

<https://doi.org/10.1038/s41746-025-01797-7>

# Predicting quality measure completion among 14 million low-income patients enrolled in medicaid

Sadiq Y. Patel<sup>1,2</sup>✉, Michael L. Barnett<sup>3,4</sup> & Sanjay Basu<sup>1,5</sup>

Low-income populations have disproportionately low completion of recommended healthcare services, from missed vaccinations to cancer screenings. While machine learning models help identify high-risk patients for targeted treatment, they have rarely been evaluated for quality measure gap completion—or among low-income populations underrepresented in typical datasets. Analyzing 14.2 million Medicaid recipients—including those excluded from electronic health records and without prior utilization—we developed models to predict gaps in nine nationally adopted quality measures, including preventive care and chronic disease management. Using clinical data to prioritize outreach, the clinical-only model improved accuracy by 32.5 percentage points (pp) over non-predictive methods such as alphabetical calling or birthday reminders (AUROC: 0.88, F1-score: 0.69). Incorporating social determinants of health data further improved performance by 2.0pp in accuracy (to 84.5%) and increased F1-score by 5.0pp (to 0.74), with no change in AUROC (area under the receiver operating characteristic curve). Compared to the clinical-only model, the SDoH model also reduced pre-existing Black–White disparities in prediction accuracy. Model performance was especially sensitive to SDoH factors like healthcare workforce and facility availability.

Decades of evidence have established that healthcare quality—from providing routine vaccinations to ensuring age-appropriate cancer screening—remains systematically worse among low-income versus higher-income patient populations<sup>1,2</sup>. Healthcare quality is an outcome shaped by structural exclusion, patient behavior, provider decisions, and systemic barriers, rather than clinical need alone. The systematically worse quality metrics among low-income populations are particularly notable among the 80 million people serviced by Medicaid, the public health care insurance program for low-income children and adults in the US, where quality care metrics are up to 50% worse than in commercially- or Medicare-insured populations<sup>3</sup>. The consequences of poor quality care are profound, contributing to persistent health disparities between socioeconomic and race/ethnic groups, and preventable morbidity and mortality among vulnerable populations<sup>4</sup>.

Recent reports from the National Academy of Medicine (NAM) have asked whether, and to what extent, data science technologies can improve, rather than worsen, quality—potentially by predicting which patients are most likely to get poor quality care, and helping care teams to proactively outreach and engage those patients<sup>5,6</sup>. The NAM reports highlighted several

unaddressed questions for data scientists working on these population health problems: First, can machine learning models meaningfully predict which patients are most likely to receive quality care, as judged by ‘closure’ of ‘care quality gaps’ such as missed vaccinations, cancer screening, or receipt of guideline-recommended medications? Second, what is the added value of incorporating social determinants of health (SDoH) data into these predictive models, particularly given the limited SDoH data available and the fact that such data are often at the area-level rather than individual-level? Finally, how can model-based targeting improve equity by prioritizing outreach to patients facing greater structural disadvantage, rather than worsening equity by reproducing existing inequalities in care access and quality?

Many current outreach strategies for addressing care quality gaps are based on logistical or administrative rules—such as contacting patients in alphabetical order or around birthdays or insurance renewal periods—rather than on predictive models<sup>7–15</sup>. Fewer than 8% of eligible Medicaid patients benefit from these efforts<sup>16,17</sup>. While such targeting approaches are simple to implement, they fail to account for which patients may be more

<sup>1</sup>Clinical Product Development, Waymark, San Francisco, CA, USA. <sup>2</sup>School of Social Policy and Practice, University of Pennsylvania, Philadelphia, PA, USA.

<sup>3</sup>Department of Health Policy and Management, Harvard T. H. Chan School of Public Health, Boston, MA, USA. <sup>4</sup>Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA. <sup>5</sup>Center for Vulnerable Populations, University of California, San Francisco, CA, USA. ✉e-mail: [sadiq.patel@waymarkcare.com](mailto:sadiq.patel@waymarkcare.com)

likely to close care quality gaps on their own, and which may need additional assistance to receive such care. They also ignore the well-documented influence of social determinants of health (SDoH)—such as transportation access and healthcare facility availability—on preventive care use among low-income populations, despite substantial investments to address these factors<sup>18–20</sup>.

Although machine learning has shown promise for identifying higher-risk patients already engaged with the healthcare system<sup>21</sup>, their potential benefits for preventive care and outreach to populations with rare and sporadic healthcare use, who are typically under-represented in electronic health data, remains unclear<sup>6</sup>. Low-income populations are under-represented in the datasets typically used to develop and validate machine learning models<sup>6</sup>.

Using administrative claims data from over 14 million Medicaid beneficiaries in the Transformed Medicaid Statistical Information System (T-MSIS)—which includes individuals with minimal or no interaction with traditional healthcare systems or electronic health records—we developed and validated machine learning models to predict receipt of nine HEDIS quality measures—including both preventive services (e.g., prenatal visits, well-child care) and condition-specific care (e.g., medication adherence, readmissions)—that are used in Medicaid performance evaluations. Our study had three aims: first, to quantify the ability of machine learning approaches to predict quality gap closure among Medicaid populations; second, to assess the added predictive value and equity implications of incorporating SDoH data into such models; and third, to explore, through model-based simulations, the magnitude of quality gap closure that could hypothetical occur with improvements to SDoH factors as illustrative counterfactual simulations. This approach helps link predictive modeling to actionable outreach strategies by estimating how model-guided targeting could shift outreach efficiency. While our simulations illustrate potential gains under improved social conditions, they are not causal estimates and should be interpreted as exploratory tools to prioritize areas for future intervention.

## Results

### Patient Characteristics

The study cohort comprised 14,178,331 Medicaid beneficiaries from 25 states and Washington D.C. who met inclusion criteria for at least one of the nine quality measures (Table 1). The population was 54.6% female and 53.7% under 18 years of age, with 62.0% residing in the West census region. The racial/ethnic distribution included 32.0% Hispanic, 30.2% non-Hispanic White, and 15.8% non-Hispanic Black participants, with 14.0% missing race/ethnicity data. Among participants, 20.8% lived below the federal poverty level, and 6.4% reported a disability.

Social determinants of health varied across the 1563 counties in our sample. County-level poverty rates ranged from 4.1% to 7.6% (mean 6.3%, interquartile range [IQR]). Environmental measures such as the percentage of days with good air quality per U.S. Air Quality Index criteria averaged 86.9% (IQR: 84.5–93.9%). The prevalence of quality measure gaps ranged from 7.3% (IQR: 6.7–7.9%) for all-cause 30-day hospital readmissions to 16.2% (IQR: 8.8–19.8%) for beta-blocker persistence after myocardial infarction (baseline rates in Supplement Methods).

### Machine Learning Model Performance

Models using only demographic and clinical predictors achieved a mean accuracy of 82.5% (range: 64.8–92.9%), mean sensitivity of 59.2% across the nine quality measures (range: 1.6–86.3%), and mean specificity of 88.2% (range: 44.2–99.9%). Performance varied by measure type without consistent patterns. For example, while statin adherence prediction for diabetes patients showed 86.7% accuracy, the same prediction for cardiovascular disease patients reached 84.0% accuracy. In maternal and child health visit measures, accuracy was 84.6% for prenatal visits, 81.0% for postpartum visits, and 64.8% for well-child visits (Fig. 1, Supplementary Table 3).

The addition of social determinants of health variables increased mean accuracy by 2.0 percentage points (pp) and sensitivity by 4.6 pp while

maintaining specificity (Fig. 1, Supplementary Table 4). Accuracy changes ranged from -2.9 pp for statin adherence prediction for diabetes patients to +8.1 pp for follow-up after emergency department visits for mental illness, sensitivity improvements from 0.2 pp for hospital readmissions to 15.9 pp for diabetes statin therapy. Key improvements in accuracy included increases of 4.9 pp for continuation of antidepressant medication, 4.3 pp for diabetes statin therapy, 3.3 pp for prenatal visits, and 2.3 pp for postpartum visits. Model specificity increased by a mean of 1.0 pp (range: -2.3 to +8.6).

