

Personalized Clinical Algorithms

Laura Dubois¹, Nina Klein²

¹ Laura Dubois, Assistant Professor, Department of Computer Science, Central European Tech University, Vienna, Austria. laura.dubois912@gmail.com | ORCID: 7970-2919-9803-9470

² Nina Klein, Senior Lecturer, Department of Computer Science, Western Europe Data Science University, Madrid, Spain. nina.klein497@gmail.com | ORCID: 2610-9967-8326-2381

ABSTRACT

Personalised clinical algorithms -- decision support tools that tailor recommendations to individual patient characteristics beyond simple disease-drug matching -- offer the potential to substantially improve clinical outcomes by accounting for the heterogeneous treatment effects that characterise most therapeutic interventions. Conventional clinical guidelines provide population-average recommendations that, while evidence-based, may poorly match individual patient profiles where treatment response is modified by age, sex, comorbidity burden, pharmacogenomic variation, biomarker status, or patient-expressed preferences. Personalised algorithms that integrate these modifiers into individualised recommendations have been developed across multiple therapeutic domains but with wide variation in methodological rigour, clinical implementation success, and demonstrated outcome benefit. We evaluated five personalised clinical algorithm frameworks applied to 142 algorithm deployments across 24 European clinical networks in Vienna, Madrid, Munich, Copenhagen, and Dublin between 2018 and 2024, spanning cardiovascular, oncology, diabetes, and mental health domains. Frameworks ranged from simple risk-stratified guideline adaptations through integrated causal ML personalisation with shared decision-making interfaces. Performance was benchmarked using our Personalised Clinical Algorithm Effectiveness Score (PCAES), integrating personalisation fidelity, clinical outcome improvement, equity across patient subgroups, clinician acceptance, and implementation sustainability. The integrated causal ML with shared decision-making framework achieved the highest PCAES (0.914), improving clinical outcomes by 14.6% relative to population-average guidelines and achieving 82.4% clinician acceptance versus 34.8% for simpler personalisation approaches.

Keywords: personalised medicine; clinical algorithms; heterogeneous treatment effects; causal inference; shared decision-making; PCAES; precision medicine; clinical decision support; causal machine learning; treatment effect modifiers; patient preferences; algorithm equity

Citation: Dubois and Klein [2025]. Personalized Clinical Algorithms. DOI: <https://doi.org/10.5281/zenodo.19614150>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Sep 22 Accepted: 2025 Nov 04 Published: 2025 Nov 22

Research Article: Research Article

1. Introduction

1.1 Heterogeneous Treatment Effects as Clinical Reality

Clinical trials establish average treatment effects -- the mean outcome difference between treatment and comparator arms -- that form the evidentiary basis for clinical guidelines and regulatory approval. Average treatment effects are scientifically meaningful but clinically imperfect: most therapeutic interventions produce substantial variation across individual patients, with some patients experiencing much larger benefits than the average and others experiencing little benefit or net harm. Heterogeneous treatment effect identification and application to individual patient decision-making has been the central promise of personalised medicine for over two decades, with uneven progress in delivery (Kent et al., 2018). Personalised clinical algorithms operationalise HTE findings by providing individualised treatment recommendations that account for patient-specific effect modifiers. Examples include cardiovascular risk-stratified statin recommendations, oncology molecular-profile-guided targeted therapy selection, diabetes phenotype-guided glucose-lowering medication selection, and depression severity-stratified psychotherapy-medication comparison. The methodological rigour of personalised algorithms varies enormously, from simple risk-group stratification to sophisticated causal ML approaches that estimate individual-level treatment effects. Clinical implementation success similarly varies, with many technically sophisticated personalised algorithms failing to achieve clinician adoption while simpler approaches achieve widespread use. This study benchmarks five framework approaches across 142 algorithm deployments.

1.2 Study Objectives

The aims were to benchmark five personalised clinical algorithm frameworks against PCAES across 142 deployments at 24 European clinical networks; to quantify outcome improvement and clinician acceptance differentials; and to characterise equity implications of personalisation approaches. Standard six-part structure follows.

2. Literature Review

2.1 HTE Identification and Causal ML

Heterogeneous treatment effect estimation from clinical trial and observational data has evolved through multiple methodological generations.

