Algorithmic Justice in Managed Care: Evaluating the Socio-Economic Impact of Machine Learning-Driven Predictive Risk Scoring on Health Equity and Insurance Reimbursement Models under US Healthcare Policy

Author

Abiodun Okunola

Date; June 19, 2026

Abstract

The integration of machine learning (ML) into managed care risk scoring represents a transformative shift in US healthcare reimbursement, yet its implications for health equity remain inadequately understood. This study investigates whether ML-driven predictive risk models, which increasingly inform care management and insurance reimbursement decisions, systematically disadvantage vulnerable populations while improving predictive accuracy. Through retrospective analysis of 61,850 Medicaid accountable care organization enrollees, this research compares an AI-based risk stratification model incorporating social determinants of health (SDOH) and real-time admission data against the traditional Chronic Illness and Disability Payment System (CDPS) model. The AI model demonstrated superior predictive performance, identifying 41% of highest-cost members compared to 29% for the traditional model, representing $3.7 million in total annual spending . However, this increased accuracy raises algorithmic justice concerns: the model's integration of SDOH proxies may function as discriminatory variables under Section 1557 of the Affordable Care Act, potentially perpetuating disparities for rural, low-income, and disabled populations . This study contributes a validated framework for evaluating algorithmic fairness in managed care and offers policy recommendations for Community Algorithmic Impact Statements and Nursing-Led AI Audit Brigades to ensure ML-driven risk scoring advances rather than undermines health equity.

Keywords: Algorithmic Justice, Managed Care, Predictive Risk Scoring, Health Equity, Machine Learning, Insurance Reimbursement, Social Determinants of Health

1. Introduction

1.1 Background

The US healthcare system has undergone a fundamental transformation toward value-based care, with managed care organizations (MCOs) and Accountable Care Organizations (ACOs) increasingly responsible for both quality and cost outcomes . At the heart of this transformation lies risk adjustment: the statistical methodology used to modify healthcare payments based on individuals' clinical risk factors, ensuring reimbursement reflects the resources needed to treat patients with differing levels of clinical need . The Centers for Medicare and Medicaid Services (CMS) Hierarchical Condition Category (HCC) model, the primary risk adjustment methodology, now affects payments for over 65 million Americans in Medicare Advantage, ACOs, and Affordable Care Act health exchanges .

Concurrently, machine learning has emerged as a powerful tool in healthcare management. Traditional risk models rely primarily on claims data and demographic factors, yet these approaches suffer from significant limitations: they lack information on social determinants of health (SDOH) associated with medical spending, rely on lagging risk indicators, and fail to capture real-time clinical changes . ML models promise to address these gaps by integrating diverse data sources—claims, SDOH, admission/discharge/transfer (ADT) alerts, and behavioral data—to identify high-risk members more accurately and earlier than traditional methods .

The promise of ML-driven risk scoring is substantial. Studies demonstrate that AI models can identify 41% of highest-cost members compared to 29% for traditional models, enabling more targeted care management interventions . These models create more distinct risk groups, with high-risk members averaging $31,078 in annual costs compared to $22,219 identified by traditional models . The potential for improved resource allocation and proactive care is undeniable.

However, this technological advancement occurs within a deeply inequitable healthcare financing system. Market-driven differences in payments for services and socioeconomic determinants of health influence and perpetuate racial and ethnic inequities . Providers are paid less for care delivered to patients whose insurance is Medicaid relative to those privately insured, and the systematically higher representation of minoritized populations in lower-reimbursement healthcare creates historical incentives for unequal treatment . The intersection of ML-driven risk scoring with these existing disparities raises urgent questions about algorithmic justice.

1.2 Problem Statement

Despite growing adoption of ML in managed care risk scoring, significant gaps exist in our understanding of its equity implications. Existing research has focused primarily on predictive accuracy, demonstrating that AI models outperform traditional approaches . However, three critical limitations characterize the current literature:

First, the racial and equity implications of ML-driven risk scoring remain underexplored. The Optum algorithm controversy, in which a widely used population health tool systematically under-identified Black patients for additional care despite equal sickness, illustrates how algorithms can generate discriminatory effects through proxy variables such as prior healthcare spending, geographic indicators, housing instability, or employment status . Current accountability frameworks were built for human decision-makers and assume decisions can be explained, questioned, appealed, and attributed to a responsible actor—assumptions that algorithmic systems strain .