Compared to a baseline simulating typical non-predictive outreach strategies—such as alphabetical calling or reminders based on enrollment dates or birthdays—the social determinants model demonstrated absolute improvements of 34.5 pps in accuracy (to 84.5% accuracy), 9.8 pp in sensitivity (to 59.8% sensitivity), and 39.3 in specificity (to 89.3% specificity; Supplementary Table 5). The clinical-only model showed improvements of 32.5, 5.2, and 38.2, respectively (to 82.5% accuracy, 55.2% sensitivity and 88.2% specificity).

In simulated care gap closure, the social determinants model achieved quality measure completion rates of 66.5% to 95.3% (mean of 82.5%) versus 28.3% to 91.9% (mean of 61.1%) for random targeting, representing absolute improvements of 6.3 to 55.0 pp (mean 26.5) from using the SDoH-enhanced model. The model simulation demonstrated a reduction in outreach attempts required per gap closed—from 5.1–17.7 to 5.1–7.5—assuming a standard 20% success rate per attempt. Under a more conservative 10% success rate, the required attempts decreased from 10.9–35.3 to 10.2–15.0 (Supplement Table 6).

### Variable Importance Analysis

Individual-level social factors had higher predictive importance than county-level measures across all outcomes (Fig. 2). Mean Gini importance scores for individual factors ranged from 0.223 (maternal/child health) to 0.754 (unnecessary care). Supplemental Security Income receipt showed the highest importance (mean: 0.426, median: 0.402), followed by household income (mean: 0.400, median: 0.394).

County-level factors had particular relevance for medication adherence (importance 0.223 for beta-blockers) and care coordination (mean: 0.155, median: 0.150), with lower importance for behavioral health (mean: 0.090), chronic disease management (mean: 0.087), maternal/child health (mean: 0.064), and unnecessary care (mean: 0.031). Specific area-level factors including provider supply, poverty rates, and facility availability showed importance scores of 0.072–0.082 (complete rankings in Supplementary Table 7).

Top features contributing to improved accuracy included individual income, reliance on public benefits, and county-level healthcare facility density. The model prioritized outreach to patients facing greater social vulnerability, rather than excluding lower-risk individuals from outreach.

### Model Sensitivity to Social Determinant Improvements

We employed a model-based simulation to explore how predicted probabilities of quality measure completion might change under hypothetical improvements in social determinants (Fig. 3). These simulations do not estimate causal effects but reflect illustrative counterfactual scenarios with alternative inputs. Reductions in county-level poverty produced the largest effect on predicted gap closure, with a mean 3.4% (median: 1.9%) relative increase in predicted quality gap closure (mean 0.43 pp improvement). Other impactful area-level changes included increased advanced practice provider supply (mean: 2.5%, median: 1.4%, 0.27 mean pp), mental health facility availability (mean: 2.9%, median: 1.2%, 0.20 mean pp), and high school completion rates (mean: 2.5%, median: 1.2%, 0.27 mean pp).

Analysis of simultaneous improvement in all social factors revealed the largest gains in measures of care overuse and adherence (Fig. 4, Supplementary Table 8). Unnecessary imaging for low back pain showed a 44.6% improvement (2.1 pp). Medication adherence measures improved by 24.5% (2.4 pp) for cardiovascular statin therapy, 22.6% (1.5 pp) for diabetes statin therapy, and 21.2% (1.6 pp) for post-heart attack beta-blockers. Prenatal visit completion increased by 12.5% (1.4 pp).

**Table 1 | Characteristics of patients receiving Medicaid in the study data, 2017–2019 (*n* = 14,178,331)\*^**

Patient-level	N	14178331
Age Category	Under 10	4202864 (29.6)
	10–17	3411572 (24.1)
	18–29	2199669 (15.5)
	30–39	1591109 (11.2)
	40–49	1171183 (8.3)
	50–64	1601268 (11.3)
	Missing	666 (< 0.1)
Sex	Female	7740904 (54.6)
	Male	6436593 (45.4)
	Missing	834 (< 0.1)
Race/ Ethnicity	Asian	764037 (5.4)
	White	4282688 (30.2)
	Multiracial	26588 (0.2)
	Black	2233315 (15.8)
	Hispanic	4531363 (32.0)
	Native American	250609 (1.8)
	Hawaiian	110971 (0.8)
	Missing	1978760 (14.0)
Census Region^	Midwest	1714996 (12.1)
	South	1278484 (16.9)
	West	2400899 (62.0)
	Northeast	8784022 (9.0)
	Missing	0 (0.0)
Household Size	Single	2042659 (14.4)
	5-Feb	1993422 (14.1)
	6 or more	751695 (5.3)
	Missing	9390555 (66.2)
Federal Poverty Line	Under 100%	2955382 (20.8)
	100–200%	489008 (3.4)
	200% or more	27732 (0.2)
	Missing	10706209 (75.5)
English Speaking	Yes	1206336 (8.5)
	No	101663 (0.7)
	Missing	12870332 (90.8)
Married	Yes	283934 (2.0)
	No	4819924 (34.0)
	Missing	9074473 (64.0)
US Citizen	Yes	11374348 (80.2)
	No	878426 (6.2)
	Missing	1925557 (13.6)
Receipt of SSI	Yes	989811 (7.0)
	No	12046639 (85.0)
	Missing	1141881 (8.1)
Receipt of SSDI	Yes	542471 (3.8)
	No	9714085 (68.5)
	Missing	3921775 (27.7)
Receipt of TANF	Yes	2202919 (15.5)
	No	8773627 (61.9)
	Missing	3201785 (22.6)

**Table 1 (continued) | Characteristics of patients receiving Medicaid in the study data, 2017–2019 (*n* = 14,178,331)\*^**

Patient-level	N	14178331
Disabled	Yes	905739 (6.4)
	No	13272592 (93.6)
	Missing	0 (0.0)
County-level	N	1563
Number of substance abuse services facilities accepting Medicaid per 100,000 population	Mean (IQR)	6.0 (0.0, 7.5)
Number of Mental health services that accept Medicaid patients per 100,000 population	Mean (IQR)	6.3 (0.9, 7.4)
Number of APRNs/PAs per 100,000 population	Mean (IQR)	2.2 (1.3, 2.8)
Number of urgent care organizations per 100,000 population	Mean (IQR)	1.3 (0.0, 2.0)
Percentage of households with public assistance income or food stamps/SNAP	Mean (IQR)	13.0 (8.7, 16.0)
Population per square mile	Mean (IQR)	232.3 (12.0, 107.8)
Percentage of population with income to poverty ratio under 0.50	Mean (IQR)	6.3 (4.1, 7.6)
Percentage of population with less than high school education	Mean (IQR)	11.6 (7.4, 14.4)
Percentage of days with good air quality	Mean (IQR)	86.9 (84.5, 93.9)

SSI social security income, SSDI social security disability income, TANF temporary assistance for needy families, APRN advanced practice registered nurse, PA physician assistant, SNAP Supplemental Nutrition Assistance Program.

\*Patient-level data comes from Chronic Conditions Warehouse (CCW) Virtual Research Data Center (VRDC). Community-level SDoH data comes from the AHRQ SDoH Database.

^We reported column percentages for all patient characteristics. Patients were enrolled for 36 months from 2017 to 2019. All patient characteristics were derived from the TAF data. States were mapped to their respective Census Regions.

