluate net benefit with updated RWD. In pharmacovigilance, HEOR can perform "quantitative benefit–risk" analyses using RWD: for instance, if an effect on mortality was seen, models can compute how high the event rate can go before the net benefit becomes unfavorable. This was done early in the SGLT2 inhibitor class: once renal outcomes were seen in trials, RWE indicated cardioprotective benefits, and HEOR models quickly incorporated both to assess value across patient profiles.

Case Examples of HEOR Adaptation

To illustrate, Table 2 summarizes several cases where real-world evidence altered the economic or clinical appraisal of a treatment. In each case, HEOR responses included re-estimating models with RWD or conducting new cost-effectiveness analyses based on observed practice patterns.

Case / Drug	Trial Findings	Real-World Findings	HEOR Adaptation
Rofecoxib (NSAID)	Trials (in low-CV-risk arthritis patients) showed pain relief but missed any increase in MI ([35] pmc.ncbi.nlm.nih.gov).	Post-market data revealed a significant uptick in myocardial infarctions and cardiovascular deaths in a broader population ([35] pmc.ncbi.nlm.nih.gov).	Real-world pharmacoepidemiology studies prompted label warnings/withdrawal. HEOR models (re)calculated cost-effectiveness by including observed CV safety risks, which rendered the drug's value negative.
SGLT-2 Inhibitors	Clinical trials (EMPA-REG, CANVAS) demonstrated reduced heart failure and renal events with no rare AEs.	RWD identified rare but serious AEs (e.g. Fournier's gangrene urinary infection) not seen in trials ([38] pmc.ncbi.nlm.nih.gov).	HEOR models updated safety protocols: added the cost and disutility of identified rare AEs based on claims data. Value models also confirmed cardiovascular benefits at

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

Case / Drug	Trial Findings	Real-World Findings	HEOR Adaptation
			population level using registry outcomes.
Multiple Myeloma (Lenalidomide/Pomalidomide regimens)	Phase III trials in selected relapsed patients reported median PFS of ~30–37 months on newer regimens.	A population cohort study found real-world PFS ~7–18 months shorter and OS substantially lower, especially in relapsed disease ([4] pmc.ncbi.nlm.nih.gov).	HEOR incorporated RWD-based hazard ratios. Cost-effectiveness analyses were re-run using real-world survival curves, yielding lower QALYs. Sensitivity analyses were conducted using both RCT and RWD estimates to bracket uncertainty.
Remdesivir (COVID-19)	An RCT (ACTT-1) showed faster time-to-recovery and a trend to lower mortality versus placebo.	A large multicenter observational cohort (with propensity matching) found similar improvements: 74% recovery vs 59% on standard care, and reduced 14-day mortality ([45] pmc.ncbi.nlm.nih.gov).	This concordance gave confidence to HEOR teams to use the trial efficacy in models; indeed, analyses for remdesivir's cost-effectiveness were done using the combined RCT/RWD evidence. RWE was also used to validate assumptions about length-of-stay reductions when projecting hospital costs.
Generic HIV Therapy	Trials showed virologic control with branded regimens, but did not address costs.	Real-world claims analyses found that switching to generic lamivudine, abacavir, efavirenz cut drug costs by ~25% without loss of efficacy ([10] pmc.ncbi.nlm.nih.gov).	HEOR budget-impact models for HIV now apply these real-world cost savings. Value assessments incorporate the lower price of generics observed in claims data, dramatically improving the estimated cost-effectiveness of HIV therapy.
Oncology Biosimilars	Originator monoclonal antibodies showed proven efficacy in trials.	Registry and administrative data demonstrated that using approved biosimilars could replicate outcomes and save millions. One analysis estimated €4.9–120 million saved per treatment course of rituximab or trastuzumab ([10] pmc.ncbi.nlm.nih.gov).	HEOR and policy models now include biosimilar pricing: cost-effectiveness models of cancer regimens plug in market prices derived from RWD (claims/health authority reports) for biosimilars, often justifying broader use of these cheaper alternatives.

Table 2. Illustrative cases where real-world experience diverged from trial data, and how HEOR responded. (Sources as cited.)

