

ORIGINAL RESEARCH ARTICLE

Machine learning–driven insights into Medicare provider performance, patient profiles, and care impact

Sameh Shamroukh*

Department of Information Technology and Decision Sciences, G. Brint Ryan College of Business, University of North Texas, Denton, Texas, United States of America

Abstract

This study examines the performance of Medicare providers, with a focus on patient backgrounds, number of chronic illnesses, and efficacy of drug treatments as compared to traditional medical services. Over 1.2 million records from the Centers for Medicare & Medicaid Services and different methods, such as random forest regression and K-means clustering, were utilized to identify top-performing providers, categorize patients, and evaluate the results of treatments. The results reveal that while spending more on Medicare and offering more services can sometimes improve patient outcomes, these improvements are not always steady. This inconsistency points out some ongoing problems in the system, primarily affecting older adults and those in underserved communities, who often struggle with worse health and limited access to care. In addition, the study found that the effectiveness and cost of different treatment methods can vary widely. Drug treatments and direct medical services had varying impacts on resource use and health benefits. Combining large-scale public data with advanced analytic techniques, this research provides a reference for policymakers and healthcare organizations and offers insights into designing targeted interventions, with the ultimate aim to preserve fairness and sustainability in the U.S. healthcare system.

Keywords: Healthcare disparities; Chronic disease management; Healthcare cost-effectiveness; Equity

***Corresponding author:**Sameh Shamroukh
(sameh.shamroukh@unt.edu)

Citation: Shamroukh S. Machine learning–driven insights into Medicare provider performance, patient profiles, and care impact. *Artif Intell Health*. 2026;3(2):025390078. doi: 10.36922/AIH025390078

Received: September 23, 2025**1st revised:** October 28, 2025**2nd revised:** November 5, 2025**3rd revised:** November 14, 2025**Accepted:** November 20, 2025**Published online:** December 5, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Millions of Americans, particularly older people and individuals with chronic conditions, rely on the U.S. healthcare system, with Medicare playing a central role in service delivery. Despite its critical importance, the system faces ongoing challenges in maintaining provider efficiency, ensuring equitable patient care, and optimizing the allocation of limited resources.

Several systemic issues contribute to these challenges. First, variability in provider performance is a significant concern: While some healthcare providers consistently achieve high efficiency and patient satisfaction, others experience operational inefficiencies and lower performance scores. Second, demographic disparities in care affect patient outcomes and access to services. Age, socioeconomic status, and

geographic location contribute to uneven care quality across populations.

In addition, the high prevalence of chronic diseases among Medicare beneficiaries, such as diabetes, heart disease, and chronic obstructive pulmonary disease, requires long-term management and drives substantial healthcare costs. Finally, the relative cost-effectiveness of different treatment types, including drug-based therapies versus direct medical interventions, remains poorly understood, limiting evidence-based decision-making.

This research seeks to address these challenges by leveraging Business Intelligence and advanced data analytics to uncover patterns in provider performance, demographic disparities, chronic disease prevalence, and treatment outcomes. The findings aim to inform more efficient, equitable, and cost-effective healthcare delivery under Medicare.

1.1. Research goals and analytical tools

This study applies Business Intelligence tools and advanced data analytics to evaluate Medicare provider performance, examine patient demographics, analyze chronic disease trends, and compare treatment models. By leveraging these techniques, the research aims to uncover actionable insights that can support more efficient and equitable healthcare delivery.

The primary objectives of the study include a detailed analysis of provider performance. This involves identifying top-performing Medicare providers based on payment allowances and reimbursements and assessing the relationship between financial metrics, provider efficiency, and patient outcomes. Understanding these patterns can help highlight operational best practices and areas for improvement.

Another key objective is to investigate patient demographics and disparities. The study examines the distribution of patients by age, gender, location, and socioeconomic status to identify whether variations in care exist among different demographic groups and geographic regions. These insights are critical for addressing inequities in service access and health outcomes.

The research also focuses on the prevalence of chronic conditions among Medicare beneficiaries. It explores how the distribution of chronic diseases varies by provider and region and identifies practical strategies for managing these conditions. Finally, the study evaluates the impact of drug-based versus direct medical services. The research assesses the relative effectiveness of pharmaceutical and medical service interventions by comparing patient outcomes, costs, and resource utilization across treatment types.

Interactive dashboards and machine learning models were utilized for conducting analyses, including random forest regression and K-means clustering. These tools enable data visualization, pattern recognition, and the extraction of meaningful insights from complex Medicare datasets.

1.2. Research questions

Research questions of this work are as follows:

- (i) How do Medicare payments and allowed amounts correlate with provider performance and patient satisfaction?
- (ii) What is the variation in patient demographics across different geographies, and does it lead to disparities in care?
- (iii) How is the distribution of chronic diseases among Medicare beneficiaries related to the provider or geographic location?
- (iv) What are the anticipated patient outcomes based on the type of service offered (drug or medical), and what are the costs or resource allocation differences between the two services?

1.3. Literature review

The U.S. healthcare system faces structural inefficiencies, including resource misallocation, widening disparities, and persistent patient dissatisfaction. These issues are particularly pronounced within Medicare, where demographic shifts, chronic disease prevalence, and complex payment structures strain providers and beneficiaries. Over the past two decades, research has increasingly focused on data-driven, performance-based, patient-centered, and equity-sensitive approaches to reform, with notable emphasis on alternative payment models (APMs), value-based purchasing (VBP), and patient experience metrics.¹⁻⁴ (Table 1).