### Model Bias and Sensitivity Analyses

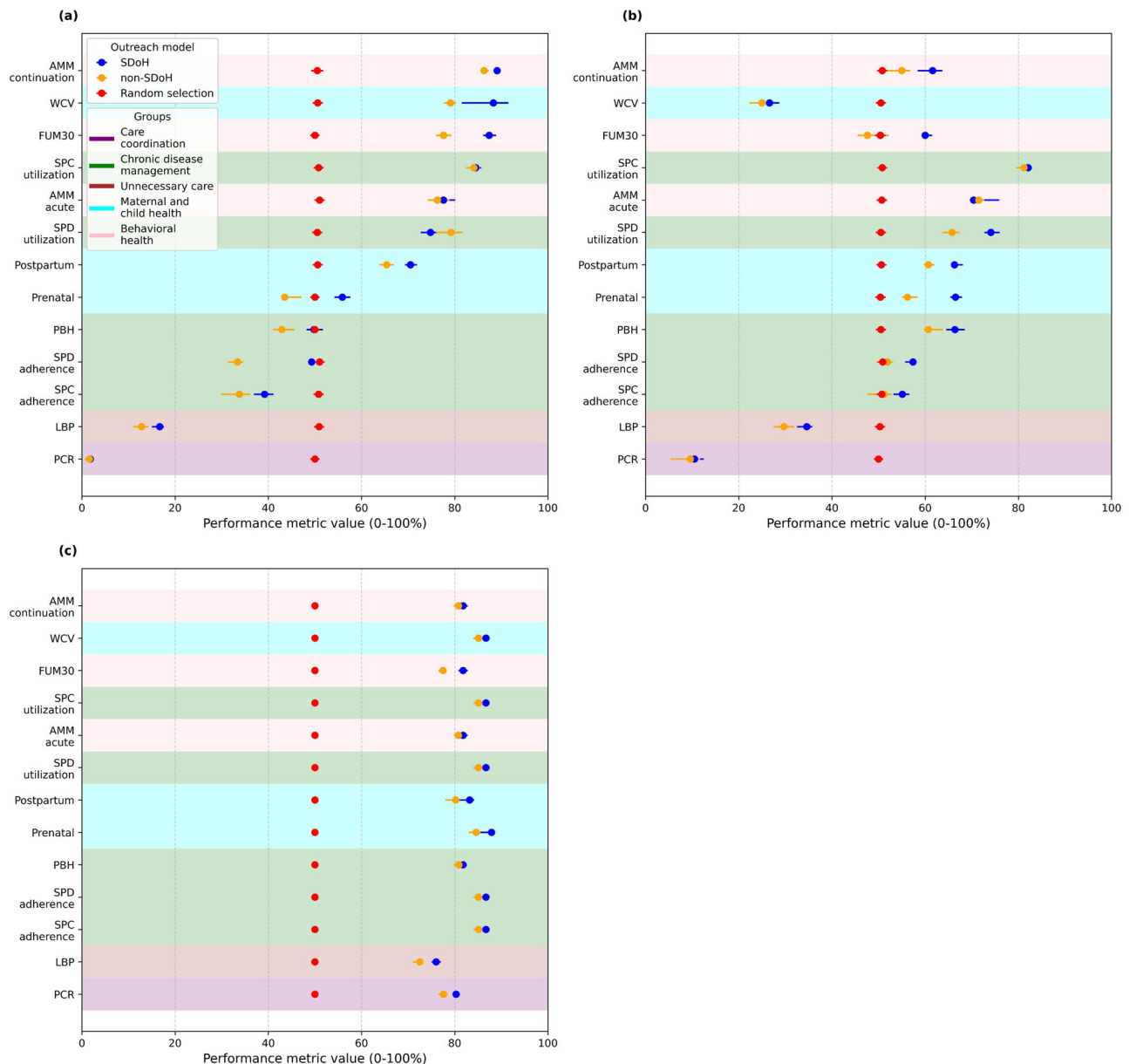
The social determinants model demonstrated equal or higher sensitivity for Black versus White patients across most measures, with prenatal visits as the sole exception. Hispanic patients showed higher sensitivity than White patients in 7 of 13 measures. The addition of social determinants data eliminated pre-existing Black-White sensitivity disparities in four measures where the clinical-only model had shown disparities (Supplementary Tables 9–10). False positive and false negative rates by race/ethnicity are presented in Supplementary Table 9. We observed comparable subgroup performance in these measures.

Sensitivity analyses revealed even distribution of quality measure completion between first and second half-year periods (Supplementary Table 11). Comparison of 36-month continuous enrollees versus one-month minimum enrollees showed no significant demographic differences (Supplementary Table 12), supporting broader generalizability of the findings.

Model performance varied across utilization strata (Supplementary Table 13). Prenatal, postpartum, AMM, and SPD1 models showed stable AUC and F1 scores across tiers, while LBP, PCR, and SPD2 models exhibited reduced sensitivity among low-utilization patients, despite high specificity.

### Discussion

In this analysis of over 14 million Medicaid beneficiaries across 25 states and Washington D.C., we quantified the potential for machine learning



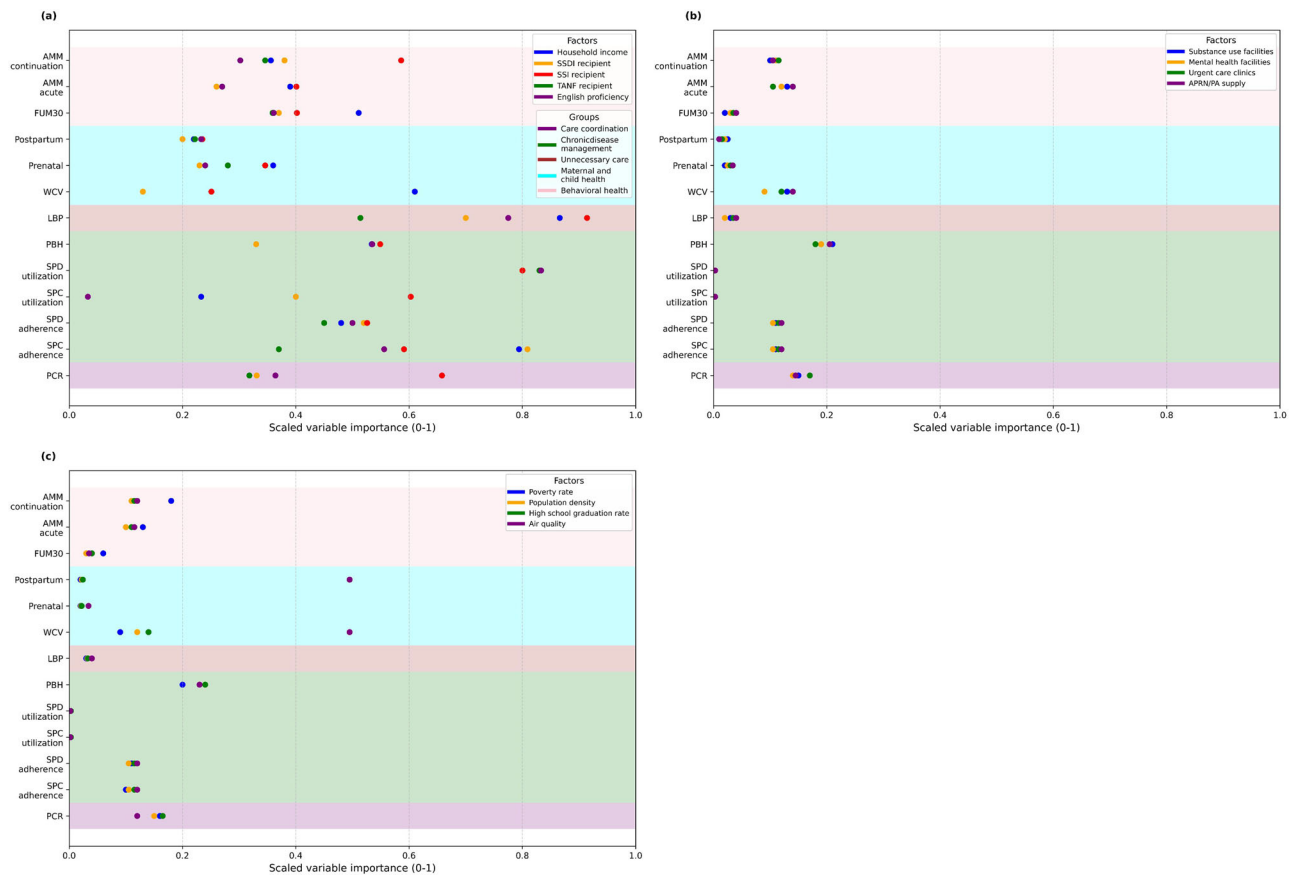
**Fig. 1 | Comparative effectiveness of non-SDoH, SDoH, and random selection models in predicting quality care gaps: sensitivity, specificity, and accuracy.** XGBoost (extreme gradient boosting), AMM (antidepressant medication management), PBH (persistence of beta-blocker treatment after a heart attack), SPC (statin therapy for patients with cardiovascular disease), SPD (statin therapy for patients with diabetes), PCR (all-cause hospital readmissions), LBP (avoidance of unnecessary imaging for routine lower back pain), FUM30 (follow-up after emergency department visits for mental illness), PPC (prenatal and postpartum care visits),