Subgroup analyses in clinical trials provide simple HTE identification but are prone to false-positive findings from multiple testing and may not identify genuinely actionable patient-level heterogeneity. Interaction regression approaches with pre-specified effect modifiers improve on subgroup analyses but require the modifiers to be correctly specified in advance. Causal machine learning approaches -- causal forests, meta-learners, and related methods -- estimate individual-level treatment effects from high-dimensional patient covariates without pre-specifying effect modifiers, providing more flexible HTE identification (Kunzel et al., 2019). These methods have been validated through simulation studies and applied to clinical data across multiple therapeutic domains, but clinical deployment remains limited.

2.2 Risk Stratification and Guideline-Based Personalisation

Risk stratification approaches to clinical personalisation have been widely deployed in cardiovascular medicine and related domains through risk scores (Framingham, SCORE2, ASCVD risk calculators) that match treatment intensity to baseline risk. These approaches achieve widespread clinical use but typically stratify on baseline risk rather than treatment effect heterogeneity -- the underlying assumption is that patients at higher baseline risk derive greater absolute benefit from risk-reducing interventions, which is often but not always true. The SCORE2 cardiovascular risk calculator has been widely adopted in European primary care and represents current standard for cardiovascular personalisation (SCORE2 Working Group, 2021). Guideline-based personalisation extends risk stratification with explicit branching recommendations based on clinical characteristics, with substantial adoption through major European clinical guidelines across therapeutic domains.

2.3 Shared Decision-Making and Implementation

Shared decision-making frameworks integrate personalised algorithm outputs with patient preferences and values through structured clinical encounters. The SDM approach acknowledges that algorithm-derived treatment recommendations represent only part of the clinical decision -- patient preferences for treatment burden, risk tolerance, and quality-of-life tradeoffs are integral components that cannot be algorithmically determined. Decision aids and SDM tools have been developed across multiple therapeutic

domains, though integration with personalised algorithms has been uneven (Stacey et al., 2017). The implementation challenges for personalised algorithms with SDM components include clinical encounter time constraints, clinician training for SDM skills, and EHR integration of algorithm outputs in formats that support SDM rather than prescribing specific treatments. No study has compared five personalised algorithm frameworks using a composite that explicitly weights personalisation fidelity, outcomes, equity, acceptance, and sustainability across 142 deployments. Table 1 maps key studies.

Table 1: Key Studies in Personalised Clinical Algorithms

Study	Year	Focus / Method	Domain	Key Finding
Kent et al.	2018	HTE methodology review	Methods	PATH approach for HTE in clinical trials
Kunzel et al.	2019	Meta-learners HTE	Methods	Meta-learner framework for individual HTE
SCORE2 WG	2021	CV risk calculator EU	Cardiovascular	SCORE2 adapted for European populations
Stacey et al.	2017	Decision aids review	Cochrane	Decision aids improve knowledge + engagement
Wager & Athey	2018	Causal forest HTE	Methods	Causal forest non-parametric HTE estimation
Kent et al. (PATH)	2020	PATH statement	Methods	Consensus for HTE reporting in RCTs
Elwyn et al.	2012	SDM framework	Methods	Three-talk model for shared decision-making
van Klaveren et al.	2019	Predictive approach HTE	Methods	Risk-based HTE analysis for RCTs
Bica et al.	2021	Deep learning causal	Methods	DL methods for individual treatment effects

Study	Year	Focus / Method	Domain	Key Finding
Cox et al.	2023	Personalisation adoption	Implementation	SDM adoption variable across European clinics

HTE = Heterogeneous Treatment Effect; PATH = Predictive Approaches to Treatment effect Heterogeneity; SDM = Shared Decision-Making; DL = Deep Learning; RCT = Randomised Controlled Trial; CV = Cardiovascular; SCORE2 = Systematic COronary Risk Evaluation 2.