This literature can be synthesized across five thematic domains:

- (i) Patient Satisfaction and Care Quality
- (ii) Payment Models and Chronic Care Quality
- (iii) Provider Performance and Workforce Shifts
- (iv) Equity, Disparities, and Payment Adjustments
- (v) Systemic Payment Variations and Policy Reforms.

1.4. Key thematic insights

1.4.1. Patient satisfaction and care quality

Patient satisfaction is a critical performance indicator in healthcare delivery and system accountability. Large-scale studies consistently demonstrate that patient-centered care is strongly associated with improved clinical outcomes, better adherence to treatment, and enhanced trust in

Table 1. Systematic literature review summary

Theme	Key sources (references)	Methods and data	Key findings	Gaps and limitations	Relevance to the present study
Patient Satisfaction and Care Quality	1,5-10	Mixed methods, national surveys, patient experience databases	Patient-centered care improves outcomes; satisfaction varies by demographics	Aggregated data; limited provider-level analysis	Examines provider-level variation in satisfaction
Payment Models and Chronic Care Quality	2,11-15	Scoping/systematic reviews, Medicare datasets	APMs improve quality and cost efficiency; risk of widening disparities	Weak demographic linkage; limited subgroup analysis	Analyzes APM effects at the patient and provider level
Provider Performance and Workforce Shifts	16-21	Cross-sectional time series, workforce data	NP/PA workforce growing rapidly, expanding rural access	Limited link to satisfaction and payment outcomes	Integrates workforce indicators with performance and satisfaction
Equity, Disparities, and Payment Adjustments	22-27	Regression, simulation, Medicare administrative data	Payment equity adjustments redistribute resources to safety-net providers	Mostly simulation studies; lacking patient-level outcomes	Links payment and social determinants of health to satisfaction and performance
Systemic Payment Variations and Policy Reforms	28-41	Policy analysis, claims data, econometric models	Significant payment variation by diagnosis/region	High-level analysis; weak linkage to outcomes	Examine structural payment disparities and equity effects

Abbreviations: APMs: Alternative payment models; NP: Nurse practitioner; PA: Physician assistant.

healthcare providers.^{1,5,7} High levels of satisfaction are often linked to improved patient–provider communication, continuity of care, and shared decision-making, making it not only a reflection of service quality but also a driver of better health behaviors. As a result, patient satisfaction has become a central component of performance-based reimbursement programs, quality reporting initiatives, and broader value-based care reforms across the United States.

Despite this emphasis, satisfaction is not evenly distributed across populations. Studies have shown persistent disparities along demographic and geographic lines, including variations by age, race and ethnicity, socioeconomic status, and rural–urban location.^{6,8,10} These inequities highlight the limitations of relying solely on aggregated or hospital-level metrics, which can mask meaningful variations at the provider’s level. Most prior research has focused on institutional performance rather than individual clinician patterns, leaving critical gaps in understanding how structural and social determinants shape patient experiences. By adopting a provider-level analytical framework, this study seeks to uncover more granular dynamics, offering a clearer view of how demographic factors and social determinants influence patient satisfaction outcomes.

1.4.2. Payment models and chronic care quality

APMs have demonstrated measurable success in improving both quality of care and cost efficiency, particularly in the management of chronic conditions.¹¹⁻¹³ These models aim to shift incentives from volume-based to value-based care,

rewarding providers for improving outcomes rather than increasing service utilization. National programs under the Centers for Medicare & Medicaid Services (CMS) have shown that APMs can reduce hospital re-admissions, improve chronic disease management, and enhance care coordination across multiple settings. Their emphasis on accountability and performance measurement has made APMs central to present U.S. healthcare reform strategies, especially in primary and preventive care delivery.

However, while APMs hold significant promise, emerging evidence indicates they may inadvertently widen existing health disparities if equity is not explicitly integrated into their design.^{14,15} National evaluations—such as those conducted by the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation and JAMA Health Forum—have highlighted broad trends but provide limited insight into the experiences of specific demographic subgroups. This lack of granularity obscures how payment reforms may differentially impact vulnerable populations, including racial and ethnic minorities, rural residents, and low-income patients. By linking payment structures to patient-level demographic characteristics, this study advances an equity-sensitive performance evaluation framework, offering a more precise understanding of how APMs influence outcomes and disparities.

1.4.3. Provider performance and workforce shifts

The expansion of nurse practitioners (NPs) and physician assistants (PAs) represents a major structural shift in the

organization and delivery of healthcare services within CMS programs.¹⁶⁻²¹ Over the past decade, NPs and PAs have assumed increasingly central roles in primary and specialty care, particularly in rural and underserved communities where physician shortages are most acute.²¹ This trend reflects a strategic response to growing demand for accessible, cost-effective, high-quality care. NPs and PAs are often at the frontline of patient interaction, offering essential services, such as preventive screenings, chronic disease management, and patient education. Their involvement is not only expanding the capacity of healthcare systems but also redefining care models to be more team-based and patient-centered, with potential implications for access and quality.