WCV (child and adolescent well-care visits). Model sensitivity (panel a), specificity (panel b), and accuracy (panel c), ordered by decreasing SDoH model performance. Results are grouped by the type of quality care gap. Predictors were measured in 2017, and quality outcomes were assessed in 2018 for measures requiring one year of data, and in 2018–2019 for those requiring two years. All patients had 36 months of continuous Medicaid enrollment. Models were developed using XGBoost for both non-SDoH and SDoH inputs.

approaches to predict quality care gap closure among populations traditionally underrepresented in healthcare analytics. Given that non-receipt of care is inherently shaped by structural inequities, we built a study sample using Medicaid eligibility files that include patients without a history of electronic health records or healthcare utilization. We found that machine learning models improved identification of patients at risk of low care quality in both preventive and therapeutic measures compared to random targeting approaches, with performance varying across different types of quality measures, including preventive care and chronic disease management—and eliminated several pre-existing Black-White racial biases in predictive model accuracy. Integrating social determinants of health (SDoH) data typically improved model predictive accuracy and sensitivity

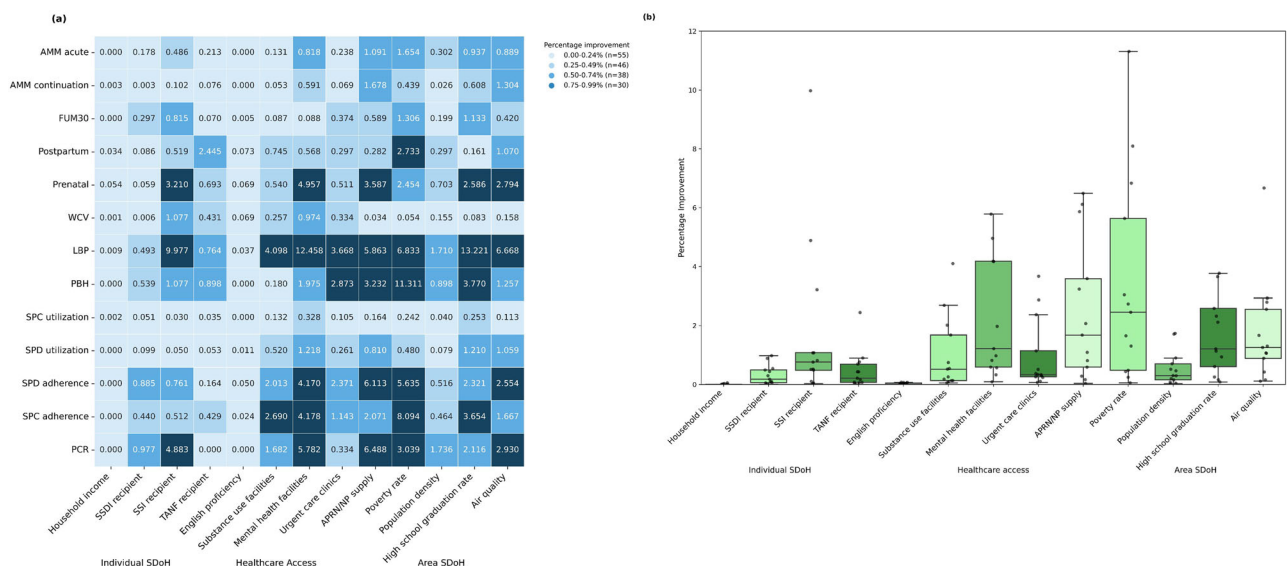
without compromising specificity. Model predictions were particularly sensitive to healthcare workforce availability and facility access variables among the SDoH variables, particularly for prenatal visit completion and medication adherence for cardiovascular disease prevention.

These findings have direct implications for resource allocation within Medicaid, which provides healthcare coverage to over one in five Americans and nearly half of U.S. births<sup>3</sup>. As Medicaid agencies and managed care organizations develop outreach strategies to improve performance on quality measures<sup>7–17</sup>, our analyses provide an evidence-based framework for patient prioritization. The incorporation of SDoH data serves two functions: improving predictive accuracy and identifying actionable intervention points for addressing systemic barriers to care. Importantly, the model is



**Fig. 2 | SDoH variable importance in prediction of quality care gaps.** APRN (advanced practice registered nurse), TANF (Temporary Assistance for Needy Families), SSI (Supplement Security Income), SSDI (Social Security Disability Insurance), AMM (antidepressant medication management), PBH (persistence of beta-blocker treatment after a heart attack), SPC (statin therapy for patients with cardiovascular disease), SPD (statin therapy for patients with diabetes), PCR (all-cause hospital readmissions), LBP (avoidance of unnecessary imaging for routine lower back

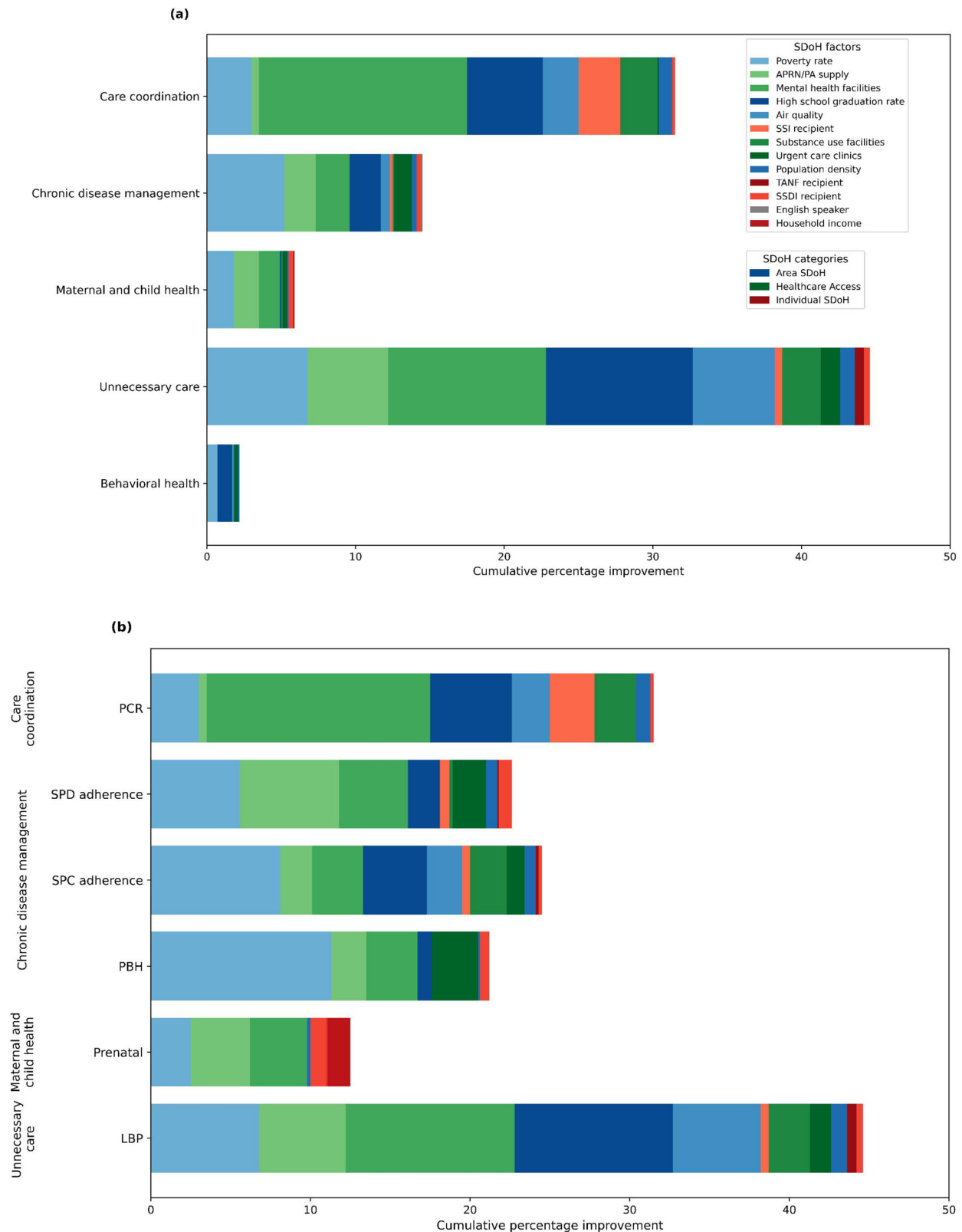
pain), FUM30 (follow-up after emergency department visits for mental illness), PPC (prenatal and postpartum care visits), WCV (child and adolescent well-care visits). Variable importance for predicting quality care gaps: individual-level SDoH (panel a), healthcare access SDoH (panel b), and area-level SDoH (panel c). Variable importance values range from 0 to 1, with higher values indicating greater contribution to model predictions.



