



Real-World Data and Causal Machine Learning to Enhance Drug Development

Chris Anagnostopoulos^{1,2} · Mihaela Van Der Schaar^{3,4} · Jean-Paul Collet^{5,6} · Ramon Hernandez Vecino⁷

Received: 3 April 2025 / Accepted: 25 July 2025
© The Author(s) 2025

Abstract

The current paradigm of clinical drug development, which predominantly relies on traditional randomized controlled trials (RCTs), is increasingly challenged by inefficiencies, escalating costs, and limited generalizability. Concurrent advancements in biomedical research, big data analytics, and artificial intelligence have allowed for the integration of real-world data (RWD) with causal machine learning (CML) techniques to address some of these limitations. This manuscript reviews the emerging role of RWD/CML in enhancing clinical research and drug development programs. By leveraging diverse data sources — including electronic health records, wearable devices, and patient registries — CML methods facilitate robust drug effect estimation, enable precise identification of responders, and support adaptive trial designs. Approaches such as advanced propensity score modelling, outcome regression, and Bayesian inference can help mitigate confounding and biases inherent in observational data, thereby strengthening the validity of causal inference. However, these innovative methodologies also face significant challenges related to data quality, computational scalability, and the absence of standardized validation protocols. Furthermore, ethical and regulatory concerns regarding model transparency and validity, data privacy, and possible algorithmic biases stress the importance of multidisciplinary collaboration and rigorous oversight. Our analysis underscores that while RWD/CML integration can enhance clinical development programs by generating more comprehensive evidence and accelerating drug innovation, its successful adoption depends on overcoming technical, operational, and scientific hurdles while maintaining a transparent approach with regulatory agencies.

Introduction

Drug development is complex, time-intensive, and costly, requiring rigorous testing to ensure efficacy and safety. The process begins with identifying promising molecules, followed by extensive clinical trials and strict regulatory

oversight from agencies like the FDA and EMA. On average, it takes 10–13 years to bring a drug to market, with only 1 in 10,000 candidates gaining approval [1–3]. Development costs range from \$1–2.3 billion [3–5], contributing to a decline in return-on-investment from 10.1% in 2010 to 1.8% in 2019 [5]. These challenges have driven efforts to streamline molecule selection and improve trial efficiency.

Randomized clinical trials (RCTs) remain the gold standard for evaluating safety and efficacy but have limitations. Phase 3 trials, despite larger samples, still struggle with diversity [6], underrepresentation of high-risk patients [7], and potential overestimation of effectiveness due to controlled conditions [8]. Small sample sizes hinder subgroup analyses and rare adverse event detection. Randomization does not guarantee perfect covariate balance [9, 10], and differential loss to follow-up may introduce biases [11]. Additionally, reliance on surrogate endpoints, such as progression free survival over overall survival, is common—70% of recent FDA oncology approvals used non-overall survival endpoints—raising concerns about real-world relevance [12].

✉ Chris Anagnostopoulos
chris.anagnostopoulos@quantumblack.com

¹ QuantumBlack, The Post Building, 100 Museum St, Holborn, London, UK

² Statistics Section, Imperial College London, London, UK

³ Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK

⁴ The Alan Turing Institute, London, UK

⁵ Evidinno Outcomes Research Inc., Vancouver, Canada

⁶ University of British Columbia (UBC), Vancouver, Canada

⁷ Department of Research and Development, Sanofi, Gentilly, France

Similar challenges apply to assessing long-term side effects and competing risks in patients with comorbidities [13].

To address the limitations of conventional trials, researchers have explored strategies such as decentralized clinical trials, master protocols (basket, umbrella, and platform trials), adaptive designs, patient-centred trials, and the integration of real-world data (RWD) to complement trials for drug development. For instance, RWD captures patient journeys, disease progression, and treatment responses, offering valuable insights into drug efficacy and safety beyond controlled trial settings [14]. It encompasses diverse sources, including electronic health records (EHRs), insurance claims, and structured patient registries—where data follow predefined protocols but without randomized interventions [15–17]. Furthermore, by generating real-world evidence (RWE), RWD plays a critical role in optimizing trial design, improving recruitment efficiency, optimal timepoints for the outcomes and identifying treatment-effect modifiers to support personalized medicine [18].

Despite its advantages, RWD presents challenges due to its observational nature and lack of randomization, making it prone to confounding and different types of biases [19]. Addressing these challenges requires advanced analytical methods that account for real-world complexities and strengthen causal validity. Causal inference has evolved from its classical epidemiological foundation—rooted in confounding factors, biases and randomized trials [20, 21]—to an essential tool in modern machine learning (ML), where prediction alone is insufficient for decision-making [22]. Causal machine learning (CML), which integrates ML algorithms with causal inference principles to estimate treatment effects and counterfactual outcomes from complex, high-dimensional data [22, 23], plays a crucial role in overcoming RWD limitations by mitigating confounding and improving inference. CML has emerged as a promising approach for drug development. Unlike traditional ML, which excels at pattern recognition, CML aims to determine how interventions influence outcomes, distinguishing true cause-and-effect relationships from correlations, a critical factor for evidence-based decision-making [23, 24].

A key application of RWD/CML is estimating causal treatment effects in real-world settings by comparing treated and untreated patients. Propensity scores, traditionally estimated using logistic regression, have been widely used to mitigate bias through inverse probability weighting, matching, or covariate adjustment [25]. However, ML methods such as boosting, tree-based models, and neural networks regularly outperform logistic regression by better handling non-linearity and complex interactions [26–29]. More recently, deep representational learning has improved propensity score estimation in high-dimensional data [30, 31], while prognostic models enhance matching

strategies, as described by Zhang et al. 2022. [32] Beyond propensity-based approaches, outcome regression models (e.g., G-computation) [33] directly adjust for confounding, leveraging RWD's large sample sizes. Advanced techniques like targeted maximum likelihood estimation [34] and doubly robust inference [35] enhance causal estimation by combining outcome and propensity models, with ML improving predictive accuracy [36, 37]. Parallel to epidemiology, econometrics—where randomized experiments are often infeasible—leverages tools like instrumental variable analysis and structural equations [38], recently equipping them with deep learning and kernel-based methods [39–41]. Additionally, graphical modeling, as proposed by Pearl [42, 43], explicitly represents causal assumptions to refine treatment effect estimation [44, 45]. Pearl's “do-calculus” aligns with potential outcomes in most cases but diverges in specific contexts, leading to ongoing academic debate [46]. Predictive modeling for clinical trial emulation has been validated by comparing its results to actual RCT outcomes, demonstrating its reliability [47].