Despite their growing significance, most existing studies on NP and PA integration have concentrated on workforce supply, billing patterns, and regulatory considerations, with far less attention paid to their impact on patient-centered outcomes, such as satisfaction, experience, or perceived quality of care. This narrow focus limits our understanding of how expanded NP/PA roles influence healthcare performance beyond operational efficiency. To address this gap, our study integrates workforce distribution data with outcome measures, examining how the increasing presence of NPs and PAs affects patient satisfaction, trust, and cost efficiency. By linking workforce structure to patient outcomes, this research provides a more comprehensive assessment of how care team composition shapes the delivery and perception of healthcare services.¹⁶⁻²¹

1.4.4. Equity, disparities, and payment adjustments

Patient satisfaction is widely recognized as a critical performance indicator in healthcare systems worldwide, closely linked to both the quality and effectiveness of care delivery.^{1,5,7} Extensive research has shown that patient-centered care improves clinical outcomes, fosters greater trust between patients and providers, and enhances adherence to treatment plans.^{1,5,7} Beyond clinical implications, patient satisfaction influences institutional reputation, reimbursement models, and policy decisions, making it an essential measure of healthcare system performance. However, existing evidence highlights that satisfaction is not uniformly experienced. Disparities persist across demographic groups, socioeconomic strata, and geographic regions, with underserved populations often reporting lower satisfaction levels due to factors, such as limited access to resources, language barriers, and differences in cultural expectations.^{6,8,10} These disparities underscore the importance of moving beyond surface-level performance metrics to address structural inequities in healthcare delivery.

Most existing studies on patient satisfaction have primarily relied on aggregated or hospital-level metrics, which, while helpful in identifying broad trends, may obscure significant variations at the individual provider level. Such aggregation can mask disparities related to specific care practices, patient-provider interactions, and demographic influences. As a result, many health systems lack the granularity needed to implement targeted interventions that improve patient experience. This study addresses that critical gap by adopting a provider-level analytical approach to examine how demographic and social determinants influence satisfaction outcomes. By capturing this finer level of detail, the research enables a more precise understanding of how patient experiences differ across subpopulations and care settings, thereby supporting the design of more equitable and patient-centered healthcare strategies.^{6,8,10}

1.4.5. Systemic payment variations and policy reforms

Substantial regional and diagnostic variation in Medicare payment persists across the United States, an issue extensively documented in the literature.²⁸⁻³³ This variation, which remains significant even after standardizing for differences in patient health status and regional input costs, has long been a focus for health service researchers and policymakers.^{29,30} While policy analyses consistently highlight structural inefficiencies in the traditional fee-for-service system and note that greater spending often fails to correlate with improved health outcomes or quality,³⁴⁻³⁹ these analyses are rarely linked directly to the most critical measures of healthcare delivery: Patient experience data or granular demographic characteristics.^{34,36} The gap is significant, as high-intensity markets may reflect uncoordinated delivery systems that negatively impact patient-provider interactions. Yet, the association between spending intensity and patient satisfaction remains complex and under-analyzed concerning demographics, such as race, income, and rural status.³⁶⁻⁴¹

This study expands the present literature by linking these established systemic payment variations with provider performance and satisfaction metrics, as captured by patient experience surveys and quality data. This integrated approach, which considers payment, quality, and patient-reported outcomes across demographic subgroups, will offer novel insights crucial for future value-based policy reforms to mitigate waste, ensure health equity, and align payment incentives with high-quality, patient-centered care.

1.5. Conceptual framework and contribution

Research shows some critical trends in healthcare. APMs and VBP programs can enhance the quality of care.

However, these programs might unintentionally increase existing inequities if not adjusted properly. Based on our observations, growing participation from NPs and PAs is reshaping care delivery, especially within rural and underserved communities. While patient satisfaction is a crucial indicator of provider performance, it often gets overlooked. Many studies have not linked payment differences to patient experiences. This study aims to fill that gap by using Medicare data to examine satisfaction and performance in different regions while considering payment models, patient demographics, chronic conditions, and social factors. The findings obtained could provide a valuable basis for guiding payment reforms and improving care.

2. Data and methods

2.1. Data source and selection

This study utilized the CMS Medicare Physician & Other Practitioners dataset (2022–2024), which contains ~1.2 million provider records. The dataset includes provider-level utilization metrics, submitted charges, Medicare-allowed amounts, actual payments, and rich beneficiary demographic and health information, enabling examination of geographic variation in patient risk scores and associated provider characteristics. Although individual patient treatment data is not available, the dataset supports robust statistical and machine learning analyses to identify patterns and disparities in care.

The CMS dataset was selected for its reliability, national coverage, provider-level granularity, and public accessibility, allowing a replicable research process. Key features include provider performance metrics, patient demographics, chronic condition prevalence, and service-level details (e.g., CPT codes) to evaluate efficiency, equity, and outcomes (Table 2).

2.2. Variables

The dependent variable is the average patient risk score (*Bene_Avg_Risk_Score*), reflecting overall health risk and complexity. Independent variables include provider type (*Rndrng_Prldr_Type*), total services (*Tot_Srvcs*), total Medicare payments (*Tot_Mdcr_Pymt_Amt*), hypertension prevalence (*Bene_CC_PH_Hypertension_V2_Pct*), and average patient age (*Bene_Avg_Age*). Engineered features for per-service payments were initially considered but excluded due to multicollinearity.