**Fig. 3 | Impact of changing one SDoH variable at a time on quality care gap completion.** APRN (advanced practice registered nurse), TANF (Temporary Assistance for Needy Families), SSI (Supplement Security Income), SSDI (Social Security Disability Insurance). Impact of individual SDoH improvements on quality care gap completion: heatmap (panel a) and boxplot by SDoH factor (panel b).

Percentage improvement reflects the relative change in predicted gap closure after improving a single SDoH variable, calculated as the difference between post-improvement and baseline predictions, divided by the baseline number of predicted gaps.





intended to support population-level outreach planning and resource prioritization, not to diagnose individual unmet need or determine clinical urgency. Rather, it offers a scalable framework for identifying groups at elevated risk of non-receipt of care, enabling more inclusive and efficient outreach strategies. The magnitude of potential improvement—particularly in medication adherence and appropriate imaging utilization—suggests

specific opportunities for intervention. The model identified disproportionate outreach to Medicaid recipients who are further from resources, affected by more profound levels of poverty, and have the lowest access to education. Low outreach would result in lower utilization in this subset. These barriers include inflexible work hours, limited transportation, and fragmented care coordination, which are common among Medicaid

**Fig. 4 | Cumulative impact of SDoH improvements on quality care gap completion.** APRN (advanced practice registered nurse), TANF (Temporary Assistance for Needy Families), SSI (Supplement Security Income), SSDI (Social Security Disability Insurance), AMM (antidepressant medication management), PBH (persistence of beta-blocker treatment after a heart attack), SPC (statin therapy for patients with cardiovascular disease), SPD (statin therapy for patients with diabetes), PCR (all-cause hospital readmissions), LBP (avoidance of unnecessary imaging for routine lower back pain), FUM30 (follow-up after emergency department visits for mental illness), PPC (prenatal and postpartum care visits), WCV (child and adolescent well-care visits). Cumulative impact of SDoH improvements on quality care

gap completion: percentage improvement by quality measure type (panel a) and by care gaps with  $\geq 10\%$  improvement (panel b). The x-axis represents the relative cumulative improvement in care gap closure after sequentially improving SDoH factors, compared to the original dataset. Improvements were calculated as the percentage reduction in predicted non-compliance. SDoH improvements are ordered based on descending median impact, as shown in Fig. 3b. Figure 3b illustrates the individual contribution of each SDoH variable to gap closure. Color coding distinguishes SDoH categories: area-level factors (blue), healthcare access factors (green), and individual-level factors (red).

recipients. However, additional implementation research is needed to determine which strategies are most effective in real-world settings. Although XGBoost performed well, the trade-offs between model complexity and interpretability are critical. Additional research should evaluate whether simpler models yield comparable performance with greater transparency for implementation.

The mitigation of algorithmic bias we observed represents a critical advancement in healthcare predictive analytics. Our SDoH-enhanced models demonstrated equal or greater sensitivity for prediction of gap closure among Black adults compared to White adults across most outcome measures, reversing pre-existing racial disparities in four measures. This finding indicates that incorporating social factors can promote more equitable care delivery<sup>22–24</sup>. However, the persistent disparity in prenatal visit prediction accuracy between Black and White patients requires attention, particularly given documented racial disparities in maternal health outcomes<sup>25</sup>. This limitation emphasizes the need for continued refinement of predictive models to advance health equity.

Several limitations merit consideration. Although our analysis included a large and diverse Medicaid population, the exclusion of states with inadequate data quality affects generalizability. People entirely excluded from Medicaid and other forms of insurance due to structural inequities may differ in unobserved ways. Our pre-pandemic analysis window also limits generalizability to current delivery conditions and requires validation with more recent data. We excluded dually-eligible Medicare-Medicaid beneficiaries due to their distinct care patterns and programs. Finally, while our sensitivity analyses support generalizability across enrollment duration, the influence of unobserved care management programs—though currently limited to fewer than 8% of patients<sup>26</sup>—remains a consideration. Our model focused on quality measures commonly included in Medicaid state with-hold contracts, excluding preventive screenings such as cervical and colorectal cancer due to inconsistent laboratory, radiology, procedural, or pathology data, which may underreport bundled or public health-provided services. Lastly, while our model incorporates a broad array of social determinants, the availability of individual-level data was limited. Key drivers of inequity—including interpersonal discrimination, historical exclusion, housing instability, and medical mistrust—were not captured in our dataset and remain essential priorities for future research<sup>27,28</sup>. Importantly, the outcome modeled—non-receipt of care—is a reflection of observed behavior and system dynamics, not necessarily a proxy for unmet clinical need or expected benefit from outreach. This distinction is critical to avoid overinterpreting model outputs as measures of individual necessity or worthiness of care.

However, we also note a key equity limitation: while Medicaid claims data allow us to include populations historically excluded from EHR-based models, the prediction of care gap closure still relies on prior utilization patterns. Individuals with minimal healthcare use—despite often facing the greatest structural barriers—may have weaker or noisier data signals, leading to lower model sensitivity in this group. As such, the model may inadvertently prioritize outreach to individuals whose patterns are more predictable rather than those with the greatest unmet need. While this approach represents a step forward in inclusivity, further efforts are needed to ensure outreach strategies reach those facing the most profound systemic exclusion.

Future research priorities include evaluating model performance in implementation settings, translating predictive insights into targeted interventions addressing both clinical and social barriers, and adapting these models for specific vulnerable populations, including dual-eligible beneficiaries. Further, future work should explore causal methods to estimate the real-world impact of structural interventions and validate model performance in operational settings. These efforts will be essential for realizing the potential of predictive analytics to enhance quality measure performance and advance health equity in Medicaid populations.

This study provides quantitative evidence for machine learning's capacity to optimize quality measure targeting among Medicaid beneficiaries. By demonstrating both the predictive value of SDoH-enhanced models and estimating the potential impact of addressing social barriers to care, our model comparisons can inform more evidence-based allocation of outreach resources to improve quality measure performance, while recognizing that actual intervention effects must be validated through implementation studies.

## Methods

### Data Sources and Study Oversight

We used Medicaid eligibility files, which contain all people covered by Medicaid whether or not they have received healthcare, to predict non-receipt of care, specifically non-closure of HEDIS quality gaps<sup>29</sup>. We specifically analyzed data from the Transformed Medicaid Statistical Information System Analytic Files (TAF) spanning 2017–2019<sup>30</sup>. The TAF data include patient demographics, eligibility information, individual-level social determinants of health metrics (e.g., Temporary Assistance for Needy Families recipient status, household income; described in detail in Supplementary Note 1), geographic information (county of residence), and comprehensive claims data for outpatient, inpatient, long-term support, pharmacy, and other healthcare services, encompassing both fee-for-service and managed care. We included data from states meeting minimum quality standards defined by Medicaid.gov's Data Quality Atlas during the study period<sup>31</sup>. State-level enrollment benchmarks, claims volume, and data completeness were assessed to ensure data quality (detailed quality criteria in Supplementary Note 2). The final analytic sample comprised 14,178,331 Medicaid beneficiaries residing across 1563 counties within 25 states and Washington, D.C. We obtained community-level social determinants of health data from the Agency for Healthcare Research and Quality (AHRQ) Social Determinants of Health Database<sup>32</sup>. This study adhered to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines (Supplementary Table 1)<sup>33</sup>.

