Advancing clinical development and evidence generation through causal AI

A targeted review of applications and implications for regulatory-grade evidence

Komaleshwari Rani, MSc[1],
Jackie Vanderpuye-Orgle, MSc, PhD[2],
Ashley E. Tate, PhD[3],
Rafael Coderch Lanau, Sr., MSc[4].

[1]Parexel International, London, United Kingdom, [2]Parexel International, Durham, NC, USA, [3]Parexel International, Amsterdam, Netherlands, [4]Parexel International, Austin, TX, USA.

MSR85

Background

Causal AI estimates cause-and-effect relationships rather than associations and it is a fundamental distinction for regulatory decision-making. Despite growing interest among regulators, sponsors, and HTA bodies, the conditions under which causal AI methods have been accepted as regulatory-grade evidence remain ill-defined and a targeted assessment is therefore needed. [1,5]

This targeted review aims to:

1. Map the landscape of causal AI applications across clinical development and real-world evidence generation, with primary focus on FDA and supplementary coverage of EMA, NICE, and Health Canada/CADTH.
2. Characterize the conditions under which causal AI methods have been accepted as regulatory-grade evidence, distinguishing between supportive and pivotal evidentiary role.
3. Identify documented cases where causal methods demonstrably influenced regulatory decisions.
4. Highlight transparency requirements, qualification challenges, and opportunities for broader regulatory adoption.

Methods

Design: Targeted review mapping the regulatory acceptance landscape of defined causal AI methods.

Methodological scope

- Tier 1: Propensity scoring, IPTW, instrumental variables, g-computation, marginal structural models.
- Tier 2: Double ML, Bayesian causal borrowing, target trial emulation, TMLE.
- Tier 3: DAGs, structural causal models, estimand framework.

Sources: Regulatory-authored peer-reviewed publications · FDA, ICH, NICE, and Health Canada/CADTH guidance documents [9,10,11,12] · Pharmacoepidemiology and biostatistics literature · Consensus reporting standards.

Inclusion: Explicit application of a defined causal AI method to a regulatory submission, HTA assessment, or clinical development context. Review period 2016–2025.

Scope limitation: This review maps regulatory acceptance conditions. It does not assess statistical validity of individual analyses.

Results

Tier 1: Conditional acceptance documented by regulatory body

- Five FDA oncology cases confirm propensity score and IPTW methods accepted as supportive evidence. [1,5]

Table 1

Drug	Method	Regulatory purpose	Status
Selumetinib (NF1)	Historical comparison	Natural history	Supportive: Served as the primary control group
Erdafitinib (urothelial)	PS weighting	Natural history	Supportive: Accepted but noted data gaps.
Pembrolizumab + lenvatinib	PS cross-trial	Contribution of effect	Exploratory: Provided insights, did not drive approval.
RCC combinations	Post-hoc PS matching	Component isolation	Exploratory: Post-hoc analysis for treatment context
Blinatumomab (ALL)	IPTW retrospective	Comparative efficacy	Supportive: Confirmed the efficacy gap vs. history

No Tier 1 method has achieved pivotal regulatory status; all documented cases remain supportive evidence. [1]

Tier 2: Methodological readiness without regulatory precedent

- Bayesian causal borrowing with propensity score weighting shows strong performance in augmenting trial control arms with real-world data, with adaptive component-wise borrowing reducing heterogeneity-related bias. [3]
- Target trial emulation provides a structured framework for designing observational studies that estimate causal intervention effects, with standardized protocols now established. [6,7]
- FDA's 2023 draft guidance on externally controlled trials explicitly references target trial emulation and causal frameworks including TMLE and G-methods as acceptable analytical approaches, requiring pre-specified protocols submitted for review. [9]

No Tier 2 method has been explicitly documented as accepted in a regulatory submission to date. [1]

Tier 3: Foundational infrastructure

- DAGs and structural causal models provide assumption identification infrastructure enabling credible Tier 1 and Tier 2 analyses. [2]
- ICH E9(R1) estimand framework, adopted by FDA (2021) and EMA (2020), establishes de facto requirements for explicit causal estimand specification and pre-specified sensitivity analyses in clinical submissions. [10]

Cases with limited transparency were consistently relegated to exploratory or contextual roles regardless of method sophistication. [1]

Table 2

.

Transparency Dimension	Regulatory Requirement	Acceptance Impact
Assumption transparency	Explicit DAG or causal diagram	Determines identifiability
Analytical transparency	Pre-specified protocol and SAP	Enables formal comparison
Reporting transparency	Full methodological disclosure	Enables independent evaluation
Data transparency	Source limitations characterized	Determines fit-for-purpose

Conclusions

- Causal AI methods have achieved a defined but limited regulatory role and accepted as supportive evidence across documented FDA oncology cases, but not yet as pivotal evidence. [1]
- The primary barrier is methodological transparency rather than sophistication. Acceptance is determined by pre-specification, explicit documentation of causal assumptions, and fit-for-purpose data quality. [1,4,9]
- Broader adoption requires formal qualification pathways. FDA draft guidance [9], TARGET Statement [4], NICE RWE Framework [11], and Health Canada/CADTH guidance [12] collectively signal international convergence on transparency as the operative standard.
- With these foundations in place, causal AI has the potential to become central to regulatory-grade evidence generation across clinical development and real-world evidence. [2,6]

REFERENCES

[1] Mishra-Kalyani PS, et al. Ann Oncol. 2022;33(4):376-383
[2] Michoel T, Zhang JD. Drug Discov Today. 2023;28(10):103737
[3] Das AC, et al. arXiv preprint arXiv:2601.03480. 2026
[4] Cashin AG, Hernán MA, et al. BMJ. 2025. doi:10.1136/bmj-2025-087179
[5] Thorlund K, et al. Clin Epidemiol. 2020;12:457-467
[6] Fu EL. J Am Soc Nephrol. 2023;34(8):1305-1314
[7] Hernán MA, Wang W, Leaf DE. JAMA. 2022;328(24):2446-2447
[8] Ren J, et al. BMC Med Res Methodol. 2023;23(1):18
[9] FDA. Externally Controlled Trials. Draft Guidance. Feb 2023
[10] ICH. E9(R1): Estimands. November 2019
[11] NICE. RWE Framework. ECD9. June 2022
[12] CADTH/Health Canada. Guidance for Reporting RWE. May 2023