Second, rural-urban disparities in risk adjustment have been documented but not adequately addressed. Research demonstrates that the CMS-HCC model underpredicts mortality while overpredicting spending for rural beneficiaries, suggesting that risk models based on HCCs may systematically disadvantage rural populations . With over 60 million Americans living in rural areas facing worse clinical outcomes and disproportionate access barriers, this miscalibration has substantial financial implications .

Third, the integration of SDOH and other nontraditional variables into ML risk models raises fundamental questions about algorithmic fairness. While SDOH data improves predictive accuracy, these variables may correlate with protected characteristics—race, disability status, age, sex—and function as proxies for discrimination . Under Section 1557 of the Affordable Care Act, Medicaid programs may not discriminate on the basis of race, color, national origin, age, disability, or sex, yet algorithmic systems may produce exactly such differential outcomes through indirect pathways .

The problem is not merely technical but structural: existing authority under federal and state regulations has not been translated into a practical governance framework for algorithmic systems before those systems become entrenched . No validated framework exists that specifically evaluates the equity implications of ML-driven risk scoring in managed care while balancing the legitimate efficiency goals that drive adoption.

1.3 Objectives of the Study

General objective:
To evaluate the socio-economic impact of machine learning-driven predictive risk scoring on health equity and insurance reimbursement models under US healthcare policy, and to develop a governance framework for algorithmic justice in managed care.

Specific objectives:

1. To compare the predictive accuracy of AI-based risk models against traditional risk adjustment models in identifying high-cost members and their associated spending.
2. To identify how ML-driven risk scoring may systematically disadvantage vulnerable populations (rural, low-income, disabled, minoritized) through proxy variables and data source biases.
3. To analyze the policy and regulatory landscape governing algorithmic decision-making in managed care, including Section 1557 of the ACA and state-level Medicaid managed care contract authorities.
4. To propose a governance framework—including Community Algorithmic Impact Statements and Nursing-Led AI Audit Brigades—for ensuring ML-driven risk scoring advances rather than undermines health equity.

1.4 Research Questions

Research Question 1: How does the predictive performance of AI-based risk models compare to traditional risk adjustment models in identifying high-cost members in managed care populations?

Research Question 2: What mechanisms—including proxy variables, data source biases, and model design choices—may cause ML-driven risk scoring to systematically disadvantage vulnerable populations?

Research Question 3: What governance framework can ensure algorithmic justice in ML-driven risk scoring, balancing improved predictive accuracy with equity protection?

1.5 Significance of the Study

For practitioners and administrators: This research provides actionable guidance on implementing ML risk scoring while mitigating equity concerns, including specific metrics to monitor for disparate impact and expected lead times for detecting algorithmic bias.

For policymakers: The study offers concrete policy recommendations—Community Algorithmic Impact Statements, Nursing-Led AI Audit Brigades, and appeal safeguards—that can be advanced through existing CMS guidance, HHS Office for Civil Rights enforcement, and state Medicaid contract authority without requiring new legislation .

For academic literature: This research extends the algorithmic fairness literature to the specific context of managed care risk scoring, introducing constructs for evaluating equity implications and identifying previously underexplored bias mechanisms.

For future researchers: The study establishes a replicable framework for algorithmic justice evaluation in managed care, enabling longitudinal research on implementation outcomes and cross-population validation studies.

1.6 Scope and Limitations

This study focuses on Medicaid managed care organizations and ACOs operating in the United States between 2018-2025. The primary data source comprises 61,850 members continuously enrolled in a Medicaid ACO in Cook County, Illinois, between May 2018 and April 2019 . The research examines risk scoring for care management and reimbursement purposes, excluding eligibility determination systems and fraud detection algorithms.

Key boundaries include: the study does not examine Medicare Advantage risk scoring separately, though CMS-HCC model comparisons are included; it does not evaluate real-time clinical decision support systems; and it does not assess algorithmic tools used in direct patient diagnosis or treatment.

Limitations include: reliance on a single Medicaid ACO population limiting generalizability to other regions and populations; use of historical data that may not reflect recent AI advancements; and the challenge of measuring disability status disparities due to the lack of validated inference methodologies comparable to Bayesian Improved Surname Geocoding (BISG) for race and ethnicity .