### Ethics Approval and Consent to Participate

This study utilized de-identified administrative claims data from the Transformed Medicaid Statistical Information System Analytic Files (TAF) spanning 2017–2019. The research protocol was reviewed and approved by the Western Institutional Review Board (Princeton, New Jersey), which granted a waiver of informed consent due to the retrospective nature of the study and the use of de-identified data. All procedures were conducted in accordance with the ethical standards of the institutional and national research committees and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

## Study Population and Follow-up

The study population included all Medicaid beneficiaries who met the standard national inclusion and exclusion criteria for at least one of the nine selected quality measures—not only the subset of patients with claims or electronic health record data. Most measures required 36 months of continuous Medicaid enrollment from 2017–2019. To assess potential selection bias from this requirement, we conducted a sensitivity analysis comparing the demographics of beneficiaries with 36 months of continuous enrollment to those with at least one month of enrollment in 2017 (results in Supplementary Table 12). We excluded beneficiaries dually enrolled in both Medicare and Medicaid, as Medicare serves as the primary payer for these individuals, resulting in potentially incomplete medical claims in TAF. Additionally, dual-eligible beneficiaries typically receive separate care management services with different outreach protocols.

## Outcomes

We assessed quality of care using the national Healthcare Effectiveness Data and Information Set (HEDIS) measures<sup>2</sup>. HEDIS comprises a standardized set of evidence-based performance measures encompassing a range of recommended services, from cancer screenings to medication adherence for chronic conditions. Our study focused on predicting non-completion of a HEDIS quality measure—specifically, the probability a patient did not receive a recommended service based on their age, biological sex, and medical history. Detailed definitions of inclusion and exclusion criteria for each measure, along with specific calculation methods following National Committee for Quality Assurance (NCQA) guidelines<sup>39</sup>, are provided in Supplementary Note 3–4.

We developed separate prediction models for nine HEDIS measures chosen based on three criteria: inclusion across state Medicaid agency quality assessments<sup>34</sup>, relevance to diverse Medicaid patient populations (pediatrics, pregnant women, older adults), and coverage of multiple quality domains (prevention, treatment, and avoidance of low-value care). The measures were: (1) child and adolescent well-care visits (WCV); (2) prenatal and postpartum care visits (PPC); (3) follow-up after emergency department visits for mental illness (FUM30); (4) avoidance of unnecessary imaging for routine lower back pain (LBP); (5) all-cause hospital readmissions (PCR); (6) antidepressant medication management (AMM); (7) persistence of beta-blocker treatment after heart attack (PBH); (8) statin therapy for patients with cardiovascular disease (SPC); and (9) statin therapy for patients with diabetes (SPD). We focused on the subset of quality measures included most commonly in Medicaid state withhold (financial penalty) specifications within contracts to health plans. These do not include some preventive screening measures—such as lead, cervical, and colorectal cancer—due to the need for laboratory, pathology, radiology, or procedural data that are only available from a biased subset of patients who have access to such services. Our goal was to ensure inclusion of patients who may have poor access to the healthcare system, thereby enhancing generalizability, and align with state policymaker priorities for quality metrics at a population level.

To validate our HEDIS measure coding and ensure alignment with standard practice, we compared aggregate state-level results from our data with publicly available reports on HEDIS outcomes among Medicaid populations from the NCQA (detailed validation results in Supplementary Note 5). To reflect the heterogeneity of these metrics, we refer to them as ‘quality measures’ throughout this manuscript, with a subset related to primary prevention.

## Predictor Variables

We constructed a comprehensive set of predictor variables from the TAF data, encompassing demographics, diagnoses, therapeutics, healthcare utilization, and social determinants of health factors. Demographic variables included age, sex, race/ethnicity (included to assess potential effects of structural racism on quality measure completion), and state of residence (using fixed effects to control for unmeasured state-level variation). We captured clinical information using standardized coding systems: Clinical

Classifications Software Refined (CCSR) for diagnoses<sup>35</sup>, Restructured Berenson-Eggers Type of Service (BETOS) for types of care<sup>36</sup>, Centers for Medicare & Medicaid Services (CMS) specialty classifications for provider specialties<sup>37</sup>, and CMS Prescription Drug Data Collection codes for medications<sup>38</sup>.

We quantified healthcare utilization through multiple metrics: counts of acute care visits (emergency department visits and hospitalizations), including ambulatory-sensitive conditions identified through the NYU Emergency Department algorithm and AHRQ Prevention Quality Indicators<sup>39,40</sup>. These methods allowed us to distinguish between emergent and non-emergent encounters, capturing both high-acuity episodes and outpatient-manageable conditions such as respiratory and gastrointestinal illnesses. To capture temporal patterns, we included the monthly rate of change in acute care visits and medication fills during 2017. We identified emergency department visits using Current Procedural Terminology, revenue, and place-of-service codes, while hospitalizations were defined as contiguous ED visits and inpatient admissions<sup>41</sup>.

We incorporated individual- and county-level social determinants of health measures based on established conceptual models linking social factors to healthcare utilization<sup>42</sup>. Individual-level measures included household size, income, English proficiency, marital status, citizenship status, and receipt of public assistance programs. County-level factors encompassed healthcare infrastructure (availability of substance use treatment facilities, mental health services, advanced practice providers, and urgent care), as well as area-level socioeconomic indicators and environmental factors (e.g., air quality, heat index; full definitions in Supplementary Note 6 and Table 2).

Because claims data capture only individuals with observed healthcare utilization, the model is limited to beneficiaries who have had at least some engagement with the healthcare system. However, the input data include individuals with minimal prior contact, and features such as missingness in clinical histories and enrollment gaps were treated as predictive signals. In line with NAM recommendations, missing data were retained as a feature rather than removed or simply imputed, enabling the model to incorporate patterns of under-documentation and exclusion.

Some individual-level social need data—such as income, education, or food insecurity—were unavailable for all members and were supplemented where possible using county-level proxies. Variables with missingness exceeding 20% were either excluded or imputed using multivariate imputation, depending on predictive importance and coverage. A detailed list of variable sources, missingness, and imputation methods is provided in Supplementary Table 2. Following National Academy of Medicine guidelines<sup>6,16</sup>, missingness itself was often retained as a feature to capture patterns of under-documentation and structural exclusion that may hold predictive value<sup>43</sup>.

## Model Development and Comparison

To evaluate the added predictive value of incorporating social determinants of health for forecasting quality measure non-completion, we developed two sets of prediction models for each of the nine outcome measures: (1) a baseline clinical model incorporating patient demographics, diagnoses, therapeutics, and healthcare utilization; and (2) an expanded social determinants model incorporating all variables from the baseline clinical model plus individual-level social factors (e.g., household income, reliance on Supplemental Security Income, Social Security Disability Insurance, Temporary Assistance for Needy Families, and English proficiency) and area-level social factors derived from patient residential FIPS county code (e.g., poverty rate, population density, and per capita rates of substance use treatment, mental health services, and urgent care facilities).

We employed an Extreme Gradient Boosting (XGBoost) algorithm for both model sets<sup>44–47</sup>. We selected XGBoost for its capacity to model non-linear relationships and interactions between diverse clinical and social features. In prior work using the same T-MSIS Medicaid dataset<sup>48</sup>, XGBoost outperformed Random Forest, logistic regression, and regularized regression in predicting acute care utilization. Given its superior empirical



performance in this context, we selected XGBoost while recognizing the trade-offs in interpretability.

To evaluate model performance and minimize overfitting, we implemented a standard 60/20/20 split for training, validation, and test sets, respectively. The validation set was used to tune hyperparameters during training, and the test set was preserved exclusively for final performance evaluation. Hyperparameters were optimized using a targeted tuning method described by Van Rijn and Hutter to enhance feature selection within the XGBoost framework (details in Supplementary Note 7)<sup>49</sup>. Although we did not implement nested cross-validation due to computational constraints within the CMS secure environment, we applied early stopping and regularization to mitigate overfitting. We acknowledge that relying on a single train/validation/test split may result in optimistic performance estimates. We benchmarked both models against a null model of random prediction using Monte Carlo simulation ( $n = 1000$  iterations).

### Performance Measures

Following standard TRIPOD guidelines, we evaluated model performance using metrics relevant to identifying patients at high risk of non-closure of a quality gap. Primary performance metrics included AUROC, F1-score, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the Matthews Correlation Coefficient (MCC), which ranges from -1 to +1 (where -1 indicates total disagreement between prediction and observation and +1 represents perfect prediction)<sup>50</sup>. We estimated 95% confidence intervals for each metric using bootstrapping with 1000 replications. Accuracy was reported for completeness but was not used as the primary metric due to class imbalance.