This article introduces the transformative potential of applying CML to RWD in the context of drug development. Rather than focusing on a single methodological innovation, it provides a comprehensive overview of the key capabilities and outputs enabled by RWD/CML, emphasizing how these can complement and extend traditional clinical research by providing richer and more actionable insights into treatment effects. To ground this discussion, we present a series of use cases that illustrate the added value of integrating RWD/CML data with RCT data. While the promise of RWD/CML is considerable, it also brings technical, scientific, ethical, and regulatory challenges. The final section of the article critically evaluates these limitations and offers pragmatic recommendations aligned with current regulatory perspectives. By framing this integration within evolving scientific and regulatory paradigms, we aim to provide readers with a conceptual foundation for understanding the opportunities and challenges of this emerging field.

Presentation of Selected Use Cases to Illustrate How CML on RWD Complement Traditional Clinical Trials Outputs and Enhance Clinical Development

In this section we illustrate how the integration of clinical trial data with RWD and CML can generate a more comprehensive and robust evidence base on drug effects. The series of cases illustrate how artificial intelligence can reveal clinically meaningful insights, for instance identifying patient subgroups or detecting delayed outcomes. These cases showcase a range of high-value outputs relevant to drug development, including trial emulation, increased efficiency in indication expansion, evaluation of treatment

transportability, optimization of dosing strategies, and the design of innovative adaptive clinical trials. In certain contexts, RWD/CML can also facilitate the development of external control arms (ECAs), offering a rigorous alternative when traditional randomized controls are not feasible. Each use case has been chosen to highlight distinct contributions of RWD/CML across various domains—such as novel study designs, enhanced decision-making processes, or the generation of new evidence that supports multiple stakeholders. To emphasize the independent value of each contribution, the cases are presented as standalone examples without attempting to establish direct connections between them.

Identifying Subgroups and Refining Treatment Responses

A key advantage of RWD/CML is their ability to identify patient subgroups that demonstrate varying responses to a specific treatment. Predictors of treatment response may include for instance biomarkers, disease severity indicators, and longitudinal health status trends [48]. RWD offers a particularly valuable source of such information, often capturing a more comprehensive view of patient health than RCTs. By integrating these findings into predictive models, future trials can be better designed to target the most responsive patient populations, improving overall efficiency and effectiveness. This framework enhances RWD applications for clinical and regulatory decision-making and holds promise in advancing precision or personalized medicine by identifying patient subgroups that benefit most from specific treatments. Bertsimas et al. (2024) [49] introduced the R.O.A.D. framework, a method for clinical trial emulation using observational data while addressing confounding bias. Applied to 779 colorectal liver metastases patients, it accurately matched the JCOG0603 trial's 5-year recurrence-free survival (35% vs. 34%). The approach uses prognostic matching and cost-sensitive counterfactual models to correct biases and identify subgroups with 95% concordance in treatment response. Traditionally, identification of patient subgroups requires testing for an interaction effect between treatment and patient attributes, which can reduce the power of the trial and may require multiple testing corrections [50]. ML models, in contrast, excel at scanning large datasets to detect complex interactions and patterns, making them particularly well-suited for discovering subpopulations with distinct responses [51–54]. In the RWD/CML approach it is also possible to deploy the outcome model's predictions as a “digital biomarker,” effectively stratifying patients based on their predicted response and optimizing trial design accordingly [55].

Combining Information from RCT and RWD for a Comprehensive Drug Effect Assessment

Beyond patient stratification, RWD/CML enhances the integration of multiple data sources, maximizing the information derived from both RCTs and RWE. For instance, while RCTs provide robust short-term efficacy and safety data under controlled conditions, they often lack long-term follow-up, which can be supplemented by observational data from RWD sources. Such an approach is particularly useful for evaluating long-term treatment effects, identifying delayed adverse events, and assessing the sustainability of a drug's benefits in real-life settings. Additionally, ECAs that incorporate both RCT and RWD data provide an alternative to traditional randomized comparisons. However, these approaches introduce unique biases, due to the absence of randomization and systematic differences between trial and real-world populations [56]. Bayesian power priors, which assign different weights to diverse evidence sources, offer a method for addressing these biases [57]. Other similar Bayesian methodologies have also been developed to integrate historical evidence into ongoing trials, even in cases where only aggregate data are available, such as in network meta-analyses [58], synthetic evidence approaches [58], or advanced Bayesian frameworks [59, 60].

Use of RWD/CML for Indication Expansion

Additionally, drugs approved for one condition often exhibit beneficial effects in other indications, and ML-assisted real-world analyses can provide early signals of such potential cross-indications [61, 62]. CML techniques (e.g., targeted maximum likelihood estimation, double machine learning, or Bayesian causal models) generate estimates of treatment effects while addressing confounding or selection biases. This can be achieved through spontaneous use tracking, assessment of incidental risk reduction in secondary conditions, or more systematic deep phenotyping approaches that integrate molecular knowledge graphs with RWD. For instance, Gao et al. (2022) [63] present KG-Predict, a knowledge graph-based computational framework for drug repurposing which enhances predictive accuracy drug discovery and repositioning.

RWD/CML for Estimating Transportability of Trial Results Across Populations

Another critical challenge in clinical research is the transportability of treatment effect estimates across different populations. Treatment effects identified in one cohort may not directly generalize to another due to differences in patient characteristics, disease prevalence, or healthcare

settings. Adjusting for these “population shifts” is essential to ensure that findings from one study setting remain relevant elsewhere [64, 65]. In ML research, this issue is known as “concept drift” or “dataset shift,” and a rich body of evidence exists on methodologies to mitigate its impact [66]. Approaches such as *modified ensemble learning* have been specifically developed for clinical datasets, enhancing the robustness of predictive models when applied across diverse populations [67, 69]. One notable application of these techniques has been in dosing prediction, where ML-based models have been adapted to account for changing patient profiles and evolving clinical practices [68].