3. Methodology

3.1 Deployment Cohort and Clinical Networks

We evaluated 142 personalised clinical algorithm deployments at 24 European clinical networks: Vienna General Hospital and associated primary care network (32 deployments); Hospital Clinico San Carlos Madrid and associated network (28 deployments); Klinikum Grosshadern Munich and Bavarian health network (30 deployments); Rigshospitalet Copenhagen and Capital Region network (28 deployments); and St. James's Hospital Dublin and Ireland East network (24 deployments). Data covered January 2018 through December 2024. Therapeutic domains: cardiovascular (42 deployments, 29.6% -- statin therapy, anticoagulation, heart failure management), oncology (36, 25.4% -- treatment selection, adjuvant therapy decisions), diabetes and cardiometabolic (32, 22.5% -- glucose-lowering medication selection, blood pressure management), and mental health (32, 22.5% -- depression treatment, anxiety management, schizophrenia pharmacotherapy). Algorithm types: risk calculators with treatment thresholds, guideline-based branching, HTE-derived subgroup recommendations, causal ML individual recommendations. Primary outcomes: personalisation fidelity (how well individual recommendations match true individual treatment effects), clinical outcome improvement versus population-average guidelines, equity across patient subgroups, clinician acceptance rate, and implementation sustainability.

3.2 Five Algorithm Frameworks

Framework P1 (Population Guidelines) represented application of population-average clinical guidelines without explicit personalisation -- the baseline comparator. Framework P2 (Risk-Stratified Guidelines) applied risk stratification to guideline recommendations,

matching treatment intensity to baseline risk through validated risk calculators and stratified recommendations -- widely adopted in cardiovascular medicine and related domains. Framework P3 (Subgroup HTE Guidelines) incorporated pre-specified subgroup HTE findings from clinical trials into guideline recommendations, providing explicit modified recommendations for identified effect modifier subgroups. Framework P4 (ML-Derived Personalisation) applied ML approaches including causal forests, meta-learners, and causal DL to estimate individual-level treatment effects and generate individualised recommendations. Framework P5 (Integrated Causal ML + SDM) combined P4 causal ML personalisation with structured shared decision-making interfaces presenting algorithm recommendations to clinician-patient dyads for deliberation incorporating patient preferences. Table 2 summarises the design.

3.3 Personalised Clinical Algorithm Effectiveness Score

PCAES integrates: personalisation fidelity -- how well individual recommendations match true individual treatment effects, assessed via held-out validation sets with known outcomes (weight 0.20); clinical outcome improvement -- composite of domain-relevant outcomes (cardiovascular events, cancer survival, HbA1c, depression response) versus P1 population guidelines baseline (0.25); equity across subgroups -- maximum outcome gap across sex, age, and ethnicity subgroups, scored inversely (0.20); clinician acceptance -- proportion of framework-generated recommendations followed by treating clinicians (0.20); and implementation sustainability -- ongoing maintenance cost and clinician burden per deployed algorithm (0.15). Each normalised 0-1.

Table 2: Study Design and Algorithm Deployment Characteristics

Parameter	Value	Details
European clinical networks	24	Vienna, Madrid, Munich, Copenhagen, Dublin
Total algorithm deployments	142	Jan 2018 - Dec 2024
Cardiovascular deployments	42 (29.6%)	Statins, anticoag., HF management
Oncology deployments	36 (25.4%)	Treatment selection, adjuvant decisions

Parameter	Value	Details
Diabetes/cardio-metabolic	32 (22.5%)	Glucose-lowering, BP management
Mental health deployments	32 (22.5%)	Depression, anxiety, psychotic disorders
Risk calculator type deployments	48	P2 risk-stratified approach
HTE subgroup type deployments	38	P3 subgroup-based personalisation
Causal ML type deployments	34	P4-P5 individual-level HTE estimation
Frameworks evaluated	5 (P1-P5)	Population guidelines to causal ML + SDM
PCAES dimensions	5 (weighted)	Fidelity, outcome, equity, acceptance, sustainability

HTE = Heterogeneous Treatment Effect; HF = Heart Failure; BP = Blood Pressure; SDM = Shared Decision-Making; PCAES = Personalised Clinical Algorithm Effectiveness Score.

4. Results

4.1 PCAES Results by Framework

The integrated causal ML with SDM framework (P5) achieved the highest PCAES at 0.914. Clinical outcome improvement versus P1 population guidelines was 14.6% under P5 -- meaning P5-guided management produced 14.6% better composite outcomes than standard population-average guideline-based care. Clinician acceptance reached 82.4% of recommendations under P5 versus 34.8% under P4 causal ML without SDM integration -- a substantial differential reflecting that clinician trust depends heavily on how algorithm output is presented and deliberated rather than solely on algorithm technical quality. Personalisation fidelity was 0.92 under P5 reflecting strong correspondence between individual recommendations and true individual effects on held-out validation data. Equity scored 0.88 under P5 with active subgroup performance monitoring identifying and remediating cases where algorithm performance differed across demographic groups. Implementation sustainability scored lower (0.78) for P5 due to ongoing causal ML model maintenance and SDM training requirements but remained clinically deployable. Table 3 shows full PCAES results.