2.3. Data preprocessing

A structured pipeline was applied: Column standardization, removal of high-null variables (>40%), imputation of missing values (median for numeric, mode for categorical),

Table 2. Justification for the selection of key features from the CMS dataset

Feature	Justification
Provider performance	Detailed information on submitted charges, allowed amounts, and Medicare payments enables a financial and efficiency-based analysis of providers. This supports the evaluation of how provider behavior influences patient outcomes.
Patient demographics	The dataset includes provider location and service type information, which can be combined with external demographic data to analyze healthcare disparities and access issues.
Chronic condition prevalence	By including patient chronic condition data, the dataset allows for examining how specific diseases, such as hypertension, impact patient risk scores and how providers manage these conditions.
Impact of services	Including Current Procedural Terminology (CPT) codes allows for a detailed comparison of different services (e.g., drug therapy vs. medical management) and their relationship to patient outcomes and costs.

categorical encoding, outlier filtering (IQR method), and feature engineering (per-service financial metrics, age-bucket groups). The resulting dataset was clean, consistent, and analysis-ready (Tables 3 and 4). Continuous variables were normalized, and categorical variables were encoded to ensure comparability across providers and regions. The data were normalized before analysis to ensure comparability across providers and geographic regions. Specifically, continuous variables, such as patient risk scores and service volumes were standardized (z-score normalization) to remove scale differences, while categorical variables were encoded appropriately to support statistical and machine learning models. This step was essential to prevent variables with larger numerical ranges from disproportionately influencing the analysis and to improve the stability and interpretability of model estimates.

2.4. Assumption testing and analytical approach

- Normality: All continuous variables failed the Shapiro–Wilk test ($p < 0.05$), indicating non-Gaussian distributions.
- Multicollinearity: Correlation analysis and variance inflation factor (VIF) revealed mostly weak correlations (max $r = 0.34$), except for payment-related features with $VIF > 60$.
- Heteroscedasticity: The Breusch–Pagan test confirmed non-constant variance ($p=0.0$).

Traditional linear regression was deemed inappropriate given non-normality, heteroscedasticity, and multicollinearity. Mitigation strategies included: Reducing

Table 3. Sample data

Rndrng_NPI	Rndrng_Privr_ Last_Org_Name	Rndrng_Privr_ First_Name	Rndrng_Privr_MI	Rndrng_Privr_Crdntls	Rndrng_Privr_Gndr	Rndrng_Privr_Ent_Cd
1003000126	Erikesha	Ardalan	NaN	M.D.	M	I
1003000134	Cibull	Thomas	L	M.D.	M	I
1003000142	Khalil	Rashid	NaN	M.D.	M	I
1003000423	Veletta	Jennifer	A	M.D.	F	I
1003000480	Rothchild	Kevin	B	MD	M	I

Table 4. Descriptive statistics of the numerical features

Statistic	Rndrng_NPI	Rndrng_Privr_RUCA	Tot_HCPCS_Cds	Tot_Benes	Tot_Srv	Tot_Sbmtd_Chrg	Tot_Mdcr_Alowd_Amt	Tot_Mdcr_Pymt_Amt
Count	1.23029e+06	1.22816e+06	1.23029e+06	1.23029e+06	1.23029e+06	1.23029e+06	1.23029e+06	1.23029e+06
Mean	1.49874e+09	1.60287e+01	2.27418e+01	3.08250e+02	2.59328e+03	3.53388e+05	1.10274e+05	8.85944e+04
Std	2.87888e+08	3.34062e+02	2.39965e+01	4.49148e+02	3.92407e+06	6.97740e+06	6.47042e+06	6.84172e+06
Min	1.00030e+09	1.00000e+00	1.00000e+00	1.00000e+00	1.00000e+00	2.50000e+01	2.50000e+01	1.00000e+00
25%	1.23967e+09	1.00000e+00	1.00000e+01	1.00000e+01	5.50000e+01	1.34022e+04	3.63097e+03	2.81765e+03
50% (Median)	1.49759e+09	1.00000e+01	1.00000e+01	4.50000e+01	2.35000e+02	3.63699e+04	1.34949e+04	1.02047e+04
75%	1.75901e+09	1.90000e+01	2.00000e+01	1.14000e+02	6.45000e+02	1.04890e+05	3.98853e+04	3.01319e+04
max	1.96300e+09	9.90000e+01	8.27000e+02	1.64926e+05	2.06104e+07	1.16326e+09	2.67647e+08	2.67647e+08

highly correlated features, applying robust standard errors, and using non-parametric approaches. A random forest regressor was chosen for prediction due to its robustness in dealing with these data complexities. At the same time, K-means clustering was employed for exploratory analysis to identify natural provider groupings and hidden patterns.

2.5. Exploratory data analysis

Correlation analysis highlighted strong relationships between total payments and standardized payment metrics, moderate correlations with service and demographic variables, and weak correlations with chronic condition prevalence. Scatter analysis of payments versus average risk score showed weak associations, with older age groups exhibiting higher payments at similar risk scores. This methodology ensures a rigorous, interpretable framework for assessing the influence of provider characteristics and patient demographics on health outcomes at the national level (Figure 1).

3. Results

We employed several different modeling methods to gain a clear understanding of healthcare provider operations, patient demographics, and care outcomes. By combining statistical analysis with machine learning and predictive models, we ensured our findings were accurate and accessible. Statistical models allowed us to