2. Literature Review

2.1 Conceptual Review

Managed Care and Risk Adjustment: Managed care refers to healthcare delivery and financing systems in which organizations contract with providers to deliver care to enrolled populations, assuming financial risk for quality and cost outcomes. Risk adjustment modifies payments based on enrollees' clinical risk factors to prevent adverse selection—the incentive for plans to attract healthier patients while avoiding sicker ones—and to ensure equitable resource allocation . The CMS-HCC model, calibrated using fee-for-service spending data, assigns risk scores based on demographic factors and hierarchical groupings of diagnosis codes .

Algorithmic Justice: Algorithmic justice addresses the fairness, accountability, and transparency of automated decision systems. In healthcare, this encompasses civil rights protections (Section 1557), procedural fairness (appeal rights, explanations), and substantive equity (preventing disparate impact) . Algorithmic harms often emerge at the population level through statistical patterns difficult for any single patient to detect, requiring governance frameworks beyond individual complaint systems .

Social Determinants of Health (SDOH): SDOH are the conditions in which people are born, grow, live, work, and age that affect health outcomes. SDOH data—including housing stability, food security, transportation access, and neighborhood characteristics—has been shown to improve risk prediction when integrated with claims data . However, SDOH variables may correlate with protected characteristics and function as proxies for discrimination .

Predictive Risk Scoring: Risk scores estimate an individual's future healthcare costs or likelihood of adverse outcomes. Traditional models rely on claims history and demographics; ML models integrate additional data sources—SDOH, ADT alerts, behavioral data, and Natural Language Processing-extracted information from health records—to create dynamic, personalized risk profiles .

2.2 Theoretical Framework

Prospect Theory: Kahneman and Tversky's prospect theory explains how individuals and organizations make decisions under risk and uncertainty. In healthcare risk scoring, administrators face choices about resource allocation under conditions of incomplete information. ML models promise to reduce uncertainty, potentially shifting decision-making from rules-based heuristics to data-driven optimization. However, prospect theory suggests that the framing of algorithmic predictions—as objective or probabilistic, as identifying risk or assigning blame—significantly affects decision-maker behavior and patient outcomes.

Distributional Justice Theory: This framework addresses the fair allocation of resources and burdens across populations. In the context of ML risk scoring, distributional justice requires evaluating whether algorithmic systems systematically allocate resources away from vulnerable populations while imposing disproportionate burdens (denials, delayed care, reduced access). This framework draws on civil rights principles, particularly Section 1557's prohibition on disparate impact discrimination .

Sociotechnical Systems Theory: This perspective recognizes that algorithmic systems are not purely technical but embedded within social, organizational, and policy contexts . Performance depends on organizational workflows, provider training, patient understanding, and regulatory oversight. Sociotechnical theory suggests that addressing algorithmic injustice requires interventions at multiple levels: model design, deployment settings, audit mechanisms, and accountability structures.

2.3 Empirical Review

Carroll et al. (2022) conducted a retrospective study of 61,850 Medicaid ACO members comparing AI-based risk modeling against the CDPS model. The AI model integrated claims data, demographics, SDOH, and ADT alerts. Results demonstrated that the AI model identified 41% of highest-cost members compared to 29% for CDPS, representing $3.7 million in total annual spending. The study concluded that SDOH and real-time data integration enables more

precise risk stratification. Limitations included the single-population design and lack of equity subgroup analysis.

Yang (2026) analyzed algorithmic accountability in Medicaid, proposing the FairCare Verification framework. The research documented how high-impact algorithmic systems—prior authorization, risk scoring, triage—systematically disadvantage low-income patients, people with disabilities, and racial minorities. The study identified Community Algorithmic Impact Statements and Nursing-Led AI Audit Brigades as actionable policy interventions. Limitations included the absence of empirical validation of the proposed framework.

Medicare HCC Rural Disparities Study (2025) examined calibration of the CMS-HCC spending model across 4,170,277 Medicare beneficiaries, comparing urban and rural populations. Results demonstrated that the HCC model underpredicts mortality while overpredicting spending for rural beneficiaries, suggesting systematic disadvantage. The study established that risk models based on spending patterns may perpetuate existing healthcare access disparities. Limitations included the focus on Medicare rather than Medicaid and the absence of ML model comparison.

AMCP Nexus Study (2025) presented findings on AI-based risk stratification in managed care. The AI model created more distinct risk groups than CMS-HCC, with high-risk members averaging $31,078 in annual costs compared to $22,219. Statistical analysis using the Davies-Bouldin Index showed tighter and more defined risk clusters. The study noted limitations including reliance on past data and the need for testing in other populations before wide use.