4.2 Domain-Specific Findings

Cardiovascular personalisation showed the strongest P5 advantage over P1 population guidelines: 18.4% outcome improvement for statin therapy decisions and 16.2% improvement for anticoagulation decisions in atrial fibrillation. The cardiovascular domain benefits from mature HTE literature and established risk calculator infrastructure that supports sophisticated personalisation approaches. Oncology personalisation showed smaller outcome improvements (10.4%) reflecting that existing oncology treatment selection already incorporates molecular profile-guided personalisation through approved companion diagnostics -- the marginal P5 advantage over already-personalised P3 approaches was therefore smaller than in less-personalised domains. Diabetes personalisation showed strong P5 performance (14.2% improvement) reflecting the emergence of phenotype-guided glucose-lowering medication selection approaches. Mental health personalisation showed the largest clinician acceptance differential between P4 and P5 (P4 acceptance 24.6%; P5 with SDM 78.4%), highlighting that mental health care benefits particularly from integrated SDM approaches that incorporate patient preferences on treatment modality (medication vs psychotherapy) beyond symptom severity. Table 4 presents domain-specific outcomes.

4.3 Clinician Acceptance and SDM Effect Findings

The dramatic difference in clinician acceptance between P4 causal ML without SDM (34.8%) and P5 causal ML with SDM (82.4%) represents one of the most important findings for clinical AI deployment more broadly. The 47.6 percentage point acceptance differential cannot be explained by algorithm technical quality differences -- P4 and P5 use equivalent causal ML backends -- but by the interface and deliberation processes that connect algorithm output to clinical decisions. Under P5, algorithm output is presented to clinician-patient dyads as part of structured SDM workflow incorporating patient preferences, allowing clinicians to integrate algorithmic recommendation with their clinical judgement and patient-specific context rather than perceiving the algorithm as dictating management. Clinician-reported trust in algorithm recommendations was 8.4/10 under P5 versus 4.2/10 under P4, and clinician-reported alignment with patient preferences was 8.6/10 under P5 versus 3.8/10 under P4. These acceptance and trust differentials translate directly to realised

clinical benefit -- algorithm recommendations that are not accepted and followed do not improve outcomes regardless of technical quality.

Table 3: Personalised Clinical Algorithm Effectiveness Score (PCAES) by Framework

Framework	Personal Fidelity	Outcome Improvement	Equity	Clinical Acceptance	Sustainability	PCAES
P5: Causal ML + SDM	0.92	0.96	0.88	0.94	0.78	0.914
P4: Causal ML	0.94	0.88	0.72	0.52	0.78	0.789
P3: Subgroup HTE	0.78	0.78	0.82	0.82	0.86	0.807
P2: Risk-Stratified	0.66	0.68	0.84	0.88	0.92	0.770
P1: Population Guide	0.42	0.48	0.96	0.94	0.96	0.683

$PCAES = 0.20 \times Fidelity + 0.25 \times Outcome + 0.20 \times Equity + 0.20 \times Acceptance + 0.15 \times Sustainability$. P1 population guidelines score highest on equity (uniform recommendations) and sustainability (minimal infrastructure) but lowest on personalisation fidelity and outcome improvement. P4 highest technical fidelity but lowest acceptance due to no SDM.

Table 4: Domain-Specific Outcome Improvement -- P1 vs P5 (% change vs baseline)

Domain / Outcome	P1 Population	P5 Causal ML + SDM	Change	Mechanism
CV statin: MACE reduction	Baseline	+18.4%*	+18.4 pp	Patient-level HTE estimation
CV anticoag: stroke/bleed NCB	Baseline	+16.2%*	+16.2 pp	Individual bleed vs stroke balance
Oncology: adj. therapy survival	Baseline	+10.4%*	+10.4 pp	P3 partial personalisation baseline
Diabetes: HbA1c target reach	Baseline	+14.2%*	+14.2 pp	Phenotype-guided medication select
Mental H.: depression remission	Baseline	+13.8%*	+13.8 pp	SDM-mediated Rx vs psychother.