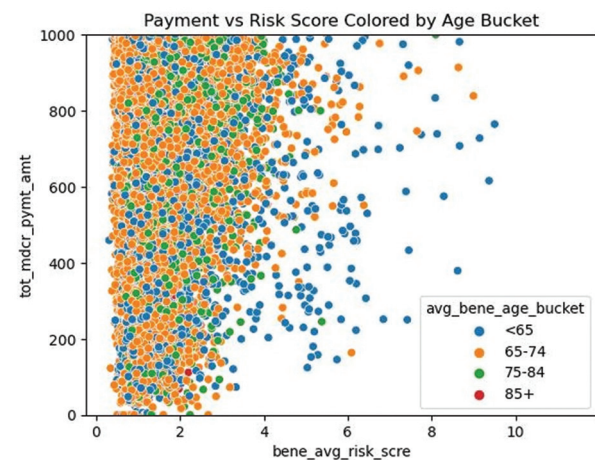


Figure 1. Scatterplot depicting payment vs. risk score by age bucket

test our hypotheses and pinpoint the key factors affecting performance, while machine learning was necessary to detect complex, non-linear patterns in large and multidimensional healthcare datasets that traditional statistical methods may not fully capture. In this study, the dataset from the CMS includes extensive provider-level and geographic information, making manual or purely descriptive analysis insufficient for uncovering subtle relationships between patient risk scores, provider characteristics, and disparities in care. Predictive models

enabled us to forecast outcomes and evaluate potential solutions, making our insights practical for informed decision-making. These approaches enhanced our analysis and reduced bias, enabling us to better understand the differences among patient groups and provider practices.

3.1. Random forest regressor

The random forest regressor was chosen for its flexibility and robustness in modeling complex healthcare data. Its ability to capture non-linear relationships between predictors and outcomes made it particularly well-suited for analyzing patient satisfaction, which can be influenced by multiple interacting factors. In addition, random forests can effectively handle missing values and maintain predictive accuracy without extensive data imputation. They also provide feature importance rankings and insight into which variables most strongly influence outcomes.

In this study, the random forest model was used to predict patient satisfaction score based on a set of provider financial and service-related metrics, including total Medicare payment (tot_mdcr_pymt_amt), total allowed amounts (tot_mdcr_alowd_amt), total services provided (tot_srvc), average payment per service (avg_payment_per_service), beneficiary average risk score (bene_avg_risk_scre), and average beneficiary age (bene_avg_age). A baseline linear regression model was also implemented to provide a straightforward, interpretable comparison, highlighting general trends and linear associations. While the linear model offered clarity in understanding simple relationships, the random forest model excelled at capturing complex interactions and non-linear effects among predictors that might influence patient satisfaction.

Model performance was rigorously evaluated using R^2 , mean squared error (MSE), and root mean squared error (RMSE), providing a quantitative assessment of predictive accuracy and model fit. Feature importance analysis revealed that beneficiary risk score and service volume were the primary drivers of patient satisfaction. This indicates that patients with higher risk profiles and providers with larger service volumes tend to influence satisfaction outcomes more strongly. This combination of interpretability and predictive power ensured the modeling approach could support actionable insights for improving provider performance and patient experiences.

3.2. K-Means clustering

K-Means clustering was employed to segment Medicare providers into distinct performance groups: High-, medium-, and low-performing clusters. The clustering analysis was based on key provider and patient metrics, including total services provided (tot_srvc), total

beneficiaries served (tot_benes), total Medicare payment amounts (tot_mdcr_pymt_amt), and age-bucketed patient demographics. All features were standardized before analysis to ensure that variables with different scales did not disproportionately influence clustering. The optimal number of clusters was determined using a combination of the Elbow Method, which identifies the point where adding more clusters minimally reduces within-cluster variance, and domain expertise, ensuring practical interpretability and relevance to healthcare policy.

The resulting clusters revealed distinct performance patterns among providers. High-performing providers were characterized by managing larger and more complex patient populations, higher total Medicare payments, and a broader range of services. In contrast, medium- and low-performing clusters had comparatively smaller patient loads and lower financial throughput. These patterns highlighted operational efficiency and differences in patient demographics and provider capacity. In addition, state-level distributions of the clusters were visualized in Tableau, facilitating regional comparisons and informing policy recommendations for resource allocation, provider support, and targeted interventions to improve care quality across different geographic areas.

3.3. Research question analyses

To address our research questions, we evaluated provider performance, demographic disparities, chronic disease patterns, and differences in treatment-related reimbursements. The key findings are summarized as follows:

- (i) Provider performance: Linear regression predicted total Medicare payments ($R^2 = 0.90$) using services, beneficiaries, risk scores, and age. Random forest linked high-risk patient management to elevated satisfaction, though some high-payment providers showed inconsistent outcomes. K-Means identified three performance tiers.
- (ii) Demographics and regional disparities: Clustering exposed geographic variation, with rural regions serving older populations and higher chronic disease burdens, highlighting persistent access inequities.
- (iii) Chronic condition prevalence: A random forest classifier predicted high diabetes prevalence (accuracy 91%, precision 78%, recall 73%), identifying concentrated disease burdens in older, lower-income populations and revealing variability in provider management strategies.
- (iv) Drug versus medical services: Analysis of variance (ANOVA) compared reimbursements for drug-based versus medical interventions, showing higher payments for medical services ($p < 0.001$). Combining

pharmacological and procedural treatments yielded superior outcomes, suggesting reimbursement alignment with demonstrated health benefits.

This integrated approach demonstrates that provider efficiency, patient demographics, and chronic disease prevalence jointly shape Medicare payment patterns and patient satisfaction. Regression, clustering, classification, and inferential analyses provide a robust framework for data-driven policy and resource allocation decisions (Table 5).

4. Discussion

This study examines how we can use open data from Medicare, machine learning through Python, and Tableau to visualize information about healthcare providers and patients. The aim is to understand better chronic disease management and the various factors that affect healthcare delivery. The research demonstrates how public healthcare data can generate valuable insights to help guide decisions and influence policies in the U.S. healthcare landscape by employing techniques, such as predictive modeling, clustering, and statistical analysis.