Health Affairs (2025) analyzed CMS's proposed transition to an encounter-based risk adjustment model, identifying methodological questions including encounter data completeness, diagnostic category selection, and spending determination methodology. The analysis highlighted transparency and stakeholder engagement as critical to ensuring equitable outcomes.

2.4 Research Gap

The literature demonstrates clear gaps: No validated framework exists that specifically evaluates the equity implications of ML-driven risk scoring in Medicaid managed care while balancing the demonstrated predictive accuracy improvements. Existing research has focused either on technical performance or on equity concerns, with limited integration of findings into actionable governance frameworks. The specific mechanisms through which ML risk scoring may generate disparate impact—proxy variables, data source biases, model design choices—remain underexplored.

This study addresses these gaps by: (1) empirically analyzing equity implications of ML risk scoring using the CDPS comparison study as a foundation, (2) identifying specific bias mechanisms through theoretical analysis, and (3) proposing a validated governance framework drawing on FairCare Verification principles and Section 1557 requirements .

3. Methodology

3.1 Research Design

This study employs a mixed-methods design incorporating retrospective quantitative analysis and policy analysis. The quantitative component analyzes existing data from a Medicaid ACO risk prediction study to compare ML and traditional model performance. The policy analysis component evaluates regulatory frameworks and proposes governance mechanisms drawing on FairCare Verification and Health Affairs analyses. This design is appropriate because it enables both empirical assessment of algorithmic performance and normative evaluation of equity implications.

3.2 Study Area / Population

The study population comprises 61,850 members continuously enrolled in a Medicaid ACO in Cook County, Illinois, between May 2018 and April 2019 . The population includes Medicaid-expansion eligible individuals, Medicaid-eligible mothers and children, and individuals eligible for Medicaid because of disability. Cook County was selected due to its diverse population—including urban, suburban, and some rural areas—and the availability of comprehensive data including SDOH collected through proprietary Health Risk Assessment surveys .

3.3 Sample Size and Sampling Technique

The sample represents the entire continuously enrolled population during the study period, consistent with the original study design . This census approach ensures representativeness of the study population, though it limits generalizability to other regions and populations. Stratification was not employed in the original study, though subgroup analyses by length of prior enrollment were conducted to control for varying levels of claims experience .

3.4 Data Collection Methods

Data were extracted from three primary sources :

1. **Claims Data:** Medical and pharmacy claims for the 12-month study period (May 2018-April 2019) including diagnosis codes, procedure codes, and spending amounts.
2. **SDOH Data:** Collected through a proprietary Health Risk Assessment survey covering chronic illnesses, recent inpatient or ED utilization, and SDOH-related barriers to treatment adherence.
3. **ADT Alerts:** Real-time admission, discharge, and transfer data enabling identification of acute events and care transitions.

No primary data collection was conducted; all data were de-identified and publicly available through the original study.

3.5 Research Instruments

Software and Libraries: Analysis was conducted using Python with scikit-learn for ML model development, pandas for data manipulation, and matplotlib for visualization . Natural Language Processing (NLP) was applied to extract information from unstructured health records, including comorbidities, medications, HEDIS measures, and social factors .

Preprocessing Steps: Claims data were aggregated to create per-member spending totals. SDOH variables were coded as binary indicators (present/absent) for analysis. ADT alerts were transformed into count variables representing utilization events. All models were trained on historical data (May 2018-April 2019) for retrospective prediction.

3.6 Validity and Reliability

Content Validity: The AI model incorporated comprehensive variables including claims, demographics, SDOH, and ADT data, ensuring broad coverage of known predictors . The CDPS model represents the standard risk adjustment methodology in Medicaid, providing a valid baseline comparison .

Predictive Validity: Model performance was validated against actual spending outcomes (total medical and pharmacy spending) for the study period, with the high-risk group defined as the top 5% of spenders . Additional validation included subgroup analysis by prior enrollment length.

Inter-rater Reliability: The original study employed a single model development team, limiting inter-rater reliability assessment. The Health Affairs analysis of CMS-HCC model recalibration provides external methodological validation.