These examples highlight how HEOR must adapt: whether by recalibrating value models to real-world risks, performing new analyses that include observed costs, or recommending different treatment strategies based on heterogeneous effectiveness. In practice, HEOR teams often present both "efficacy-based" and "effectiveness-based" scenarios in dossiers to payers, enabling stakeholders to see the impact of the trial–practice gap on conclusions.

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

Data Analysis: Evidence Integration and Outcomes

Beyond single examples, HEOR involves rigorous data analysis to quantify the impact of trial–real-world variance. This includes meta-analyses, systematic reviews, and modeling efforts that synthesize multiple sources.

For instance, the systematic review by Hong *et al.* quantified treatment effect ratios between observational studies and RCTs. Across 74 comparisons, about 80% showed overlapping confidence intervals (no significant difference in relative effect) ([5] pmc.ncbi.nlm.nih.gov). However, 20% were meaningfully different – often because the observational cohorts included patients or contexts quite unlike any trial. Crucially, this review emphasized understanding *why* discrepancies occur, recommending further research into confounders versus true population effects. In HEOR terms, this translates into doing meta-regression or sensitivity analyses. If modeling a new oncology drug, for example, analysts might use a Bayesian framework or probabilistic sensitivity analysis to "shrink" RCT-based estimates toward RWD-based priors, reflecting uncertainty about external validity.

Administrative claims and EHR databases are another cornerstone. HEOR analysts routinely mine these sources to **validate model cohorts**. Before finalizing a cost-effectiveness model, one might query a large insurance database for patients meeting the inclusion criteria to check how many would have been excluded (e.g. how many have renal insufficiency or are outside the age range). If the trial would have enrolled only 10% of real patients with the disease, that signals a major external validity concern. Sensitivity analyses can then apply lower efficacy to account for the additional risk factors seen in practice. Moreover, observational *phase IV* studies (sometimes conducted by manufacturers or academic groups) can measure the drug's effectiveness with real practice patterns; these results feed directly into HEOR. For example, when a drug has an accelerated approval based on surrogate endpoints, HEOR teams track subsequent real-world studies to update long-term outcome estimates.

Advanced statistical techniques aid this evidence integration. Modern HEOR often employs machine learning algorithms to extract insights from RWD. For example, natural language processing can identify patient characteristics from free-text notes, enabling more accurate patient matching. Causal inference methods (e.g. **targeted maximum likelihood estimation** or machine-learning-powered propensity scores) are used to adjust for confounding in RWD studies (apices.es). These methods improve confidence that RWE estimates are valid. Furthermore, simulation methods like discrete-event simulation or microsimulation can represent patient heterogeneity explicitly, drawing input distributions from RWD. For treatments with genetic or biomarker targets, real-world prevalence of those markers (often gleaned from genomic datasets) is used to weight model subgroups. All this sophisticated analytics helps bridge the efficacy vs effectiveness divide quantitatively.

The U.S. -based ICER provides a concrete example of this data-driven HEOR. Lee *et al.* (2021) analyzed ICER's economic models from 2014–2019 and reported that on average about 33% of all model inputs came from RWE ([8] pmc.ncbi.nlm.nih.gov). However, the types of inputs differed greatly: nearly 29% of RWE inputs were used for **disease progression** and about 21% for **resource use and costs** ([8] pmc.ncbi.nlm.nih.gov), whereas only ~1–2% of inputs for effectiveness or adverse events came from RWE. This shows a methodological pattern: in the ICER framework, RWE mainly modifies the context (how fast disease worsens, how much healthcare is consumed) rather than the drug's intrinsic potency. Nonetheless, ICER and other HTA groups *acknowledge* that RWE is increasingly expected. Indeed, draft guidance from the FDA now allows manufacturers to present RWE in payer communications, and HTA guidelines explicitly recommend using RWE "to improve evidence synthesis" when RCTs are incomplete ([46] pmc.ncbi.nlm.nih.gov).

Despite these advances, HEOR analysts must remain cautious. Bias and data quality issues in RWD can mislead models if unaddressed ([40] pubmed.ncbi.nlm.nih.gov). Key weaknesses include incomplete records, coding errors, and confounders. To mitigate this, HEOR practice now often involves *triangulation* of evidence: checking RWE against multiple sources (e.g. registries, claims, even external control arm construction) and requiring transparency in assumptions. Structured reporting guidelines (CHEERS for economic evaluation, STROBE for observational studies) help ensure that RWD-derived analyses are trustworthy enough to drive policy.