RWD/CML for Optimizing Dosage Determination

Finally, another crucial contribution of RWD/CML in advancing clinical trials consists in enhancing dosage optimization for improved efficacy and safety. Dosing decisions are traditionally based on phase 1 clinical data and preclinical models, which may lack the precision needed for real-world patient populations. By integrating real-world pharmacokinetic and pharmacodynamic data with predictive models, RWD-driven dose optimization strategies can improve the safety and efficacy of drug treatments. In certain therapeutic areas, natural variation in drug exposure post-approval can be analysed to refine dosing strategies and ensure optimal therapeutic effects while minimizing adverse events [69].

RWD/CML for Planning More Efficient Clinical Trials

The use of RWD is also central to describing natural history and prognostic research, which provides essential context for drug development. These studies offer insights into disease prevalence, severity, progression, and unmet medical needs. CML information is particularly valuable for refining trial eligibility criteria, identifying potential external controls, and guiding future modelling efforts. The ability to track real-world disease trajectories enables a deeper understanding of risk factors [70] and allows for the modelling of drug effects in complex clinical scenarios. Additionally, RWD/CML facilitates the study of competing risks, polypharmacy effects, and drug-drug interactions—critical considerations that are often underrepresented in RCTs, where patients typically have fewer comorbidities and concomitant medications [71].

RWD/CML for Designing Efficient Adaptive Trial and Sequential Learning

Adaptive clinical trial designs represent another major innovation facilitated by RWD/CML. Increasingly, clinical

development programs incorporate adaptive strategies such as interim futility analyses [72] and dynamic enrolment adjustments [73]. Conceptually, the entire drug development process can be viewed as a sequential learning experiment, where each trial phase informs the next. RWD/CML plays a crucial role in this iterative process, allowing protocols to be refined in real-time based on accumulating data. In statistical and CML research, adaptive experiments have been formalized through frameworks such as multi-armed bandits [74] and Markov decision processes [75], which have gained widespread application in multiple domains, including clinical trials [76–78]. These approaches enable treatment allocation or enrolment criteria to be dynamically modified as new evidence emerges. In dose-finding phase 1 studies, for instance, ML-driven adaptive methodologies have been successfully applied to continuously reassess safety and efficacy, optimizing dosing regimens in real-time [79]. Hüyük et al. 2024 [80], demonstrated how ML-based optimization can refine trial decision-making, continuously evaluating whether a study should proceed, be adjusted, or be terminated early based on efficacy trends. Bayesian methodologies further support these dynamic decision processes, with growing regulatory acceptance of Bayesian designs in adaptive clinical trials [81–83], including applications in ECAs [84, 85]. In short, CML brings flexibility, precision, and learning capacity to adaptive trial design, turning RWD into a strategic resource for guiding and refining trials far beyond classical methods.

RWD/CML Generating External and Synthetic Control Arms

RWD/CML can play a key role in the generation of ECAs as valid comparators of single arm trials [86, 87]. The recent publication of guidelines by regulatory agencies on ECAs and synthetic control arms (SCAs) is an official recognition, which will likely facilitate their future use in clinical research, especially in oncology [88]. By leveraging data from real-world patient populations as a comparator group, ECAs or SCAs allow single-arm trials to generate comparative effectiveness estimates, particularly useful for rare diseases where randomization may be infeasible [18, 86, 89]. CML in this context plays a critical role to mitigate possible selection biases and control confounding factors [90, 91]. These controls can also be used internally for decision-making, enabling pharmaceutical companies to retrospectively enrich trial results with additional comparator data or to simulate potential future trial outcomes. The use of ECAs helps optimize go/no-go decisions, fine-tune upcoming trials, and refine statistical power calculations [86].

Using RWD/CML for Target Trial Emulation

Target trial emulation (TTE) is another emerging application of RWD/CML that transforms clinical trials, using a combined analysis of RWD and RCT data. TTE allows researchers to replicate or complement RCT results with RWD, while maintaining strict comparability in terms of inclusion/exclusion criteria and outcome definitions. The comparator arm can be simulated using real-world patients, whereas advanced modelling techniques can help model the treatment arm. This approach has proven valuable in improving alignment between RCT and real-world findings, reducing discrepancies caused by differences in study design [92]. At the same time, RWD can be leveraged to explore alternative hypotheses beyond the original scope of an RCT, therefore complementing traditional trial results. For instance to evaluate long-term outcomes, assess treatment effects in broader populations, or study rare subgroups where RCT evidence may be lacking [93]. The ability to emulate future trials also enhances clinical development by supporting interim decision-making, such as inclusion/exclusion refinements [94] and futility assessments [95].

Challenges and Recommendations Regarding the Use of RWD/CML for Drug Development

The integration of AI into drug development is a groundbreaking advancement, with CML emerging as a powerful tool driving progress. The promise of RWD/CML quickly captured the attention of the pharmaceutical industry primarily due to their potential to address the inherent limitations of traditional clinical trials. However, the rapid rise of RWD/CML has been met with caution from regulatory agencies like the FDA and EMA. Additionally, the widespread adoption of RWD/CML faces additional technical and scientific challenges, such as the RWD quality and the complexity of algorithms requiring sophisticated programming. Furthermore, ethical concerns regarding data privacy, bias, and transparency continue to raise important questions about the responsible use of these technologies in drug development. The next section explores the technical, operational, and scientific challenges, followed by a discussion on the regulatory perspective of RWD/CML, as well as ethical and sociocultural considerations.

Technical/Operational and Scientific Challenges Faced by RWD/CML

CML applied to RWD faces several technical limitations that stem from both the nature of the data and the complexity of the models used. A major challenge is scalability and computational efficiency, as some CML methods require

significant computational resources to process large, high-dimensional datasets typical of RWD. Organizations must invest in the infrastructure necessary to process and analyse RWD. One important concern regarding RWD/CML is the quality of the data used as RWD databases often suffer from missing data, inconsistent or incorrect coding, measurement errors, incomplete longitudinal follow-up, and biased patient selection. These limitations can undermine model validity and raise significant regulatory concerns. Currently, best practices emphasize a thorough documentation of source databases before their use, a resource-intensive process that requires transparent communication [96–98]. The development of standardized guidelines regarding technical equipment and operation is necessary to guide organizations that wish to enter this new discipline.