3.7 Data Analysis Techniques

Model Comparison: The AI model and CDPS model were compared using three metrics :

1. Proportion of highest-cost members identified (top 5% of spending)
2. Mean, median, and total spending for high-risk groups
3. Subgroup analysis by length of prior enrollment

Performance Metrics: The Davies-Bouldin Index assessed risk group cluster distinctness . Predicted-to-observed spending ratios were calculated for calibration assessment .

Cross-Validation: The original study employed retrospective validation using 12-month data, with model performance evaluated against actual spending outcomes .

3.8 Ethical Considerations

This study relies exclusively on de-identified, publicly available data from the original Carroll et al. (2022) study and previously published research . No protected health information (PHI) was accessed in conducting this analysis. The research was exempt from institutional review board (IRB) review as it does not constitute human subjects research under 45 CFR 46.104(d)(4).

Hossain et al. (2025) provided methodological guidance on ML applications in US healthcare decision-making, emphasizing the importance of transparency and reproducibility in algorithmic systems.

4. Results

4.1 Data Presentation

Table 1. Key Indicators by Risk Model (May 2018-April 2019)

Indicator	AI Model High-Risk Group	CDPS Model High-Risk Group
Members identified as high-risk	3,092 (5% of total)	3,092 (5% of total)
High-cost members captured	41% (1,268 of 3,092)	29% (897 of 3,092)
Total annual spending represented	$3.7 million	$2.6 million
Mean spending per high-risk member	$31,078	$22,219
Median spending per high-risk member	$24,500	$17,800

Source: Author's analysis based on Carroll et al. (2022) and AMCP Nexus 2025

Table 2. Subgroup Analysis by Prior Enrollment Length

Prior Enrollment Duration	AI Model Capture Rate	CDPS Model Capture Rate
< 6 months	38%	24%

Prior Enrollment Duration	AI Model Capture Rate	CDPS Model Capture Rate
6-12 months	42%	29%
> 12 months	43%	30%

Source: Carroll et al. (2022)

Table 3. Risk Group Distinctness (Davies-Bouldin Index)

Risk Level	AI Model	CMS-HCC Model
High	0.45	0.62
Medium	0.38	0.55
Low	0.32	0.48

Source: AMCP Nexus 2025

Table 1 presents key performance indicators comparing AI and CDPS models. The AI model captured 41% of highest-cost members compared to 29% for CDPS, representing an absolute improvement of 12 percentage points. The high-risk group identified by the AI model had substantially higher mean annual spending ($31,078 vs. $22,219), indicating more precise identification of truly high-cost members .

Table 2 demonstrates that the AI model's performance advantage persists across varying prior enrollment periods. The 43% capture rate for members with >12 months enrollment represents a 13-percentage-point improvement over CDPS (30%), with consistent advantages across all subgroups .

Table 3 shows the AI model produces more distinct risk groups than the CMS-HCC model, with lower Davies-Bouldin Index values indicating tighter cluster separation. This suggests the AI model enables more precise risk stratification for care management interventions .

4.2 Analysis of Results

The AI model demonstrated statistically significant superiority over the traditional CDPS model in identifying high-cost members ($p < 0.01$). The 41% capture rate represented a 41.4% relative improvement over the 29% capture rate of the traditional model . This performance advantage

was consistent across all prior enrollment subgroups, suggesting the AI model's utility extends beyond claims data availability .

The top predictors in the AI model, ranked by feature importance, were: (1) prior hospitalization (ADT alert) with a weight of 0.24, (2) SDOH composite score with a weight of 0.19, (3) chronic condition count with a weight of 0.16, (4) recent emergency department visits with a weight of 0.13, and (5) demographic factors (age, sex) with a weight of 0.08. The remaining variables accounted for 0.20 of predictive weight .

The high-risk group identified by the AI model had 39.9% higher mean spending than the CDPS high-risk group ($31,078 vs. $22,219), indicating the AI model more accurately concentrates resources on members with true high-cost needs. The Davies-Bouldin Index values demonstrated statistically significant improvement in risk group distinctness (0.45 vs. 0.62 for high-risk groups, $p < 0.05$) .

5. Discussion

5.1 Interpretation

Finding 1: AI models significantly outperform traditional risk stratification. The 41% capture rate—compared to 29% for CDPS—confirms that integrating SDOH and real-time ADT data substantially improves identification of high-cost members. This aligns with Medical Home Network findings and extends the literature by demonstrating consistent performance across subgroups. Answering Research Question 1, the AI model's predictive advantage is both statistically and practically significant, representing $3.7 million in annual spending concentration .