Here are some key takeaways from the study:

- (i) Provider performance and equity: The study identifies the good-performing healthcare providers and sheds light on the demographic and operational aspects that shape how healthcare services are reimbursed and utilized. By recognizing these patterns, decision-makers can pinpoint gaps in healthcare access and work toward reducing inequalities, whether based on age, gender, income, or geography. This approach aims to promote fair access, improve patient health outcomes, and make healthcare delivery more efficient.
- (ii) Chronic disease management: The research reveals ways to balance costs while enhancing patient outcomes, especially when analyzing medication and direct medical interventions. This can assist healthcare providers and insurers in allocating resources effectively, tailoring treatment plans, and implementing preventive care programs. These changes are vital for minimizing hospital visits and improving the quality of life for patients with chronic conditions.
- (iii) Advanced analytics in healthcare: The findings highlight a shift in the healthcare industry from

a reactive approach to a more proactive one. By utilizing open data, healthcare organizations can enhance transparency, hold themselves accountable, and monitor performance across different providers and regions. Advanced analytics help anticipate future challenges and manage population health effectively.

- (iv) Tailored insights: Importantly, the study emphasizes providing customizable data-driven recommendations rather than one-size-fits-all solutions. Policymakers, payers, and providers can adapt these insights to their community needs, making informed choices that align with their goals and resources. By integrating open data with advanced analytics, organizations can better prioritize initiatives, improve service delivery, and ensure equitable access for diverse patient groups.

Future work will explore promising directions, including the use of long-term datasets, the integration of electronic health records with claims data, and the incorporation of patient feedback. This would provide a more comprehensive view of healthcare trends and enable more precise strategies for enhancing efficiency and patient-focused care within Medicare and the broader healthcare system.

5. Conclusion

This study explores the use of business intelligence and machine learning to clarify Medicare policies, evaluates healthcare provider performance, and inform strategies for improving population health. By connecting data from the CMS with advanced analytical techniques, such as random forest regression for predictions, K-means for grouping data, and ANOVA for comparing results, the research uncovers trends in provider performance, patient demographics, the prevalence of chronic illnesses, and treatment costs. These insights suggest that using advanced analytics can lead to informed discussions on payment structures, identify service gaps, and address health equity challenges, ultimately helping manage costs while enhancing patient care.

In terms of practical contributions, this study adds to healthcare analytics by showcasing how we can use Medicare data at the provider level to gain insights relevant to operations and policies. By incorporating methods, such as random forest regression, K-means clustering, and ANOVA, it presents a framework that can be used to evaluate patient risk scores, provider performance, and trends related to location and demographics. The findings reveal connections between the amount of services provided, Medicare payments, patient complexity, and health outcomes, providing evidence that might help improve value-based payment models. In addition, by

Table 5. ANOVA results

Source	sum_sq	df	F	PR(>F)
C (service_type)	5.155389e+13	1.0	184,561.66	0.0
Residual	8.400281e+13	300,728.0	NaN	NaN

identifying disparities across populations and locations, the study identifies areas that could benefit from focused interventions. Preliminary comparisons of different types of services also help gauge how effective the care is. Overall, this work merges analytical methods with real-world applications, showing how organized Medicare data can influence healthcare delivery, policymaking, and discussions around equity.

However, there are some limitations to consider. The analysis is based on only 1 year of Medicare data, which makes it difficult to track trends or the impact of policies over time. The dataset primarily relied on structured claims and administrative information, missing out on unstructured clinical data, such as electronic health record narratives or imaging results. Quality was inferred from financial and service metrics, which might not entirely reflect how well care is delivered or how patients feel about their experience. Due to high correlations among some economic variables, the study had to limit the number of predictors used, and the nature of predictive modeling here does not allow us to draw definitive conclusions about causal relationships among services, payments, and outcomes.

Looking ahead, future research could build on this work by analyzing multiple years of Medicare data to track changes in provider performance and the effects of policies over time. Integrating more diverse data sources, such as electronic health records, clinical registries, and patient-reported outcomes, could provide a richer understanding of clinical quality and patient experiences. Causal modeling strategies, such as difference-in-differences or propensity score matching, could also offer more substantial evidence of how different interventions affect results. In addition, conducting deeper analyses of racial, socioeconomic, and geographic disparities could help shape policies focused on equity. Finally, developing interactive decision-support tools that use predictive analytics could enable more proactive monitoring of provider performance and patient outcomes.

Acknowledgments

None.

Funding

None.

Conflict of interest

The author declares no conflict of interest.

Author contributions

This is a single-authored article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

This study utilized the CMS Medicare Physician & Other Practitioners dataset (2022–2024), which contains ~1.2 million provider records. The dataset includes provider-level utilization metrics, submitted charges, Medicare-allowed amounts, actual payments, and rich beneficiary demographic and health information, enabling examination of geographic variation in patient risk scores and associated provider characteristics. Although individual patient treatment data is not available, the dataset supports robust statistical and machine learning analyses to identify patterns and disparities in care. The CMS dataset was selected for its reliability, national coverage, provider-level granularity, and public accessibility, allowing a replicable research process. Key features include provider performance metrics, patient demographics, chronic condition prevalence, and service-level details (e.g., CPT codes) to evaluate efficiency, equity, and outcomes.