Beyond technical challenges, the scientific limitations of using RWD/CML stem from gaps in expertise and the absence of universally accepted best practices. The successful application of CML requires interdisciplinary knowledge that spans epidemiology, statistics, AI, and domain-specific expertise in healthcare or life sciences. Loftus (2024) [99] argues that progress in causal modeling has been hindered by scientific perfectionism and the lack of a human-centric approach. Hernán (2018) [100] emphasizes that causal analyses typically require not only good data and algorithms but also domain expert knowledge. Unlike traditional statistical methods that have well-defined guidelines for causal inference, CML is still evolving, and consensus on best practices for using these methods with RWD is still evolving. This gap in knowledge increases the risk of improper model selection, overfitting, and misattribution of causal effects. One important takeaway from this situation is the critical importance of rigorous validation of the CML generated models to ensure reliability, accuracy, and scientific integrity [101]. Baweja et al. 2023 [102] highlighted the challenges data scientists face in keeping up with the fast-paced advancements in ML, suggesting that human factors methods could be applied to address these difficulties. Addressing these scientific barriers requires comprehensive training programs, the development of standardized guidelines, and fostering collaboration across disciplines to build the necessary expertise for advancing CML applications in real-world settings.

Regulatory Perspective on Developing RWD/CML

The regulatory review of AI-generated results presents a significant factor shaping the path forward for this emerging technology. Both the FDA and EMA have been actively addressing the integration of AI and RWD in drug development by progressively developing frameworks and guidelines to balance innovation with regulatory rigor. The FDA

first acknowledged the utility of RWE/RWD to complement clinical trials in 2017, particularly for complementing clinical trials in specialized contexts such as oncology [98]. However, the use of CML on RWD introduces regulatory concerns, particularly regarding the transparency of AI-driven models. Many CML models function as “black boxes,” offering little insight into their internal decision-making processes. This opacity raises concerns about the reproducibility of scientific findings and the trustworthiness of AI-generated evidence, especially in regulatory decision-making, where understanding the rationale behind conclusions is crucial [103]. Recognizing these challenges, the FDA released its first set of recommendations on AI in drug development in January 2025. This guidance introduced a risk-based framework for sponsors to assess and establish the credibility of AI models within defined contexts of use, outlining necessary steps to ensure model reliability and applicability [104]. In Europe, the EMA published a reflection paper on AI in drug development in September 2024, addressing key issues such as data quality, model validation, and the ethical implications of AI deployment in clinical research. This document underscores the agency’s commitment to ensuring that AI applications are both scientifically sound and ethically justified [105]. These regulatory efforts reflect a broader commitment to balancing patient safety, product efficacy, and innovation in the rapidly evolving field of AI-driven drug development. They also underscore the responsibility on the research teams to develop high-quality models that are transparent and can be reproducible—an effort that, in practice, requires addressing the challenges outlined in section “*Technical/Operational and Scientific Challenges*”.

Ethical Challenges in RWD/CML Integration

The integration of RWD and CML into drug development also raises significant ethical concerns, particularly regarding privacy, consent, and data ownership [106, 107]. Since RWD is often derived from EHRs, insurance claims, and other sensitive sources, compliance with privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. and the General Data Protection Regulation (GDPR) in Europe, is essential to protect patient confidentiality. However, the use of large-scale datasets required for AI and CML modelling, along with extensive exploratory analyses of the variable relationships through causal networks, increases the risk of re-identifying anonymized patient records, potentially leading to privacy breaches [108]. Additionally, ambiguities surrounding data ownership—whether it belongs to patients, healthcare institutions, or third-party entities—can lead to disputes over access, usage rights, and potential commercialization,

further complicating the ethical landscape [108]. Establishing robust ethical frameworks is crucial to ensuring that AI-driven healthcare solutions uphold ethical principles while enhancing patient care [109].

A final critical barrier to AI adoption in drug development is sociocultural, driven by skepticism among healthcare professionals and patients toward AI-generated recommendations. Concerns over the fairness, transparency, and potential biases of algorithmic decisions can foster public reluctance and mistrust, slowing widespread acceptance. Addressing these challenges requires clear communication, robust accountability measures, and proactive efforts to ensure diverse representation in research. Moreover, since RWD encompasses individuals from diverse cultural, socioeconomic, and healthcare backgrounds, CML models must be designed to account for these differences, ensuring that treatment recommendations are both equitable and culturally relevant. Building trust in AI-driven applications depends on ongoing collaboration with healthcare providers, active patient engagement, and a steadfast commitment to transparency.

Conclusions

This paper has highlighted the rapid development of CML and RWD as complementary tools in drug development, addressing limitations of traditional clinical trials. CML enables unbiased drug effect estimation, responder subgroup identification, and long-term efficacy and safety assessment. It also supports clinical trial adaptation and external control generation. However, widespread adoption is hindered by the limited quality of some source-databases, scarce relevant technical expertise, lack of standardized guidelines, and universally accepted best practices. Overcoming these barriers requires multidisciplinary collaboration to establish rigorous methodologies, develop operational guidelines, and integrate these approaches into widely used software packages. Regulatory agencies are progressively providing guidance and structured initiatives to evaluate these technologies responsibly. Achieving broader acceptance and trust, however, requires transparent communication with stakeholders regarding how these models are developed, tested, and applied. These methods should only be implemented by institutions with the necessary infrastructure and expertise to ensure methodological rigor, transparency, and reproducibility. By fostering stakeholder engagement and establishing rigorous, evidence-based implementation strategies, these innovative approaches can complement RCTs and accelerate the future of drug development programs and regulatory science.

Acknowledgements We extend our gratitude to Dr. Mir Sohail Fazeli from Evidinno Outcomes Research Inc. (Vancouver, BC, Canada) for his insightful feedback and help in drafting and finalizing this manuscript.

Author Contributions C.A was the primary contributor to the manuscript, responsible for drafting most of its content. R.H.V provided the foundational idea that gave direction to the work and was responsible for placing the current position in historical perspective. M.V.D.S offered deep technical expertise and her perspective on the cutting edge of machine learning research in this domain. J.P.C contributed expertise in clinical research, causality assessment and drug development. All authors collaborated closely and co-edited the manuscript, collectively contributing to its quality and overall positioning.