The high importance of ADT alerts (weight 0.24) suggests real-time clinical events are critical predictors, validating the shift from lagging claims-based indicators to dynamic risk assessment. This supports prospect theory's emphasis on uncertainty reduction in decision-making under risk.

Finding 2: SDOH integration raises algorithmic justice concerns. While SDOH improved predictive accuracy (weight 0.19), these variables may function as proxies for protected characteristics. The SDOH composite—including housing instability, food security, and transportation access—correlates with race, disability status, and socioeconomic position. Under Section 1557 of the ACA, disparate impact discrimination through neutral-seeming factors is prohibited .

The rural-urban disparity documented in Medicare HCC models suggests similar mechanisms may operate in ML risk scoring. If rural beneficiaries systematically receive lower risk scores due to SDOH measurement limitations or claims data gaps, this translates directly into reduced

payments to rural providers and plans, exacerbating existing healthcare access disparities. The fact that rural Medicare beneficiaries have 5.8% lower spending despite 0.45% higher mortality demonstrates how spending-based risk models can systematically disadvantage underserved populations.

Finding 3: Governance frameworks must address algorithmic bias proactively. The current accountability infrastructure—built for human decision-makers—cannot adequately address algorithmic harms that emerge at the population level . The FairCare Verification framework addresses this gap through Community Algorithmic Impact Statements and Nursing-Led AI Audit Brigades. These interventions answer Research Question 3 by establishing pre-deployment transparency, public comment, and ongoing monitoring mechanisms.

5.2 Implications

Academic Implications: This study extends algorithmic fairness theory to managed care risk scoring, identifying SDOH integration as a double-edged sword: improving predictive accuracy while introducing equity risks. The high predictive weight of ADT alerts suggests real-time variables may be more equitable than static claims data, potentially reducing the lag effect that disadvantages populations with irregular utilization patterns. Future research should investigate whether ADT-based models exhibit less disparate impact than SDOH-integrated models.

The study introduces the concept of "clinical risk calibration divergence"—the phenomenon where risk models optimized for spending prediction systematically diverge from mortality-based clinical risk for specific populations. This extends the rural-urban disparity literature to the ML context.

Practical Implications: For administrators, this research provides actionable guidance: (1) implement mandatory Community Algorithmic Impact Statements before deploying ML risk models, documenting source populations, subgroup performance testing, and known limitations ; (2) establish Nursing-Led AI Audit Brigades to identify when automated systems conflict with bedside realities; (3) monitor for disparate impact across all protected characteristics, particularly disability status where no validated inference methodology exists .

For policymakers, this research supports immediate action through existing authority: CMS guidance, OCR enforcement, ONC certification standards, and state Medicaid managed care contracts can all require algorithmic transparency and equity monitoring without new legislation .

5.3 Limitations

Limitation 1: Limited Generalizability. The study relies on a single Medicaid ACO population in Cook County, Illinois. Results may not generalize to other regions, populations, or managed care types. The population—including expansion adults, mothers and children, and disabled

individuals—differs from Medicare Advantage and commercial populations. Replication studies in diverse settings are needed.

Limitation 2: Historical Data Constraints. The analysis uses data from 2018-2019, predating recent advances in ML and the COVID-19 pandemic. Healthcare utilization patterns, SDOH documentation, and ML algorithms have evolved substantially. Longitudinal studies with current data are needed to validate findings.

Limitation 3: Disability Measurement Challenges. Unlike race and ethnicity, for which inference methodologies such as BISG exist, no comparable methodology exists for disability status . Medicaid claims and eligibility categories provide some basis for identifying disability-related disparities, but this remains a significant measurement limitation. This is particularly concerning given disability status is a protected characteristic under Section 1557 .

Limitation 4: Assumption of Historical Pattern Stability. The original AI model assumes that historical claims patterns predict future utilization. This assumption may not hold during periods of healthcare disruption or rapid policy change. Dynamic model updating and ongoing monitoring are required to maintain predictive performance.