Multiple Perspectives on the Variance

Understanding why real-world drug performance diverges from trials requires perspectives from all stakeholders.

- **Clinicians** emphasize *patient heterogeneity and practice variability*. They note that trial patients "are not likely identical" to those seen day-to-day ([20] pmc.ncbi.nlm.nih.gov). Clinicians also critically assess whether trial protocols (e.g. dosing schedules or monitoring frequency) can be replicated. As Zarbin (2019) comments, clinicians worry if trial findings are **reproducible** with diverse patients ([47] pmc.ncbi.nlm.nih.gov). Practicing doctors also value RWE that addresses practical questions: does this drug improve the outcomes I care about, like hospitalization rates or function? If RCTs suggest a physiological benefit but an RWE study shows modest real-world effectiveness, physicians may adjust their prescribing or patient selection.

- **Patients and Advocacy Groups** often push for inclusive evidence. They point out that trials exclude many real patients (e.g. elderly or minorities), and ask whether the drug will work for "people like me." In response, regulators and HEOR have begun to emphasize subgroup analyses and broader trial designs to capture diverse experiences. Moreover, through patient-reported outcomes (PROMs), RWD sometimes reveals that a treatment's quality-of-life impact is smaller (or larger) than expected. The trend toward patient-centric endpoints in trials (PROMs, daily functioning) reflects this perspective.

- **Regulators** historically focused on RCTs for approval, but are increasingly open to RWE. Agencies have funded initiatives and issued guidance on when and how real-world data can support labeling or post-market requirements ([15] pmc.ncbi.nlm.nih.gov) ([48] pmc.ncbi.nlm.nih.gov). They often require risk management plans and real-world safety monitoring precisely because they understand trials might miss some effects ([7] pmc.ncbi.nlm.nih.gov) ([23] pmc.ncbi.nlm.nih.gov). However, regulators still scrutinize RWE quality heavily. The FDA's RWE Program (2018) and EMA's RWE roadmap both highlight the need for "regulatory-grade" observational research, with clear standards for design and analysis. For example, the FDA now explicitly allows RWE to justify new indication approvals (as it did with palbociclib for male breast cancer) ([16] pmc.ncbi.nlm.nih.gov), but only if the data are reliable. In summary, regulators aim to *merge* trial and real-world insights: they use RCTs for definitive efficacy, but expect RWE to fill gaps in populations, duration, or contexts that trials couldn't cover.

- **Payers and Health Technology Assessors (HTA)** are highly sensitive to the trial–practice gap because it directly affects coverage and spending. For payers, the ultimate question is *"What will happen to my costs and patient outcomes if we pay for this drug?"* If RWE suggests lower effectiveness, payers may demand price discounts or coverage restrictions. Indeed, outcomes-based contracting is one mechanism explicitly designed to share risk: pharmaceutical companies agree to rebates or price adjustments based on real-world performance (e.g. achieving certain clinical endpoints in the covered population) ([49] www.pharmacytimes.com). A Pharmacy Times analysis highlights that *"these conversations often require real-world evidence, which are data and studies that show how (and whether) a particular drug performs in typical patients and usual care settings."* ([50] www.pharmacytimes.com). Payers have begun including conditional coverage (e.g. pay only if follow-up registry shows benefit) to manage uncertainty. Surveys (e.g. of U.S. payers) indicate strong support for RWE: most respondents believe health decisions should be informed by RWE alongside RCTs ([51] pmc.ncbi.nlm.nih.gov). HTA bodies like NICE incorporate RWE into their appraisals, especially for real-world resource use and when trial evidence is immature (www.nice.org.uk) ([52] pmc.ncbi.nlm.nih.gov). For example, NICE's RWE framework explicitly states that RWD can "improve our understanding of … the effects of interventions on patient and system outcomes in routine settings" (www.nice.org.uk).