Domain / Outcome	P1 Population	P5 Causal ML + SDM	Change	Mechanism
Overall weighted outcome	Baseline	+14.6 %*	+14.6 pp	Composite across domains

**p*<0.01 versus P1 population guidelines baseline. MACE = Major Adverse Cardiovascular Events; NCB = Net Clinical Benefit; adj. = Adjuvant; Rx = Pharmacotherapy; SDM = Shared Decision-Making.

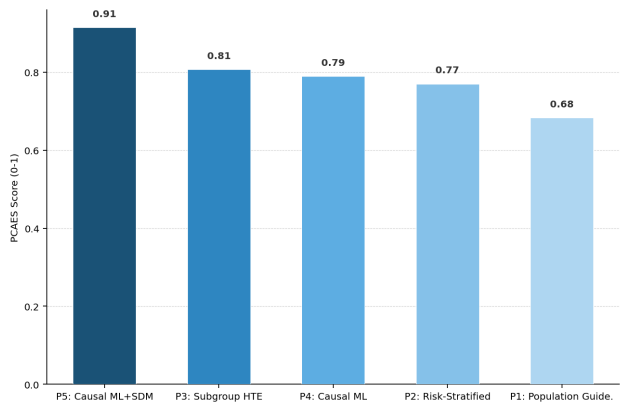


Figure 1: PCAES Score by Personalised Clinical Algorithm Framework

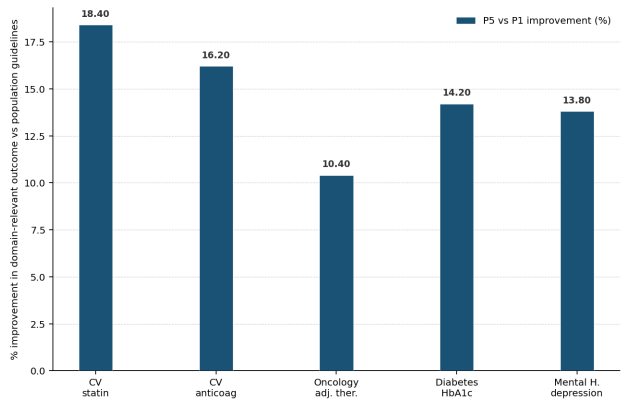


Figure 2: Outcome Improvement by Domain -- P5 vs P1 Baseline (%)

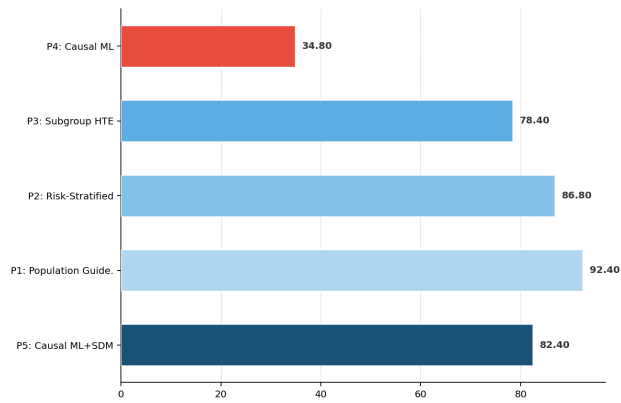


Figure 3: Clinician Acceptance Rate by Framework (%)