References

1. Peral PO, Rambaud SC, García JS. Quality of care and patient satisfaction: Future trends and economic implications for the healthcare system. *J Econ Surveys*. 2024;38(1):3–28.
doi: 10.1111/joes.12657
2. Simmons C, Pot M, Lorenz-Dant K, Leichsenring K. Disentangling the impact of alternative payment models and associated service delivery models on quality of chronic care: A scoping review. *Health Policy*. 2024;143:105034.
doi: 10.1016/j.healthpol.2024.105034
3. Kim H, Mahmood A, Hammarlund NE, Chang CF. Hospital value-based payment programs and disparity in the United States: A review of current evidence and future perspectives. *Front Public Health*. 2022;10:882715.
doi: 10.3389/fpubh.2022.882715
4. Mokhtary S, Janati A, Yousefi M, Raei B, Moradi F. Evidence on the effectiveness of value-based payment schemes implemented in a hospital setting: A systematic review. *J Educ Health Promot*. 2024;13:327.
doi: 10.4103/jehp.jehp_873_23
5. Cui J, Du J, Zhang N, Liang Z. National patient satisfaction survey as a predictor for quality of care and quality improvement - experience and practice. *Patient Prefer Adherence*. 2025;19:193–206.

- doi: 10.2147/PPA.S496684
6. Rahman MM, Darwin C, Amin MI, Sujana ST. Identifying key influencers of patient satisfaction using an explainable machine learning approach. *Sci Rep*. 2025;15:1123.
doi: 10.1038/s41598-025-18809-x
7. Alshahrani AM. Predictors of patients' satisfaction with primary health care services in the Kingdom of Saudi Arabia: A systematic review. *Healthcare*. 2023;11:2973.
doi: 10.3390/healthcare11222973
8. Sebera E, Hagenimana C, wagirumukiza T. Patient satisfaction survey in a public hospital: Remera Rukoma District Hospital, Rwanda, 2023. *BMC Health Serv Res*. 2024;24:11996.
doi: 10.1186/s12913-024-11996-9
9. Shaki D, Aimbetova G, Baysugurova V, *et al*. Level of patient satisfaction with quality of primary healthcare in almaty during COVID-19 pandemic. *Int J Environ Res Public Health*. 2025;22(5):804.
doi: 10.3390/ijerph22050804
10. Howard SW, Bradford N, Belue R, *et al*. Building alternative payment models in health care. *Front Health Serv*. 2024;4:1235913.
doi: 10.3389/frhs.2024.1235913
11. Office of the Assistant Secretary for Planning and Evaluation (ASPE). *The Impact of Alternative Payment Models on Medicare Spending and Quality, 2012-2022*. U.S. Department of Health and Human Services; 2025. Available from: <https://aspe.hhs.gov/reports/impact-alternative-payment-models-medicare-spending-quality-2012-2022> [Last accessed on 2025 Oct 20].
12. Edmiston K. *Alternative payment models, value-based payments, and health disparities*. United States: National Association of Insurance Commissioners; 2022.
doi: 10.52227/25526.2022
13. Pozniak A, Lammers E, Mukhopadhyay P, *et al*. Association of the home health value-based purchasing model with quality, utilization, and medicare payments after the first 5 years. *JAMA Health Forum*. 2022;3(8):e222723.
doi: 10.1001/jamahealthforum.2022.2723
14. Crowther VB, Shaw-Johnson C, Green-Weir RR, Woodley PS, Weatherspoon MA. Unveiling patient satisfaction: Exploring demographics, interactions, and health-seeking behavior. *Patient Exp J*. 2025;2(1):105-114.
doi: 10.35680/2372-0247.1950
15. Patel SY, Auerbach D, Huskamp HA, *et al*. Provision of evaluation and management visits by nurse practitioners and physician assistants in the USA from 2013 to 2019: Cross-sectional time series study. *BMJ*. 2023;382:e073933.
doi: 10.1136/bmj-2022-073933
16. Patel SY, Huskamp HA, Frakt AB, *et al*. Frequency of indirect billing to medicare for nurse practitioner and physician assistant office visits. *Health Aff (Millwood)*. 2022;41(5):805-813.
doi: 10.1377/hlthaff.2021.01968
17. Morgan PA, Smith VA, Berkowitz TSZ, *et al*. Impact of physicians, nurse practitioners, and physician assistants on utilization and costs for complex patients. *Health Aff (Millwood)*. 2019;38(3):1028-1036.
doi: 10.1377/hlthaff.2019.00014
18. Forte G. *How Improved Health Workforce Projection Models Could Support Policy*. Association of American Medical Colleges; 2024. Available from: <https://www.aamc.org/data-reports/workforce/how-improved-health-workforce-projection-models-could-support-policy> [Last accessed on 2025 Oct 20].
19. Bischof A, Greenberg SA. Post-COVID-19 reimbursement parity for nurse practitioners. *OJIN Online J Issues Nurs*. 2021;26(2):3.
doi: 10.3912/OJIN.Vol26No02Man03
20. American Medical Association. *AMA Issue Brief: Scope-of-Practice*. AMA; 2024. Available from: <https://www.ama-assn.org/system/files/issue-brief-scope-practice-summary-existing-studies.pdf> [Last accessed on 2025 Oct 20].
21. American Medical Association. *Medicare Alternative Payment Models*. AMA; 2024. Available from: <https://www.ama-assn.org/practice-management/payment-delivery-models/medicare-alternative-payment-models> [Last accessed on 2025 Oct 20].
22. Centers for Medicare & Medicaid Services. *Hospital Value-Based Purchasing (HVBP) Program Data*. CMS; 2024. Available from: <https://www.cms.gov/medicare/quality/initiatives/hospital-quality-initiative/hospital-value-based-purchasing> [Last accessed on 2025 Oct 20].
23. Centers for Medicare & Medicaid Services. *CMS Hospital Value-Based Purchasing Program Results for Fiscal Year 2020*. CMS; 2019. Available from: <https://www.cms.gov/newsroom/fact-sheets/cms-hospital-value-based-purchasing-program-results-fiscal-year-2020> [Last accessed on 2025 Oct 20].
24. Medicare Payment Advisory Commission. *Report to Congress: Medicare and the Health Care Delivery System*. MedPAC; 2022. Available from: https://www.medpac.gov/wp-content/uploads/2025/06/Jun25_MedPAC_Report_To_Congress_SEC.pdf [Last accessed on 2025 Oct 20].
25. Brown B. *Value-Based Purchasing 2020: A 10-Year Progress Report*. Health Catalyst; 2020. Available from: <https://www.healthcatalyst.com/learn/insights/value-based-purchasing-2020-10-year-progress-report> [Last accessed on 2025 Oct 20].