We compared the predictive power of the social determinants model to random selection targeting for closing care gaps. Using model-specific sensitivity and specificity values, we estimated open gap rates, effective closure rates, and the number of outreach attempts required to close one gap, assuming a typical 20% success rate per outreach attempt in engaging patients to close their care gaps<sup>51,52</sup>. This analysis provides a population-level estimate of the social determinants model's potential impact on improving quality measure completion rates compared to random targeting. To reflect the uncertainty in outreach success, we conducted a sensitivity analysis assuming lower success rates (5%, 10%, and 15%), reported in Supplement Table 6.

### Variable Importance

To understand the relative contribution of individual-level and area-level social determinants features in predicting quality measure non-completion, we assessed feature importance using the Gini index. Calculated within the XGBoost framework, the Gini index quantifies the average gain in purity (reduction in variance) achieved by splitting data based on a given feature across all decision trees in the ensemble. Features with higher Gini index values are considered more influential in the model's predictions. For each of the nine outcome measures, we ranked all features (clinical and social determinants variables) by their Gini importance scores. To facilitate comparison across outcome measures and between feature types, we normalized the Gini importance scores to a 0–100 scale by dividing each score by the maximum Gini importance observed across all features for that specific outcome measure<sup>53</sup>. We then examined the top ten features for each outcome measure to identify the most salient clinical and social factors associated with quality measure non-completion.

### Assessing the Potential Impact of Social Determinants Improvement

To explore how model predictions vary under hypothetical improvements in social determinants of health, we conducted model-based simulations (Supplementary Note 8). These simulations do not estimate causal effects but provide illustrative counterfactual scenarios with changes in input features. We compared predicted probabilities of quality measure completion before and after hypothetically improving each social determinant variable,

simulating a scenario with reduced social barriers. For the nine county-level variables, we first predicted quality measure completion using the held-out test set. We then created a modified version of this test set, where each member's county-level social measures were set to their 75th percentile value within our sample of 1563 counties. Values already at or above the 75th percentile remained unchanged. We selected this percentile a priori to represent substantial, but achievable, improvements in county-level social conditions.

For the five individual-level social variables (household income, reliance on Supplemental Security Income, Social Security Disability Insurance, Temporary Assistance for Needy Families, and English proficiency), we simulated improvement by shifting members from the lowest category to the next highest category. Using a dataset that incorporated all transformations (both county-level improvements to the 75th percentile and individual-level category shifts), we then re-generated model predictions to examine how estimated outcome probabilities shifted under hypothetical improvements.

We conducted two sensitivity analyses: (1) univariate analyses, adjusting each social determinant variable individually to isolate its effect, and (2) a multivariate analysis concurrently adjusting all social variables to estimate their combined impact. For each outcome measure and analysis, we calculated the relative and absolute percentage point change in predicted quality gap closure attributable to the hypothetical social determinant improvements.

### Evaluation of Potential Bias and Model Robustness

To assess potential biases and the robustness of our findings, we conducted several analyses (Supplementary Note 9). First, we evaluated racial/ethnic biases in the predictive models using the equalized odds method<sup>54</sup>. This approach examines whether the models exhibit differential predictive performance across racial/ethnic subgroups. Specifically, equalized odds assesses whether the probability of a prediction (i.e., receiving preventive care) is the same or different between groups, among those with the true outcome (i.e., they actually received the care). This method is particularly valuable for healthcare applications, as it ensures balanced true positive rates across racial/ethnic groups, preventing systematic under-identification of quality measure needs in historically marginalized populations.

We also conducted a sensitivity analysis using six-month intervals for defining quality measure completion (Supplementary Note 10). This addressed the potential for unobserved time-varying confounding due to care management programs, which may intensify outreach later in the year based on eligibility file updates. To examine potential selection bias introduced by the continuous enrollment criteria (36 months), we compared the baseline demographics of the included sample to those of the broader Medicaid population in our dataset enrolled for at least one month in 2017. This comparison evaluated the generalizability of our findings to a less stringently defined population and assessed the likelihood of biased predictions for those outside our 36-month sample.

To evaluate the robustness of model performance across subpopulations with more versus less data availability, we conducted a sensitivity analysis stratifying patients into low, medium, and high utilization tiers based on the volume of claims observed in the baseline period (Supplementary Note 10). For each model, we computed standard performance metrics (e.g., AUC, F1, sensitivity, specificity) separately within each utilization stratum. This allowed us to assess whether performance was disproportionately driven by high-utilization patients and to identify potential limitations in generalizability to patients with sparse data.

### Data availability

The datasets analyzed during this study are not publicly available due to restrictions on sharing patient-level Medicaid data containing protected health information. Researchers interested in accessing these data can apply through the Centers for Medicare and Medicaid Services (CMS). The application process includes Institutional Review Board (IRB) approval and use of a secure CMS data portal.

## Code availability

The code used to construct and evaluate the models is implemented in PySpark (version 3.2.1) and is publicly available at: <https://github.com/sadiqypatel/quality-measure-completion-prediction>.