Funding No external funding was utilized for manuscript conceptualization, development and revisions. Editorial support was provided by Oxford PharmaGenesis (Oxford, UK) and Evidinno Outcomes Research Inc. (Vancouver, BC, Canada), funded by Sanofi.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests Chris Anagnostopoulos is a consultant with QuantumBlack– AI by McKinsey, a global consulting firm that serves a broad range of healthcare/life sciences companies across industry and governments, on topics related to AI, machine learning, drug discovery and development and digital transformations. Mihaela Van Der Schaar reports employments with the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK and with The Alan Turing Institute, London, UK. Jean-Paul Collet reports employment with Evidinno Outcomes Research Inc. and Ramon Hernandez Vecino is a Sanofi employee and may hold shares and/or stock options in the company. The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hardman TC, Aitchison R, Scaife R, Edwards J, Slater G. The future of clinical trials and drug development: 2050. *Drugs Context*. 2023;12doi. <https://doi.org/10.7573/dic.2023-2-2>.
- Taylor KPF, Cruz M, Ronte H, Haughey J. Intelligent clinical trials: transforming through AI-enabled engagement. 2020.
- Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm Sinica B* Jul. 2022;12(7):3049–62. <https://doi.org/10.1016/j.apsb.2022.02.002>.
- Farid SS, Baron M, Stamatis C, Nie W, Coffman J. Benchmarking biopharmaceutical process development and manufacturing cost contributions to R&D. *mAbs*. Jan-Dec. 2020;12(1):1754999. <https://doi.org/10.1080/19420862.2020.1754999>.
- Schlender M, Hernandez-Villafuerte K, Cheng CY, Mestre-Ferrandiz J, Baumann M. How much does it cost to research and develop a new drug? A systematic review and assessment. *PharmacoEconomics* Nov. 2021;39(11):1243–69. <https://doi.org/10.1007/s40273-021-01065-y>.
- Clark LT, Watkins L, Piña IL et al. May. Increasing diversity in clinical trials: overcoming critical barriers. *Current problems in cardiology*. 2019;44(5):148–72. <https://doi.org/10.1016/j.cpcardi.2018.11.002>.
- Taira DA, Ranken MS, Seto BK, et al. Representation of native Hawaiian and Pacific Islander individuals in clinical trials. *JAMA Netw Open*. 2024;7(10):e2442204–2442204. <https://doi.org/10.1001/jamanetworkopen.2024.42204>.
- Bahnam P, Hanzel J, Ma C, et al. Most placebo-controlled trials in inflammatory bowel disease were underpowered because of overestimated drug efficacy rates: results from a systematic review of induction studies. *Journal Crohn's & Colitis* Apr. 2023;3(3):404–17. <https://doi.org/10.1093/ecco-jcc/jjac150>.
- Eichler HG, Sweeney F. Feb. The evolution of clinical trials: can we address the challenges of the future? *Clinical trials* (London, England). 2018;15(1_suppl):27–32. <https://doi.org/10.1177/1740774518755058>.
- Suissa S. Run-in bias in randomised trials: the case of COPD medications. *The Eur Respiratory Journal* Jun. 2017;49(6). <https://doi.org/10.1183/13993003.00361-2017>.
- Molero-Calafell J, Burón A, Castells X, Porta M. Intention to treat and per protocol analyses: differences and similarities. *Journal Clin Epidemiology* Sep. 2024;173:111457. <https://doi.org/10.1016/j.jclinepi.2024.111457>.
- Abi Jaoude J, Kouzy R, Ghabach M et al. Food and Drug Administration approvals in phase 3 Cancer clinical trials. *BMC Cancer*. 2021/06/12 2021;21(1):695. <https://doi.org/10.1186/s12885-021-08457-5>.
- Bellot A, Schaar M. Multitask boosting for survival analysis with competing risks. presented at: Proceedings of the 32nd International Conference on Neural Information Processing Systems; 2018; Montréal, Canada.
- Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Intern Medicine* Dec. 2016;1(12):1826–33. <https://doi.org/10.1001/jamainternmed.2016.6008>.
- Boulanger V, Schlemmer M, Rossov S, Seebald A, Gavin P. Establishing patient registries for rare diseases: rationale and challenges. *Pharmaceutical Medicine* Jun. 2020;34(3):185–90. <https://doi.org/10.1007/s40290-020-00332-1>.
- Chen J, Ho M, Lee K, et al. The current landscape in biostatistics of real-world data and evidence: clinical study design and analysis. *Stat Biopharm Res*. 2021;02/01:15:1–28. <https://doi.org/10.1080/19466315.2021.1883474>.
- Real-World FDA. Evidence. <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence> Accessed Feb 20, 2025. <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>.
- Dagenais S, Russo L, Madsen A, Webster J, Becnel L. Use of real-world evidence to drive drug development strategy and inform clinical trial design. *Clinical Pharmacol Therapeutics* Jan. 2022;111(1):77–89. <https://doi.org/10.1002/cpt.2480>.
- Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight - reconsidering the use of race correction in clinical algorithms. *The New Engl J Medicine* Aug. 2020;27(9):874–82. <https://doi.org/10.1056/NEJMms2004740>.

20. Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol.* 1974;10/01(66). <https://doi.org/10.1037/h0037350>.
21. Miguel A, Hernan, Robins JM. Causal Inference What If. 1st ed. 2020.
22. Moccia C, Moirano G, Popovic M, et al. Machine learning in causal inference for epidemiology. *European J Epidemiology* Oct. 2024;39(10):1097–108. <https://doi.org/10.1007/s10654-024-01173-x>.
23. Sanchez P, Voisey JP, Xia T, Watson HI, O'Neil AQ, Tsafaris SA. Causal machine learning for healthcare and precision medicine. *Royal Soc Open Sci.* 2022;9(8):220638. <https://doi.org/10.1098/rsos.220638>.
24. IBM. What is machine learning? <https://www.ibm.com/think/topics/machine-learning> Accessed Feb 18, 2025. Accessed Feb 18, 2025. <https://www.ibm.com/think/topics/machine-learning>
25. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55. <https://doi.org/10.1093/biomet/70.1.41>.
26. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. *Statistics Medicine* Feb. 2010;10(3):337–46. <https://doi.org/10.1002/sim.3782>.
27. McCaffrey DF, Ridgeway G, Morral AR. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods* Dec. 2004;9(4):403–25. <https://doi.org/10.1037/1082-989x.9.4.403>
28. Setoguchi S, Schneeweiss S, Brookhart MA, Glynn RJ, Cook EF. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology Drug Safety* Jun. 2008;17(6):546–55. <https://doi.org/10.1002/pds.1555>.
29. Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal Clin Epidemiology* Aug. 2010;63(8):826–33. <https://doi.org/10.1016/j.jclinepi.2009.11.020>.
30. Ghosh S, Bian J, Guo Y, Prosperi M. Deep propensity network using a sparse autoencoder for estimation of treatment effects. *Journal Am Med Inf Association: JAMIA* Jun. 2021;12(6):1197–206. <https://doi.org/10.1093/jamia/ocaa346>
31. Weberpals J, Becker T, Davies J, et al. Deep learning-based propensity scores for confounding control in comparative effectiveness research: a large-scale, real-world data study. *Epidemiology (Cambridge, Mass).* 2021;1(3):378–88. <https://doi.org/10.1097/de.0000000000001338>
32. Zhang Y, Yang S, Ye W, Faries DE, Lipkovich I, Kadziola Z. Practical recommendations on double score matching for estimating causal effects. *Statistics Medicine* Apr. 2022;15(8):1421–45. <https://doi.org/10.1002/sim.9289>.
33. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics - Theory and Methods.* 1991/01/01 1991;20(8):2609–2631. <https://doi.org/10.1080/03610929108830654>
34. Laan MJvd, Rubin D. Targeted maximum likelihood learning. *Int J Biostatistics.* 2006;2(1). <https://doi.org/10.2202/1557-4679.1043>.
35. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. *American J Epidemiology* Apr. 2011;1(7):761–7. <https://doi.org/10.1093/aje/kwq439>
36. Hill JL. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics.* 2011/01/01 2011;20(1):217–240. <https://doi.org/10.1198/jcgs.2010.08162>
37. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol.* 2017;185(1):65–73. <https://doi.org/10.1093/aje/kww165>
38. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health.* 1998;19:17–34. <https://doi.org/10.1146/annurev.publhealth.19.1.17>.
39. Guo Z, Small DS. Control function instrumental variable estimation of nonlinear causal effect models. *J Mach Learn Res.* 2016;17(1):3448–82.
40. Singh R, Sahani M, Gretton A. Kernel instrumental variable regression. *Proceedings of the 33rd International Conference on Neural Information Processing Systems.* Curran Associates Inc.; 2019:Article 413.
41. Wu A, Kuang K, Xiong R, Wu F. Instrumental variables in causal inference and machine learning: A survey. *ArXiv.* 2022;abs/2212.05778.
42. Almond R, Graphical Models SL, Lauritzen, Oxford University Press. 1996. No. of pages: 298. ISBN 0-19-852219-3. *Statistics in medicine.* 1999;18(21):2983–2984. [https://doi.org/10.1002/\(SICI\)1097-0258\(19991115\)18:21%3C2983::AID-2013SIM198%3E3.0.CO;2%2013A](https://doi.org/10.1002/(SICI)1097-0258(19991115)18:21%3C2983::AID-2013SIM198%3E3.0.CO;2%2013A)
43. Pearl J. Causal diagrams for empirical research. *Biometrika.* 1995;82(4):669–88. <https://doi.org/10.1093/biomet/82.4.669>.
44. Glymour C, Zhang K, Spirtes P. Review of causal discovery methods based on graphical models. *Front Genet.* 2019;10:524. <https://doi.org/10.3389/fgene.2019.00524>.
45. Pamfil R, Sriwattanaworachai N, Desai S et al. DYNOTEARS: Structure Learning from Time-Series Data. presented at: Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics; 2020; Proceedings of Machine Learning Research. <https://proceedings.mlr.press/v108/pamfil20a.html>
46. Dawid A, Senn S. Personalised Decision-Making without Counterfactuals. 2023.
47. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA.* 2022;328(24):2446–7. <https://doi.org/10.1001/jama.2022.21383>
48. Qian Z, Alaa A, Bellot A, Schaar M, Rashbass J. Learning Dynamic and Personalized Comorbidity Networks from Event Data using Deep Diffusion Processes. presented at: Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics; 2020; Proceedings of Machine Learning Research. <https://proceedings.mlr.press/v108/qian20a.html>
49. Bertsimas D, Koulouras A, Nagata H, et al. The R.O.A.D. to clinical trial emulation. 2024.
50. Bender R, Lange S. Adjusting for multiple testing—when and how? *Journal Clin Epidemiology* Apr. 2001;54(4):343–9. [https://doi.org/10.1016/s0895-4356\(00\)00314-0](https://doi.org/10.1016/s0895-4356(00)00314-0).
51. Zhang Y, Bellot A, Schaar M. Learning overlapping representations for the estimation of individualized treatment effects. presented at: proceedings of the twenty third international conference on artificial intelligence and statistics; 2020; Proceedings of Machine Learning Research. <https://proceedings.mlr.press/v108/zhang20c.html>
52. Alaa AM, Schaar Mvd. Bayesian inference of individualized treatment effects using multi-task gaussian processes. *ArXiv.* 2017;abs/1704.02801.
53. Bica I, Alaa AM, Lambert C, van der Schaar M. Learning: Current and Future Methods to Address Underlying Challenges. *Clinical Pharmacol therapeutics* Jan. 2021;109(1):87–100. <https://doi.org/10.1002/cpt.1907>. From Real-World Patient Data to Individualized Treatment Effects Using Machine.
54. Logan BR, Sparapani R, McCulloch RE, Laud PW. Decision making and uncertainty quantification for individualized treatments using bayesian additive regression trees. *Statistical methods in*