IntuitionLabs   IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

- **Industry (Pharma/Biotech)** recognizes the disparity as both a challenge and an opportunity. New drug programs increasingly plan for RWE generation from the outset. HEOR and RWE teams collaborate on early modeling: before launch, companies use RWD to refine trial eligibility or pricing scenarios (as highlighted by Apices) (apices.es). Post-approval, firms invest in disease registries, claims analyses, and even pragmatic trials to demonstrate value in practice. The goal is to sustain a drug's market access by proving cost-effectiveness with real patients. When RWE reveals a positive surprise (e.g. off-label benefits), companies may seek indication expansions via real-world studies. Conversely, if RWE uncovers hidden risks, companies often revise risk-management plans or explore narrower labeling. Notably, legislation like the 21st Century Cures Act has empowered companies to engage payers by discussing RWE and HEOR data beyond the label ([53] www.pharmacytimes.com). This regulatory change, coupled with negative press about pricing, has pushed firms to use RWE proactively in negotiating formularies and outcomes contracts.

In sum, **all stakeholders** are converging on the importance of real-world performance. Clinicians and patients want evidence applicable to everyday care; payers demand proof of value in their populations; and regulators require RWE to ensure safety and generalizability. HEOR acts as the integrative bridge among these viewpoints, supplying the analytic methods and evidence synthesis to reconcile trial results with practice realities.

Discussion: Implications and Future Directions

The gap between trial efficacy and real-world effectiveness has profound implications for healthcare decision-making:

- **Access and Coverage:** If drugs underperform outside trials, payers may restrict coverage or require risk-sharing schemes. Conversely, robust RWE showing strong real-world benefits can justify broader access or higher valuations. For patients, this means that trial excitement may translate into delayed or conditional access depending on emerging real-world data. Industry must therefore craft launch strategies that include RWE commitments.

- **Pricing and Value-Based Models:** Discrepancies lead to interest in value-based pricing. Payers increasingly tie payment to real-world outcomes, effectively making price a function of the observed effectiveness ([49] www.pharmacytimes.com). Example: oncology drugs may have formulary agreements where a refund is triggered if patients do not achieve a survival benchmark in the first year. These models rely on accurate RWE measurement. In the future, *indication-based pricing* may also grow: the same drug could be priced differently for different conditions based on real-world performance in each indication, as Leonard observed ([54] www.pharmacytimes.com).

- **Methodological Evolution:** The need to learn from RWD is driving innovations in study design. We have seen pragmatic trials become more common, and regulators offering clear pathways for RWE acceptability. Tools like the PRECIS-2 instrument help investigators design trials along the explanatory–pragmatic continuum ([30] pmc.ncbi.nlm.nih.gov). Statistical advances (propensity methods, emulation frameworks) are being standardized by communities like ISPOR/ICMJE. Looking ahead, the integration of AI into HEOR workflows (as noted by Apices) promises to accelerate data curation and confounder adjustment (apices.es). Machine learning models are already being tested for predicting individual patient trajectories from complex RWD.

- **Data Infrastructure Expansion:** Healthcare systems worldwide are building the infrastructure for RWE. Initiatives like nationwide registries (for stroke, cancer, heart disease) and data linkages (EHR networks, claims) will feed more real-time evidence. Notably, digital health (wearables, smartphones) is creating continuous data streams (apices.es). Wearable sensors can track blood pressure, glucose, activity and more; linking these data to conventional RWD could transform outcome measurement (e.g. dosing effects on daily mobility). HEOR will incorporate such digital endpoints into models, enabling evaluation of interventions on "time-out-of-hospital" or "functional status" in ways trials rarely measure.

- **Heterogeneity and Personalized Policies:** As we gather RWE on diverse populations, HEOR can support more personalized coverage. If RWD shows that only patients with biomarker X truly benefit from a drug, payers might restrict payment to that subgroup. This aligns with the shift toward precision medicine. In some cases, HEOR might even advocate for "coverage with evidence development" (CED) – allowing access while collecting RWE in a systematic registry to update assessments.

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

- **Regulatory and HTA Guidelines:** Agencies are actively updating guidance. The NICE RWE framework (2022) is one example; in the EU, EMA and national HTA bodies are drafting RWE standards for reimbursement submissions. It is likely that in the future, health economics evaluation sections will routinely be expected to include RWE scenarios. Failures to consider real-world variation may be viewed as incomplete analyses.