Received: 3 March 2025; Accepted: 13 June 2025;  
Published online: 02 July 2025

## References

- Zhou, F. et al. Health and economic benefits of routine childhood immunizations in the Era of the Vaccines for Children Program - United States, 1994-2023. *MMWR Morb. Mortal. Wkly. Rep.* **73**, 682–685 (2024).
- Philipson, T. J., Durie, T., Cong, Z. & Fendrick, A. M. The aggregate value of cancer screenings in the United States: full potential value and value considering adherence. *BMC Health Serv. Res.* **23**, 829 (2023).
- Kaiser Family Foundation. What is Medicaid? (2021). <https://www.kff.org/health-policy/101-medicaid/>. Accessed December 1, 2024.
- Farley, T. A., Dalal, M. A., Mostashari, F. & Frieden, T. R. Deaths preventable in the U.S. by improvements in use of clinical preventive services. *Am. J. Prev. Med.* **38**, 600–609 (2010).
- Aggarwal, N. et al. *Advancing artificial intelligence in health settings outside the hospital and clinic*. NAM Perspectives. Discussion Paper, 2020. Available from: <https://nam.edu/advancing-artificial-intelligence-in-health-settings-outside-the-hospital-and-clinic/>. Accessed December 1, 2024.
- National Academy of Medicine; The Learning Health System Series, Whicker D, Ahmed M, Israni ST, Matheny M, eds. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington (DC): National Academies Press (US); August 2, 2023.
- UCSF Health. *Cancer Screening with Automated Outreach*. CipherHealth. Available from: <https://cipherhealth.com/resource/case-study/improve-the-reach-and-cost-effectiveness-of-panel-management-calls-for-cancer-screening-with-automated-outreach-ucsf-health/>. Accessed December 1, 2024.
- Hall, A. K., Cole-Lewis, H. & Bernhardt, J. M. Using text message reminders in health care services: A narrative literature review. *mHealth* **1**, 21 (2015). <https://www.sciencedirect.com/science/article/pii/S2214782918300022> Available from.
- CipherHealth. *Using Automated Preventive Outreach to Close Care Gaps and Reengage Patients in Their Care*. CipherHealth Blog. Available from: <https://cipherhealth.com/blog/using-automated-preventive-outreach-to-close-care-gaps-and-reengage-patients-in-their-care/>. Accessed December 1, 2024.
- MagView. *Appointment Reminders & Texts Messaging in Healthcare*. Available from: <https://magview.com/blog/text-messaging-in-healthcare-and-why-it-makes-perfect-sense-for-breast-centers/>. Accessed December 1, 2024.
- CipherHealth. *Using Automated Preventive Outreach to Close Care Gaps and Reengage Patients in Their Care* [PDF]. Available from: [https://cipherhealth.com/wp-content/uploads/2021/11/CipherHealth\\_NY\\_Health\\_System\\_Close\\_Care\\_Gaps.pdf](https://cipherhealth.com/wp-content/uploads/2021/11/CipherHealth_NY_Health_System_Close_Care_Gaps.pdf). Accessed December 1, 2024.
- Dialog Health. *Automate Healthcare with Text Appointment Reminders*. Available from: <https://www.dialoghealth.com/post/automate-healthcare-with-text-appointment-reminders>. Accessed December 1, 2024.
- CipherHealth. *CipherHealth's Outreach Solutions Reduce Hospital Readmissions by 56 Percent*. CipherHealth Case Study. Available from: <https://cipherhealth.com/resource/case-study/cipherhealths-outreach-solutions-reduce-hospital-readmissions-by-56-percent/>. Accessed December 1, 2024.
- Guy, R. et al. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. *PLoS One* **7**, e44423 (2012). <https://pmc.ncbi.nlm.nih.gov/articles/PMC3419880/> Available from.
- CipherHealth. *How Automated Preventive Outreach Can Help Close Care Gaps and Reengage Patients in Their Care*. CipherHealth Blog. Available from: <https://cipherhealth.com/blog/how-automated-preventive-outreach-can-help-close-care-gaps-and-reengage-patients-in-their-care/>. Accessed December 1, 2024.
- National Academy for State Health Policy. Data strategies to understand and address health disparities (2024). <https://nashp.org/data-strategies-to-understand-and-address-health-disparities/>. Accessed December 1, 2024.
- U.S. Preventive Services Task Force. Integrating evidence-based clinical and community strategies to improve health (2024). <https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/>. Accessed December 1, 2024.
- Artiga, S. & Hinton, E. Beyond health care: the role of social determinants in promoting health and health equity. Kaiser Family Foundation (2018). <https://www.kff.org/racial-equity-and-health-policy/issue-brief/beyond-health-care/>. Accessed October 5, 2024.
- Baum, A., Batniji, R., Ratcliffe, H., DeGosztonyi, M. & Basu, S. Supporting rising-risk Medicaid patients through early intervention. *NEJM Catal. Innov. Care Deliv.* **5**, 11 (2024).
- Hogg-Graham, R. et al. Technology, data, people, and partnerships in addressing unmet social needs within Medicaid managed care. *BMC Health Serv. Res.* **24**, 368 (2024).
- California Department of Health Care Services. California Advancing and Innovating Medi-Cal (CalAIM) Consolidated Quality Strategy (2022). <https://www.dhcs.ca.gov/services/Documents/Formatted-Combined-CQS-2-4-22.pdf>. Accessed October 4, 2024.
- Beidler, L. B. et al. Addressing social needs in clinical settings: early lessons from accountable health communities. *Popul. Health Manag.* **26**, 283–293 (2023).
- Evans, W. N. et al. Reducing readmissions by addressing the social determinants of health. *Am. J. Health Econ.* **7**, 1–40 (2021).
- Mohottige, D. et al. Residential structural racism and prevalence of chronic health conditions. *JAMA Netw. Open* **6**, e2348914 (2023).
- Njoku, A., Evans, M., Nimo-Sefah, L. & Bailey, J. Listen to the whispers before they become screams: addressing Black maternal morbidity and mortality in the United States. *Healthcare* **11**, 438 (2023).
- Bestsennyy O., Chmielewski M., Koffel A., Shah A. *The Untapped Potential of Payer Care Management*. McKinsey & Company; 2021. Available from: <https://www.mckinsey.com/industries/healthcare/our-insights/the-untapped-potential-of-payer-care-management>. Accessed December 1, 2024.
- Silvestri, D. et al. Factors associated with disparities in hospital readmission rates among US adults dually eligible for Medicare and Medicaid. *JAMA Health Forum* **3**, e214611 (2022).
- Radley, D. C. et al. Advancing racial equity in U.S. health care: The Commonwealth Fund 2024 state health disparities report. The Commonwealth Fund (2024).
- National Committee for Quality Assurance. HEDIS Measures (2024). <https://www.ncqa.org/hedis/measures/>. Accessed September 12, 2024.
- Centers for Medicare & Medicaid Services. Transformed Medicaid Statistical Information System Analytic Files (2024). <https://www.medicaid.gov/medicaid/data-systems/macbis/taf/>. Accessed September 12, 2024.
- Centers for Medicare & Medicaid Services. DQ Atlas (2024). <https://www.medicaid.gov/dq-atlas/>. Accessed September 12, 2024.
- Agency for Healthcare Research and Quality. Social Determinants of Health Database (2024). <https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html>. Accessed September 12, 2024.
- Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD). *Ann. Intern. Med.* **162**, 55–63 (2015).
- Healthcare Management Associates. *HMA Databook*. Bellevue, WA: Healthcare Management Associates; 2025.

35. Agency for Healthcare Research and Quality. *Clinical Classifications Software Refined (CCSR)*. Available from: [https://hcup-us.ahrq.gov/toolssoftware/ccsr/ccs\\_refined.jsp](https://hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp). Accessed December 1, 2024.
36. Centers for Medicare & Medicaid Services. *Restructured BETOS Classification System*. Available from: <https://data.cms.gov/provider-summary-by-type-of-service/provider-service-classifications/restructured-betos-classification-system>. Accessed December 1, 2024.
37. Centers for Medicare & Medicaid Services. *CMS Specialty Codes/Healthcare Provider Taxonomy Crosswalk*. 2003. Available from: <https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/MedicareProviderSupEnroll/Downloads/provider-taxonomy-crosswalk.pdf>. Accessed December 1, 2024.
38. Centers for Medicare & Medicaid Services. *Prescription Drug Data Collection*. 2024. Available from: <https://www.cms.gov/marketplace/about/oversight/other-insurance-protections/prescription-drug-data-collection-rxdc>. Accessed December 1, 2024.
39. Johnston, K. J., Allen, L., Melanson, T. A. & Pitts, S. R. A “Patch” to the NYU Emergency Department Visit Algorithm. *Health Serv. Res.* **52**, 1264–1276 (2017).
40. Agency for Healthcare Research and Quality. Guide to Prevention Quality Indicators: Hospital Admission for Ambulatory Care Sensitive Conditions. AHRQ publication no. 02-R0203 (2001).
41. Patel, S. Y. et al. Association between telepsychiatry capability and treatment of patients with mental illness in the emergency department. *Psychiatr. Serv.* **73**, 403–410 (2022).
42. Marmot M., Wilkinson R., eds. *Social Determinants of Health*. Oxford University Press; 2005.
43. Groenwold, R. H. H. Informative missingness in electronic health record systems: the curse of knowing. *Diagn. Progn. Res.* **4**, 8 (2020).
44. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (2016).
45. Herrin, J. et al. Comparative effectiveness of machine learning approaches for predicting gastrointestinal bleeds in patients receiving antithrombotic treatment. *JAMA Netw. Open* **4**, e2110703 (2021).
46. Irvin, J. A. et al. Incorporating machine learning and social determinants of health indicators into prospective risk adjustment for health plan payments. *BMC Public Health* **20**, 608 (2020).
47. Inoue, T. et al. XGBoost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury. *Neurotrauma Rep.* **1**, 8–16 (2020).
48. Patel, S. Y., Baum, A. & Basu, S. Prediction of non-emergent acute care utilization and cost among patients receiving Medicaid. *Sci. Rep.* **14**, 824 (2024).
49. Van Rijn, J. N. & Hutter, F. Hyperparameter importance across datasets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2367–2376 (2018).
50. Chicco, D. & Jurman, G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Min.* **16**, 4 (2023).
51. Coronado, G. D. et al. Randomized controlled trial of advance notification phone calls vs text messages prior to mailed fecal test outreach. *Clin. Gastroenterol. Hepatol.* **19**, 2353–2360 (2021).
52. Coronado, G. D. et al. The reach of calls and text messages for mailed FIT outreach in the PROMPT stepped-wedge colorectal cancer screening trial. *Cancer Epidemiol. Biomark. Prev.* **33**, 525–533 (2024).
53. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4765–4774 (2017).
54. Huang, J., Galal, G., Etemadi, M. & Vaidyanathan, M. Evaluation and mitigation of racial bias in clinical machine learning models: scoping review. *JMIR Med. Inform.* **10**, e36388 (2022).

## Acknowledgements

No external funding or grants supported this work.

## Author contributions

Drs. Patel and Basu led the study design, with all three authors (Drs. Patel, Basu, and Barnett) collaborating on analysis, interpretation, and manuscript development.

## Competing interests

Dr. Patel receives salary support and stock options from Waymark. Dr. Basu receives grants from the National Institutes of Health and the Centers for Disease Control and Prevention, personal fees from the University of California San Francisco, salary support from HealthRight360 and Waymark, and holds stock options in Collective Health and Waymark. Dr. Barnett serves as a paid scientific consultant for the State of California on risk stratification algorithms in Medicaid. The authors declare no pending patents or patent applications related to this work.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41746-025-01797-7>.

**Correspondence** and requests for materials should be addressed to Sadiq Y. Patel.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025