- medical research. Apr. 2019;28(4):1079–93. <https://doi.org/10.1177/0962280217746191>
55. Tao W, Concepcion AN, Vianen M et al. Feb. Multiomics and Machine Learning Accurately Predict Clinical Response to Adalimumab and Etanercept Therapy in Patients With Rheumatoid Arthritis. *Arthritis & rheumatology* (Hoboken, NJ). 2021;73(2):212–222. <https://doi.org/10.1002/art.41516>
 56. Adamson BJS, Ma X, Griffith SD, Sweeney EM, Sarkar S, Bourla AB. Differential frequency in imaging-based outcome measurement: bias in real-world oncology comparative-effectiveness studies. *Pharmacoepidemiology Drug Safety* Jan. 2022;31(1):46–54. <https://doi.org/10.1002/pds.5323>.
 57. Lu N, Wang C, Chen WC, et al. Propensity score-integrated power prior approach for augmenting the control arm of a randomized controlled trial by incorporating multiple external data sources. *Journal Biopharm Statistics* Jan. 2022;2(1):158–69. <http://doi.org/10.1080/10543406.2021.1998098>.
 58. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. *Medical Decis Making: Int J Soc Med Decis Making* Jul. 2013;33(5):679–91. <https://doi.org/10.1177/0272989x13485156>.
 59. Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* Sep. 2011;67(3):1047–56. <https://doi.org/10.1111/j.1541-0420.2011.01564.x>.
 60. Jenkins DA, Hussein H, Martina R, Dequen-O'Byrne P, Abrams KR, Bujkiewicz S. Methods for the inclusion of real-world evidence in network meta-analysis. *BMC Med Res Methodol*. 2021;2021/10/09(1):207. <https://doi.org/10.1186/s12874-021-01399-3>.
 61. Tanoli Z, Vähä-Koskela M, Aittokallio T. Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opin Drug Discovery* Sep. 2021;16(9):977–89. <https://doi.org/10.1080/17460441.2021.1883585>.
 62. Gurbuz O, Alanis-Lobato G, Picart-Armada S, et al. Knowledge graphs for indication expansion: an explainable target-disease prediction method. *Front Genet*. 2022;13:814093. <https://doi.org/10.3389/fgene.2022.814093>
 63. Gao Z, Ding P, Xu R, KG-Predict. A knowledge graph computational framework for drug repurposing. *J Biomed Inform*. 2022. <https://doi.org/10.1016/j.jbi.2022.104133>. /08/01/2022;132:104133.
 64. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *American J Epidemiology* Jul. 2010;1(1):107–15. <https://doi.org/10.1093/aje/kwq084>.
 65. Hartman E, Grieve R, Ramsahai R, Sekhon J. From Sample Average Treatment Effect to Population Average Treatment Effect on the Treated: Combining Experimental with Observational Studies to Estimate Population Treatment Effects. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 06/01. 2015;17810.1111/rssa.12094
 66. Lu J, Liu A, Dong F, Gu F, Gama J, Zhang G. Learning under concept drift: A review. *IEEE Trans Knowl Data Eng*. 2019;31(12):2346–63. <https://doi.org/10.1109/TKDE.2018.2876857>.
 67. Anagnostopoulos C, Gramacy RB. Information-theoretic data discarding for dynamic trees on data streams. *Entropy*. 2013;15(12):5510–35.
 68. Chan AJ, Schaar M. Synthetic model combination: an instance-wise approach to unsupervised ensemble learning. presented at: Proceedings of the 36th International Conference on Neural Information Processing Systems; 2022; New Orleans, LA, USA.
 69. Pan S, Tsakok T, Dand N, et al. Mar. Using real-world data to guide ustekinumab dosing strategies for psoriasis: a prospective pharmacokinetic-pharmacodynamic study. *Clinical and translational science*. 2020;13(2):400–409. <https://doi.org/10.1111/cts.12725>
 70. Seeff LB. Nov. Natural history of chronic hepatitis C. *Hepatology* (Baltimore, Md). 2002;36(5 Suppl 1):S35–46. <https://doi.org/10.1053/jhep.2002.36806>
 71. Burdett N, Vincent AD, O'Callaghan M, Kichenadasse G. Competing risks in older patients with cancer: a systematic review of geriatric oncology trials. *Journal Natl Cancer Institute* Aug. 2018;1(8):825–30. <https://doi.org/10.1093/jnci/djy111>
 72. Lai TL, Liao OY. Efficient adaptive randomization and stopping rules in Multi-arm clinical trials for testing a new treatment. *Sequential Anal*. 2012;31(4):441–57. <https://doi.org/10.1080/07474946.2012.719433>.
 73. Park JJ, Thorlund K, Mills EJ. Critical concepts in adaptive clinical trials. *Clin Epidemiol*. 2018;10:343–51. <https://doi.org/10.2147/clep.S156708>.
 74. Bouneffouf D, Rish I, Aggarwal C. Survey on Applications of Multi-Armed and Contextual Bandits. 2020:1–8.
 75. White DJ. A Survey of Applications of Markov Decision Processes. *Journal of the Operational Research Society*. 1993/11/01 1993;44(11):1073–1096. <https://doi.org/10.1057/jors.1993.181>
 76. Bazrafshan N, Lotfi M. A finite-horizon Markov decision process model for cancer chemotherapy treatment planning: an application to sequential treatment decision making in clinical trials. *Ann Oper Res*. 2020;12/01(295). <https://doi.org/10.1007/s10479-020-03706-5>.
 77. Varatharajah Y, Berry B. A Contextual-Bandit-Based approach for informed Decision-Making in clinical trials. *Life* (Basel, Switzerland). Aug. 2022;21(8). <https://doi.org/10.3390/life12081277>
 78. Villar SS, Bowden J, Wason J. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Stat Science: Rev J Inst Math Stat*. 2015;30(2):199–215. <https://doi.org/10.1214/14-sts504>.
 79. Aziz M, Kaufmann E, Riviere M-K. On multi-armed bandit designs for dose-finding clinical trials. *J Mach Learn Res*. 2021;22(1):14.
 80. Alihan Hüyük ZQ, Mihaela van der Schaar. Adaptive Experiment Design with Synthetic Controls. Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. 2024.
 81. Berry DA. Bayesian clinical trials. *Nature Reviews Drug Discovery* Jan. 2006;5(1):27–36. <https://doi.org/10.1038/nrd1927>.
 82. Green PJ, Łatuszyński K, Pereyra M, Robert CP. Bayesian computation: a summary of the current state, and samples backwards and forwards. *Statistics and Computing*. 2015/07/01 2015;25(4):835–862. <https://doi.org/10.1007/s11222-015-9574-5>
 83. Gupta SK. Use of bayesian statistics in drug development: advantages and challenges. *International J Appl & Basic Med Research* Jan. 2012;2(1):3–6. <https://doi.org/10.4103/2229-516x.96789>.
 84. Nikolakopoulos S, van der Tweel I, Roes KCB. Dynamic borrowing through empirical power priors that control type I error. *Biometrics* Sep. 2018;74(3):874–80. <https://doi.org/10.1111/biom.12835>.
 85. Viele K, Berry S, Neuenschwander B, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical statistics*. Jan-Feb. 2014;13(1):41–54. <https://doi.org/10.1002/pst.1589>
 86. Seeger JD, Davis KJ, Iannacone MR, et al. Methods for external control groups for single arm trials or long-term uncontrolled extensions to randomized clinical trials. *Pharmacoepidemiology Drug Safety* Nov. 2020;29(11):1382–92. <https://doi.org/10.1002/pds.5141>.
 87. Davi R, Mahendraratnam N, Chatterjee A, Dawson CJ, Sherman R. Informing single-arm clinical trials with external controls.