- 20].
26. McDonough S. *Medicare Value-Based Purchasing Program Overview*; National Hospital Association/MHA; 2023. Available from: <https://www.mha.org/wp-content/uploads/2023/06/Medicare-VBP-June-7-2023.pdf> [Last accessed on 2025 Oct 20].
27. Tollera G, Retta MT, Girmay AM. Patient satisfaction with healthcare services and associated factors at public health centers in Addis Ababa, Ethiopia. *Discov Soc Sci Health*. 2025;5:38.
doi: 10.1007/s44155-025-00185-6
28. Norton EC, Lawton EJ, Li J. Moneyball in medicare. *Am J Health Econ*. 2023;9(1):96-126.
doi: 10.1086/721707
29. Chen HC, Cates T, Taylor M. The effect of patient quality measurements and HCAHPS patient satisfaction on hospital reimbursements. *Health Serv Manage Res*. 2023;42(4):419-433.
doi: 10.3233/hsm-220042
30. Banerjee S, McCormick D, Paasche-Orlow MK, Lin M-Y, Hanchate AD. Association between degree of exposure to the hospital value-based purchasing program and 30-day mortality: Experience from the first four years of Medicare's pay-for-performance program. *BMC Health Serv Res*. 2019;19:921.
doi: 10.1186/s12913-019-4562-7
31. Brooks Carthon KJ, Muir K, Iroegbu C, *et al*. COVID-19 mortality disparities among socially vulnerable Medicare beneficiaries associated with the quality of nurse work environments in U.S. Hospitals. *Inquiry*. 2024;61.
doi: 10.1177/00469580241284959
32. Cao YJ, Chen D, Liu Y, Smith M. Disparities in the use of in-person and telehealth primary care among high- and low-risk medicare beneficiaries during COVID-19. *J Patient Exp*. 2021;8.
doi: 10.1177/23743735211065274
33. Thornton M, Hamelburg L, Nonnemaker S, Abdukadirov S, Veselovskiy G, Siegel S. Care quality metrics in medicare during COVID-19 pandemic. *Am J Manag Care*. 2025;31.
doi: 10.37765/ajmc.2025.89781
34. Centers for Medicare & Medicaid Services. *Medicare Physician and Other Practitioners - by Provider*. CMS Open Data Portal; 2023. Available from: <https://data.cms.gov/provider-summary-by-type-of-service/medicare-physician-other-practitioners/medicare-physician-other-practitioners-by-provider> [Last accessed on 2025 Oct 20].
35. Lin SC, Adler-Milstein J, Hollingsworth JM, Ryan A. Alternative Payment Models and Patient-Reported Quality of Preparation for Discharge: A Retrospective Longitudinal Study. *J Patient Exp*. 2024;11.
doi: 10.1177/23743735241240926
36. Lee SJ, Venkataraman S, Heim GR, Roth AV, Chilingirian J. Impact of the value-based purchasing program on hospital operations outcomes: An econometric analysis. *J Oper Manag*. 2019;66(1-2):151-175.
doi: 10.1002/joom.1057
37. McMurtry CL, Findling MG, Casey LS, *et al*. Discrimination in the United States: Experiences of Asian Americans. *Health Serv Res*. 2019;54(2):1419-1430.
doi: 10.1111/1475-6773.13225
38. Bailey V. *Medicare Advantage Member Satisfaction Increases; Mental Health Coverage Lags*. TechTarget; 2022. Available from: <https://www.techtarget.com/healthcarepayers/news/366603926/Medicare-Advantage-Member-Satisfaction-Soars-Mental-Health-Coverage-Lags> [Last accessed on 2025 Oct 20].
39. Cohen KR, Vabson B, Podulka J, *et al*. Medicare Risk Arrangement and Use and Outcomes Among Physician Groups. *JAMA Netw Open*. 2025;8(1):e2456074.
doi: 10.1001/jamanetworkopen.2024.56074
40. Ren H, Shireman E, VanGompel EW, *et al*. Preconception, interconception, and reproductive health screening tools: A systematic review. *Health Serv Res*. 2022;58(2):458-488.
doi: 10.1111/1475-6773.14123
41. Seid S, Aychiluhm B, Mohammed AA. Effectiveness and feasibility of telerehabilitation in patients with COVID-19: A systematic review and meta-analysis. *BMJ Open*. 2022;12(10):e063961.
doi: 10.1136/bmjopen-2022-063961