- **Limitations and Cautions:** Recognizing the value of RWE does not mean trials will be supplanted. The gold-standard evidence hierarchy remains, especially for internal validity. RWE studies can suffer from data quality gaps and bias. HEOR continues to emphasize process rigor: for instance, NICE and ISPOR both articulate standards for using observational data in comparative effectiveness, including transparency about data provenance and sensitivity analyses ([46] pmc.ncbi.nlm.nih.gov) (www.nice.org.uk). In the longer term, methodological research is needed to quantify when purely trial-based models lead us astray and how best to combine evidence (e.g. Bayesian meta-models, hierarchical blending).

In conclusion, the divergence between trial and real-world performance is both a challenge and an opportunity. For patients and payers, it highlights that **clinical trial results are only part of the story**. For researchers and policy-makers, it means embracing a mixed-evidence paradigm: using RCTs for internal validity and RWE for context and continuity. HEOR – through advanced analytics, economic modeling, and outcome research – provides the toolkit for integrating these sources. As data ecosystems grow and analytic methods improve, our ability to predict a drug's **true** value and optimize its use will continue to evolve.

Conclusion

Drugs do often perform differently in routine care than in the tightly controlled environment of clinical trials. This report has explored the many facets of this phenomenon. We have shown that differences in patient selection, adherence, treatment administration, follow-up, and outcome measurement – along with statistical issues of bias – all contribute to what is known as the efficacy–effectiveness gap ([1] pmc.ncbi.nlm.nih.gov) ([2] pmc.ncbi.nlm.nih.gov). Empirical studies across diseases consistently document that real-world outcomes may be less favorable than trial results (though sometimes better in certain subgroups) ([4] pmc.ncbi.nlm.nih.gov) ([32] pmc.ncbi.nlm.nih.gov). These gaps are not merely academic: they have real implications for safety (leading to new FDA warnings), clinical guidelines, and especially for economic value.

Health Economics and Outcomes Research is explicitly designed to address this gap. By leveraging real-world data – from electronic health records, claims, registries, and even digital devices – HEOR brings the "world outside trials" into analyses. Through modeling adjustments, pragmatic studies, comparative-effectiveness research, and advanced statistics, HEOR strategies recalibrate trial findings to reflect actual practice. Economic models now routinely incorporate RWE for baseline risks, costs, and patient heterogeneity (apices.es) ([8] pmc.ncbi.nlm.nih.gov), and they often present scenarios with both trial-based and real-world-based efficacy estimates. In doing so, HEOR provides a more complete picture of a drug's benefits, risks, and value.

Looking ahead, the use of real-world evidence in health decision-making will only grow. International regulatory agencies (FDA, EMA, NICE, etc.) have signaled a clear intent to institutionalize RWE in approval and reimbursement processes ([15] pmc.ncbi.nlm.nih.gov) (www.nice.org.uk). Technological advances (integrated data platforms, AI analytics) will make the synthesis of trial and real data more seamless. The key will be maintaining scientific rigor: ensuring that RWE methods are robust and that conclusions drawn are reproducible and fair.

In the end, bridging the trial–real-world divide demands a cultural shift in evidence assessment. Stakeholders must view RCTs and RWE as complementary: *efficacy* data from trials should be understood in context by *effectiveness* data from observation. HEOR stands at that interface – tasked with translating efficacy into effectiveness in a way that guides policy, optimizes patient care, and ensures sustainable healthcare value.

References

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

- Chen D. *Real-world studies: bridging the gap between trial-assessed efficacy and routine care*. J Biomed Res. 2022;36(3):147–154. ([1] pmc.ncbi.nlm.nih.gov) ([7] pmc.ncbi.nlm.nih.gov)

- Visram A et al. *Comparing the clinical trial efficacy versus real-world effectiveness of treatments for multiple myeloma: a population-based study*. Haematologica. 110(1):228–233, 2024. ([4] pmc.ncbi.nlm.nih.gov) ([18] pmc.ncbi.nlm.nih.gov)

- Lee WJ, et al. *Use of real-world evidence in economic assessments of pharmaceuticals in the United States*. J Manag Care Spec Pharm. 2021;27(1):105–115. ([8] pmc.ncbi.nlm.nih.gov) ([55] pmc.ncbi.nlm.nih.gov)

- Real-world Evidence and Health Economics and Outcomes Research: From Nice to Have to Non-negotiable. 2024; Apices white paper. (apices.es) (apices.es)

- Zarbin MA. *Real Life Outcomes vs. Clinical Trial Results*. J Ophthalmic Vis Res. 2019;14(1):88–92. ([6] pmc.ncbi.nlm.nih.gov) ([56] pmc.ncbi.nlm.nih.gov)

- Ruggeri M, et al. *Real-world studies in HIT* (NMJ synonyms)   med device - guidelines (2021). (discussing FDA guidance on RWE) ([15] pmc.ncbi.nlm.nih.gov).