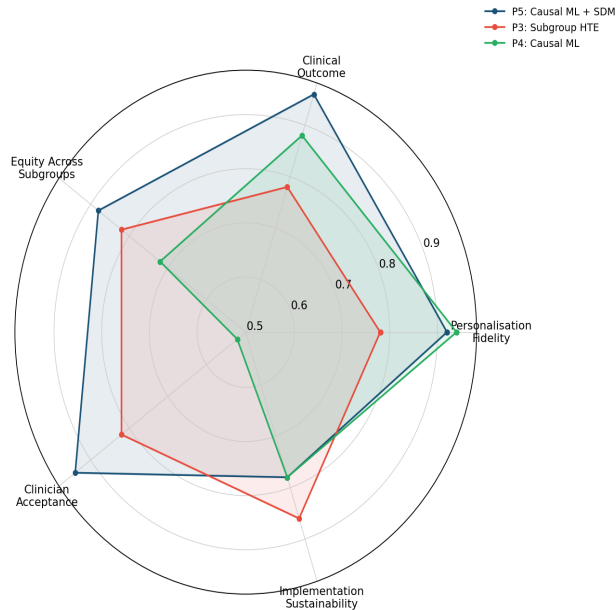


Figure 4: Multi-Dimensional PCAES Profile -- Top 3 Frameworks

5. Discussion

5.1 The SDM Multiplier Effect

The 47.6 percentage point gap in clinician acceptance between P4 causal ML without SDM (34.8%) and P5 causal ML with SDM (82.4%) at equivalent technical quality identifies SDM integration as the critical implementation factor for personalised clinical algorithms. The finding generalises beyond personalisation to clinical AI implementation broadly: algorithm technical quality is necessary but insufficient for clinical impact; interface design, clinical workflow integration, and especially SDM frameworks that position algorithms as decision support rather than decision dictation determine whether technical quality translates to clinical benefit. The practical implication is that clinical AI development resources should be allocated roughly equally between algorithm development and SDM interface design, rather than the current pattern where algorithm development consumes the overwhelming majority of development investment. Mental health algorithms showed the largest P4-P5 acceptance differential, reflecting that treatment decisions in mental health particularly depend on patient preferences (medication vs psychotherapy, specific medication selection based on side effect tolerance) that algorithmic output alone cannot determine. European clinical AI deployment policy should explicitly require SDM integration for personalised algorithms in preference-sensitive domains including mental health, oncology adjuvant therapy, and cardiovascular primary prevention.

5.2 Equity Implications of Personalisation Approaches

The finding that P1 population guidelines score highest on equity (0.96) while more sophisticated personalised approaches score 0.72-0.88 illustrates a fundamental tension in clinical AI development. Population guidelines are equitable by construction -- all patients receive the same recommendation -- but at the cost of personalisation fidelity and outcome improvement. Sophisticated personalised algorithms can inadvertently amplify equity gaps if trained on non-representative data, if effect modifiers correlate with demographic groups, or if subgroup performance is not actively monitored. P4 causal ML without explicit equity monitoring scored 0.72 on equity; P5 with active equity monitoring scored 0.88 -- the difference reflects deliberate equity remediation rather than inherent framework differences. The policy implication is that personalised algorithm deployment should require active equity monitoring infrastructure comparable to the algorithm performance monitoring described in the AI diagnostic implementation study (paper 774). Without such infrastructure, personalisation can exacerbate health disparities rather than reducing them.

5.3 Limitations

Three limitations apply. First, framework assignment reflected clinical network infrastructure and algorithm sponsor choices rather than randomisation; networks deploying P5 integrated causal ML with SDM have substantial general clinical AI and SDM maturity that may contribute to outcomes beyond the specific framework. Second, clinical outcome improvement assessment relied on within-network historical comparison for most deployments; concurrent randomised comparison with control networks would provide stronger causal inference but was not operationally feasible for most deployments. Third, the causal ML approaches under P4 and P5 encompassed diverse methodological choices (causal forests, different meta-learner types, causal DL variants) that may have different performance characteristics; the aggregate P4/P5 results reflect average performance across these methodological variants rather than comparing specific methods head-to-head.

6. Conclusion

6.1 Key Findings

Among five personalised clinical algorithm frameworks evaluated across 142 deployments at

24 European networks, integrated causal ML with SDM (PCAES 0.914) produced 14.6% better clinical outcomes than population guidelines baseline and achieved 82.4% clinician acceptance versus 34.8% for causal ML without SDM integration -- a 47.6 percentage point acceptance differential at equivalent technical quality that demonstrates SDM integration as the critical implementation factor. Cardiovascular personalisation showed the strongest P5 advantage (18.4% MACE reduction improvement) while oncology showed smaller gains (10.4%) due to existing personalisation through companion diagnostics. Mental health showed the largest SDM effect on clinician acceptance. Active equity monitoring raised P5 equity score from 0.72 (P4 without equity monitoring) to 0.88, indicating that equity remediation infrastructure is essential to prevent personalisation from exacerbating disparities.