- Nature Reviews Drug Discovery Dec. 2020;19(12):821–2. <https://doi.org/10.1038/d41573-020-00146-5>.
88. FDA. Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products> Accessed Feb 25, 2025. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products>
 89. Mack C, Christian J, Brinkley E, Warren EJ, Hall M, Dreyer N. When context is hard to come by: external comparators and how to use them. *Therapeutic innovation & regulatory science*. Nov. 2019;5:2168479019878672. <https://doi.org/10.1177/2168479019878672>
 90. Han B, Zhan J, John Zhong Z, Liu D, Lindborg S. Covariate-adjusted borrowing of historical control data in randomized clinical trials. *Pharmaceutical Statistics* Jul. 2017;16(4):296–308. <http://doi.org/10.1002/pst.1815>.
 91. Lambert J, Lengline E, Porcher R, Thiébaud R, Zohar S, Chevret S. Enriching single-arm clinical trials with external controls: possibilities and pitfalls. *Blood Adv*. 2022;12/19(7). <https://doi.org/10.1182/bloodadvances.2022009167>.
 92. Wang SV, Schneeweiss S, Initiative R-D. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA*. 2023;329(16):1376–85. <https://doi.org/10.1001/jama.2023.4221>.
 93. Thompson D. Jan. Replication of randomized, controlled trials using Real-World data: what could go wrong? *Value in health: the journal of the international society for pharmacoeconomics and outcomes research*. 2021;24(1):112–5. <https://doi.org/10.1016/j.jval.2020.09.015>
 94. Kwee SA, Wong LL, Ludema C, et al. Target trial emulation: A design tool for cancer clinical trials. *JCO Clin Cancer Informatics* Jan. 2023;7:e2200140. <https://doi.org/10.1200/cci.22.00140>.
 95. Martínez-Alés G, Domingo-Relloso A, Quintana-Díaz M, Fernández-Capitán C, Hernán MA. Thromboprophylaxis with standard-dose vs. flexible-dose heparin for hospitalized COVID-19 patients: a target trial emulation. *Journal Clin Epidemiology* Nov. 2022;151:96–103. <https://doi.org/10.1016/j.jclinepi.2022.08.006>.
 96. Ding J, Li X. An Approach for Validating Quality of Datasets for Machine Learning. 2018:2795–2803.
 97. Suján M, Smith-Frazer C, Malamateniou C, et al. Validation framework for the use of AI in healthcare: overview of the new British standard BS30440. *BMJ Health & Care Informatics* Jun. 2023;30(1). <https://doi.org/10.1136/bmjhci-2023-100749>.
 98. FDA. Framework for FDA's real-world evidence program. <https://www.fda.gov/media/120060/download> Accessed Feb 20, 2025. Accessed Feb 20, 2025. <https://www.fda.gov/media/120060/download>
 99. Loftus J. Position: The Causal Revolution Needs Scientific Pragmatism. 2024.
 100. Hernán MA. The C-Word: scientific euphemisms do not improve causal inference from observational data. *American J Public Health* May. 2018;108(5):616–9. <https://doi.org/10.2105/ajph.2018.304337>.
 101. Tsopra R, Fernandez X, Luchinat C, et al. A framework for validating AI in precision medicine: considerations from the European ITFoC consortium. *BMC Med Inf Decis Mak*. 2021;201/10/02(1):274. <https://doi.org/10.1186/s12911-021-01634-3>.
 102. Baweja JA, Fallon CK, Jefferson BA. Opportunities for human factors in machine learning. *Front Artif Intell*. 2023;6:1130190. <https://doi.org/10.3389/frai.2023.1130190>.
 103. Farhud DD, Zokaei S. Ethical issues of artificial intelligence in medicine and healthcare. *Iranian J Public Health Nov*. 2021;50(11):i–v. <https://doi.org/10.18502/ijph.v50i11.7600>.
 104. FDA. FDA Proposes Framework to Advance Credibility of AI Models Used for Drug and Biological Product Submissions. <https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions> Accessed Feb 23, 2025. Accessed Feb 23, 2025. <https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions>
 105. EMA. Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf Accessed Feb 20, 2025. Accessed Feb 20, 2025. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf
 106. Boudi AL, Boudi M, Chan C, Boudi FB. Ethical challenges of artificial intelligence in medicine. *Cureus Nov*. 2024;16(11):e74495. <https://doi.org/10.7759/cureus.74495>.
 107. Solà-Morales O, Sigurðardóttir K, Akehurst R et al. Data Governance for Real-World Data Management: A proposal for a checklist to support decision making. *Value Health*. 2023/04/01/2023;26(4, Supplement):32–42. <https://doi.org/10.1016/j.jval.2023.02.012>
 108. Naik N, Hameed BMZ, Shetty DK, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility?? *Front Surg*. 2022;9:862322. <https://doi.org/10.3389/fsurg.2022.862322>.
 109. Rahmattalabi A, Xiang A. Promises and Challenges of Causality for Ethical Machine Learning. <https://montrealethics.ai/promises-and-challenges-of-causality-for-ethical-machine-learning/> Accessed Feb 21, 2025. Accessed Feb 21, 2025. <https://montrealethics.ai/promises-and-challenges-of-causality-for-ethical-machine-learning/>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.