- NICE. *Introduction to real-world evidence in NICE decision making: NICE RWE framework*. NICE Corporate, 2022. (www.nice.org.uk) (www.nice.org.uk)

- Pipeline Complex: *Pragmatic Clinical Trials and Real-World Evidence: An Introduction*. Am J Nurs. 2025;125(2):56–58. ([43] pmc.ncbi.nlm.nih.gov) ([57] pmc.ncbi.nlm.nih.gov)

- Marzano L, et al. *Exploring the discrepancies between clinical trials and real-world data: A small-cell lung cancer study*. Clin Transl Sci. 2024;17(8):e13909. ([29] pmc.ncbi.nlm.nih.gov) ([58] pmc.ncbi.nlm.nih.gov)

- Hong YD, et al. *Comparative effectiveness and safety of pharmaceuticals assessed in observational studies compared with RCTs*. BMC Med. 2021;19:307. ([5] pmc.ncbi.nlm.nih.gov) ([59] pmc.ncbi.nlm.nih.gov)

- Parody-Rúa E, et al. *Economic evaluations informed exclusively by real-world data: A systematic review*. Int J Environ Res Public Health. 2020;17(4):1171. ([60] www.mdpi.com)

- Leonard D. *Value Depends on Real-World Evidence*. Pharm Times. 2017. ([50] www.pharmacytimes.com) ([49] www.pharmacytimes.com)

- Chou R, et al. *FDA real-world evidence framework*. Clin Pharmacol Ther. 2020;107(4):843–852. (guidance document) ([15] pmc.ncbi.nlm.nih.gov)

- Other relevant sources as cited above (which include FDA/EMA guidelines and domain-specific RWE studies).

External Sources

[1] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:use,7...

[2] https://pmc.ncbi.nlm.nih.gov/articles/PMC11902901/#:~:metic...

[3] https://pmc.ncbi.nlm.nih.gov/articles/PMC11902901/#:~:PCTs%...

[4] https://pmc.ncbi.nlm.nih.gov/articles/PMC11694101/#:~:The%2...

[5] https://pmc.ncbi.nlm.nih.gov/articles/PMC8647453/#:~:Overa...

[6] https://pmc.ncbi.nlm.nih.gov/articles/PMC6388532/#:~:Appli...

[7] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:endpo...

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

[8] https://pmc.ncbi.nlm.nih.gov/articles/PMC10390921/#:~:We%20...

[9] https://pmc.ncbi.nlm.nih.gov/articles/PMC10390921/#:~:RWE%2...

[10] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:affor...

[11] https://pmc.ncbi.nlm.nih.gov/articles/PMC8799209/#:~:In%20...

[12] https://pmc.ncbi.nlm.nih.gov/articles/PMC8647453/#:~:There...

[13] https://pmc.ncbi.nlm.nih.gov/articles/PMC6388532/#:~:consi...

[14] https://pubmed.ncbi.nlm.nih.gov/26797239/#:~:Backg...

[15] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:Havin...

[16] https://pmc.ncbi.nlm.nih.gov/articles/PMC8799209/#:~:Cures...

[17] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:2018,...

[18] https://pmc.ncbi.nlm.nih.gov/articles/PMC11694101/#:~:for%2...

[19] https://pmc.ncbi.nlm.nih.gov/articles/PMC11694101/#:~:basel...

[20] https://pmc.ncbi.nlm.nih.gov/articles/PMC6388532/#:~:Repli...

[21] https://pmc.ncbi.nlm.nih.gov/articles/PMC11694101/#:~:for%2...

[22] https://pmc.ncbi.nlm.nih.gov/articles/PMC11306525/#:~:poten...

[23] https://pmc.ncbi.nlm.nih.gov/articles/PMC6388532/#:~:the%2...

[24] https://pmc.ncbi.nlm.nih.gov/articles/PMC8647453/#:~:Healt...