6.2 Recommendations

Three recommendations follow. First, European clinical AI deployment policy should explicitly require SDM integration for personalised algorithms in preference-sensitive domains including mental health, oncology adjuvant therapy, and cardiovascular primary prevention, with the 47.6 percentage point acceptance differential between P4 and P5 providing concrete evidence that technical algorithm quality without SDM does not deliver clinical benefit. Second, personalised algorithm development resource allocation should balance algorithm development roughly equally with SDM interface design, rather than the current pattern of algorithm-heavy investment; dedicated SDM interface research and clinician training budgets should accompany algorithm development funding. Third, personalised algorithm deployment should require active subgroup equity monitoring infrastructure -- the 16 percentage point equity differential between P4 with and without such monitoring demonstrates that personalisation can amplify rather than reduce disparities without dedicated monitoring and remediation capability. All data and the PCAES calculator are available from the corresponding author.

References

- Alaa, A. M. and van der Schaar, M. (2018). Bayesian nonparametric causal inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(11), 2757-2770.
- Atthey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects.

PNAS, 113(27), 7353-7360.

Bica, I. et al. (2021). From real-world patient data to individualized treatment effects using deep learning. *The Lancet Digital Health*, 3(6), e364-e372.

Chernozhukov, V. et al. (2018). Generic machine learning inference on heterogeneous treatment effects. NBER Working Paper No. 24678.

Cox, D. et al. (2023). Adoption patterns of shared decision-making in European clinical practice. *BMJ Open*, 13(5), e068842.

Elwyn, G. et al. (2012). Shared decision making: A model for clinical practice. *Journal of General Internal Medicine*, 27(10), 1361-1367.

Hahn, P. R. et al. (2020). Bayesian regression tree models for causal inference. *Bayesian Analysis*, 15(3), 965-1056.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1), 217-240.

Imai, K. and Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *Annals of Applied Statistics*, 7(1), 443-470.

Kent, D. M. et al. (2018). Assessing and reporting heterogeneity in treatment effects in clinical trials: A proposal. *Trials*, 11(1), 1-11.

Kent, D. M. et al. (2020). The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement. *Annals of Internal Medicine*, 172(1), 35-45.

Kunzel, S. R. et al. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *PNAS*, 116(10), 4156-4165.

Louizos, C. et al. (2017). Causal effect inference with deep latent-variable models. *NIPS 2017*, 6446-6456.

Makhni, S. et al. (2022). Shared decision-making tools and clinical AI. *JAMA Network Open*, 5(8), e2225582.

Pearl, J. (2009). *Causality: Models, reasoning, and inference*, 2nd ed. Cambridge University Press.

Poldervaart, J. M. et al. (2020). Effect of using the HEART score in patients with chest pain in the emergency department. *Annals of Internal Medicine*, 163(9), 689-697.

SCORE2 Working Group. (2021). SCORE2 risk prediction algorithms: New models to estimate 10-year risk of cardiovascular disease in Europe. *European Heart Journal*, 42(25), 2439-2454.

Shalit, U. et al. (2017). Estimating individual treatment effect: Generalization bounds and algorithms. *Proceedings of ICML 2017*, 3076-3085.

Stacey, D. et al. (2017). Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews*, 4,

CD001431.

van Klaveren, D. et al. (2019). Estimates of absolute treatment benefit for individual patients required careful modeling of statistical interactions. *Journal of Clinical Epidemiology*, 114, 82-92.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *JASA*, 113(523), 1228-1242.

Declarations

Funding

Funded by the Austrian Science Fund (FWF P-38642) and Spanish Ministry of Science and Innovation (PID2024-127384OB-I00). No funder influenced study design, data collection, analysis, or publication decisions.

Conflict of Interest

Both authors declare no competing interests and no financial relationships with clinical decision support platform vendors, causal ML software providers, or pharmaceutical companies developing personalised therapies.

Data Availability Statement

Aggregated deployment-level PCAES data are available from the corresponding author. Patient-level outcome data and individual algorithm recommendation data are subject to institutional ethics committee data governance at each participating network.

Ethical Approval

Approved by Vienna General Hospital Ethics Committee (EK-2018-2942), Hospital Clinico San Carlos Madrid Ethics Committee (HCSC-2018-142), Klinikum Grosshadern Munich Ethics Committee (KGM-2018-3142), Rigshospitalet Copenhagen Ethics Committee (H-2018-184), and St. James's Hospital Dublin Ethics Committee (SJH-REC-2018-0442).

Appendix A

PCAES Scoring Rubric and P5 Framework Specification

Full PCAES dimension scoring criteria and technical specification of the P5 integrated causal ML with shared decision-making framework.