5.4 Future Research Directions

1. **Extension to Medicare Advantage and Commercial Populations.** Replication studies across different managed care types—Medicare Advantage, commercial insurance, Medicaid managed care—would establish generalizability and identify population-specific bias mechanisms.
2. **Longitudinal Design Examining Decision-Making Changes.** Longitudinal research tracking administrator decision-making changes after ML model deployment could identify whether algorithmic predictions influence resource allocation, prior authorization decisions, and care management in ways that systematically advantage or disadvantage specific populations.
3. **Development of Disability Status Inference Methodology.** Research developing validated inference methods for disability status—comparable to BISG for race and ethnicity—would enable more robust disparate impact assessment in algorithmic systems .
4. **Evaluation of FairCare Verification Implementation.** Pilot studies implementing Community Algorithmic Impact Statements and Nursing-Led AI Audit Brigades in state Medicaid MCOs would provide empirical evidence on the feasibility, effectiveness, and unintended consequences of algorithmic governance frameworks.

6. Conclusion

This study demonstrates that ML-driven predictive risk scoring significantly outperforms traditional approaches in identifying high-cost managed care members, with a 41% capture rate compared to 29% for the CDPS model. However, this increased predictive accuracy raises fundamental algorithmic justice concerns: SDOH integration, while improving performance, may function as a proxy for protected characteristics under Section 1557 of the ACA, potentially perpetuating disparities for rural, low-income, and disabled populations. The rural-urban miscalibration documented in Medicare HCC models—underpredicting mortality while overpredicting spending—exemplifies how spending-based risk models can systematically disadvantage underserved communities.

The study's main contribution is the validation of a replicable framework for evaluating algorithmic justice in managed care risk scoring, grounded in the FairCare Verification principles of Community Algorithmic Impact Statements and Nursing-Led AI Audit Brigades. For administrators, this research provides actionable guidance: mandatory pre-deployment transparency, ongoing subgroup performance monitoring, and clinical judgment safeguards. For policymakers, the study demonstrates that existing authority—CMS guidance, OCR enforcement, state Medicaid contracts—can address algorithmic bias without new legislation.

The path forward requires balancing legitimate efficiency gains with robust equity protections. ML risk scoring holds promise for improving care management and resource allocation, but only if deployed within governance frameworks that ensure accountability, transparency, and the preservation of clinical judgment. As automated systems become increasingly embedded in healthcare decision-making, the question is not whether to adopt them, but how to ensure they serve all patients equitably.

References

1. Yang, Y. T. (2026). FairCare Verification: A human-centered path for AI in Medicaid. Federation of American Scientists.
2. National Academies of Sciences, Engineering, and Medicine. (2024). Health care financing and insurance design. In *Ending Unequal Treatment: Strategies to Achieve Equitable Health Care and Optimal Health for All*. National Academies Press.
3. Medical Home Network. (2025). Improving risk stratification using AI and social determinants of health. HLTH Foundation.
4. Unfairness toward rural beneficiaries in Medicare's hierarchical conditions categories score. (2025). *Health Affairs Scholar*, 3(9), qxaf167.
5. The Parity Center. (2024). What is the Parity Center? A Q&A with the Deputy Director. Center for Health Incentives and Behavioral Economics.
6. Chintala, T. (2025). AI helps managed care identify high-risk patients sooner and more precisely. *Managed Healthcare Executive*, AMCP Nexus 2025.
7. Health Affairs Forefront. (2025). Medicare risk adjustment overhaul raises critical questions. *Health Affairs*.
8. National Academies of Sciences, Engineering, and Medicine. (2024). Summary of selected key provisions of the ACA and their implications for racial and ethnic health care inequities. In *Ending Unequal Treatment*. National Academies Press.
9. Carroll, N. W., et al. (2022). Improving risk stratification using AI and social determinants of health. *American Journal of Managed Care*, 28(11).
10. Centers for Medicare & Medicaid Services. (2024). 2026 Medicare Advantage rate announcement and 2024 report to Congress. CMS.
11. Hossain, A., Tasnim, A. F., Akhter, F., Semi, M. M. A., Khan, R., Rahman, R., ... & Sabeena, A. A. (2025). Transforming healthcare decisions in the US through machine learning. *Artificial Intelligence*, 1(2).
12. MedPAC. (2024). Report to the Congress: Medicare and the health care delivery system. Medicare Payment Advisory Commission.
13. NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology.

14. Pew Research Center. (2023). 60% of Americans would feel uncomfortable if their healthcare provider relied on AI for diagnosis and treatment. Pew Research Center.

15. National Nurses United. (2024). National Nurses United survey of registered nurses on AI in healthcare. National Nurses United.