[25] https://pubmed.ncbi.nlm.nih.gov/26797239/#:~:Resul...

[26] https://pubmed.ncbi.nlm.nih.gov/26797239/#:~:Concl...

[27] https://pmc.ncbi.nlm.nih.gov/articles/PMC6388532/#:~:repli...

[28] https://pmc.ncbi.nlm.nih.gov/articles/PMC11306525/#:~:is%20...

[29] https://pmc.ncbi.nlm.nih.gov/articles/PMC11306525/#:~:the%2...

[30] https://pmc.ncbi.nlm.nih.gov/articles/PMC11902901/#:~:While...

[31] https://pmc.ncbi.nlm.nih.gov/articles/PMC11694101/#:~:cohor...

[32] https://pmc.ncbi.nlm.nih.gov/articles/PMC8799209/#:~:Two%2...

[33] https://pmc.ncbi.nlm.nih.gov/articles/PMC8799209/#:~:treat...

[34] https://pmc.ncbi.nlm.nih.gov/articles/PMC11694101/#:~:from%...

[35] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:fore....

[36] https://pmc.ncbi.nlm.nih.gov/articles/PMC10390921/#:~:,info...

[37] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:The%2...

[38] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:Anoth...

[39] https://pubmed.ncbi.nlm.nih.gov/30816067/#:~:impos...

[40] https://pubmed.ncbi.nlm.nih.gov/30816067/#:~:long,...

[41] https://pmc.ncbi.nlm.nih.gov/articles/PMC11306525/#:~:The%2...

[42] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:Simil...

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

[43] https://pmc.ncbi.nlm.nih.gov/articles/PMC11902901/#:~:Pragm...

[44] https://pmc.ncbi.nlm.nih.gov/articles/PMC11902901/#:~:%28Ho...

[45] https://pmc.ncbi.nlm.nih.gov/articles/PMC9179114/#:~:In%20...

[46] https://pmc.ncbi.nlm.nih.gov/articles/PMC10390921/#:~:A%20r...

[47] https://pmc.ncbi.nlm.nih.gov/articles/PMC6388532/#:~:Repro...

[48] https://pmc.ncbi.nlm.nih.gov/articles/PMC8799209/#:~:envir...

[49] https://www.pharmacytimes.com/view/value-depends-on-realworld-evidence#:~:Payer...

[50] https://www.pharmacytimes.com/view/value-depends-on-realworld-evidence#:~:with%...

[51] https://pmc.ncbi.nlm.nih.gov/articles/PMC10390921/#:~:A%20r...

[52] https://pmc.ncbi.nlm.nih.gov/articles/PMC10390921/#:~:for%2...

[53] https://www.pharmacytimes.com/view/value-depends-on-realworld-evidence#:~:The%2...

[54] https://www.pharmacytimes.com/view/value-depends-on-realworld-evidence#:~:70,2...

[55] https://pmc.ncbi.nlm.nih.gov/articles/PMC10390921/#:~:A%20r...

[56] https://pmc.ncbi.nlm.nih.gov/articles/PMC6388532/#:~:Repro...

[57] https://pmc.ncbi.nlm.nih.gov/articles/PMC11902901/#:~:metic...

[58] https://pmc.ncbi.nlm.nih.gov/articles/PMC11306525/#:~:INTRO...

[59] https://pmc.ncbi.nlm.nih.gov/articles/PMC8647453/#:~:about...

[60] https://www.mdpi.com/1660-4601/17/4/1171#:~:throu...

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

IntuitionLabs - Industry Leadership & Services

North America's #1 AI Software Development Firm for Pharmaceutical & Biotech: IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

Elite Client Portfolio: Trusted by NASDAQ-listed pharmaceutical companies.

Regulatory Excellence: Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

Founder Excellence: Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

Custom AI Software Development: Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

Private AI Infrastructure: Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

Document Processing Systems: Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

Custom CRM Development: Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

AI Chatbot Development: Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

Custom ERP Development: Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

Big Data & Analytics: Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

Dashboard & Visualization: Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

AI Consulting & Training: Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at https://intuitionlabs.ai/contact for a consultation.

IntuitionLabs - AI Software for Pharma & Biotech

Efficacy vs. Effectiveness: Why Trial Results Vary in Practice

Page 20 of 20

DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by Adrien Laurent, a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.