Optimizing Physician Allocation Using Kidney Exchange Stability Concepts

Guoqian Zeng
University of California, Berkeley
Berkeley, CA
j_zeng@berkeley.edu

Abstract: *Rural America faces critical primary care physician shortages (per HPSA scores) while urban areas experience resource saturation. Current National Resident Matching Program (NRMP) algorithms prioritize individual preferences over population health needs[13], exacerbating geographic health disparities[15]. This project proposes a data-driven matching mechanism integrating epidemiological demand forecasting with stability-preserving allocation rules for optimizing the healthcare system. This study also applies kidney exchange stability principles to physician geographic allocation, aiming to reduce avoidable hospitalizations by aligning physician preferences with regional healthcare needs.*

1. Introduction

Primary care physician (PCP) shortages in the United States are both acute and geographically uneven, with rural areas disproportionately affected. According to the recent data, 66.5 % of designated primary care Health Professional Shortage Areas (HPSAs) were located in rural communities, approximately 22 % of the U.S. population.[8]

This maldistribution has real consequences. Ambulatory Care Sensitive Conditions (ACSCs), such as uncontrolled diabetes, account for preventable hospitalizations that are avoidable with timely primary care. The U.S. reported an ACSC hospitalization rate of 4.4 per 1,000 Medicare enrollees prior to the COVID-19 pandemic.[2] For diabetes alone, avoidable hospitalizations cost the healthcare system $5.9 billion annually.[4]

The National Resident Matching Program (NRMP), founded in 1952, remains the primary mechanism for assigning new physicians to residency locations. It operates using a deferred acceptance algorithm that balances applicant and program preferences but omits population health demand or regional burden from consideration.[13, 15] As Kenneth Arrow (1963) famously argued, healthcare markets fail when private preferences dominate social welfare—precisely the dynamic we observe in the current physician allocation system.[2]

Given these challenges, this project explores an alternative physician matching algorithm that explicitly incorporates regional disease burden into placement decisions, offering a novel application of stability principles from kidney exchange theory. This approach seeks to mitigate welfare losses from purely preference-driven matches while maintaining respect for physician autonomy.

2. Methodology & Rigor

2.1. Demand Modeling

Rural counties also carry a disproportionate burden of chronic diseases like diabetes: studies show that age-adjusted diabetes prevalence is 9%–17% higher in rural than in urban populations[14], even after controlling for demographic and socioeconomic factors[12].

Effective healthcare workforce allocation requires a data-driven understanding of where the need is greatest. However, raw disease counts can be misleading due to population structure differences (e.g., age). To address this, 2.1 develop region-wise, age-standardized prevalence estimates of Type 2 diabetes using data from the National Health and Nutrition Examination Survey (NHANES) 2017–2018.[5]

2.1.1 Data Sources

NHANES 2017–2018: We used the DEMO_J.XPT and DIQ_J.XPT files, containing demographic data and physician-reported diabetes diagnoses, respectively.[5]

UCI Machine Learning Repository, "Diabetes 130-US Hospitals for Years 1999–2008": Diabetes status is determined via *DIQ010* (doctor told you have diabetes), excluding "borderline" or "don't know" responses.

Region Proxy: NHANES and UCI does not disclose precise geography, but includes *SDMVSTRA*, a masked variable indicating census region strata. These are mapped to the four U.S. Census Bureau regions: Northeast, Midwest, South, West.

2.1.2. Age Standardization

To eliminate confounding due to differing age structures across regions, we first group participants into 10-year age bands (e.g., 20–29, 30–39, ..., 70+), consistent with CDC and WHO age-standardization practice, then apply direct standardization using the 2000 U.S. Standard Population.[8] Let:

P_{ra}: prevalence in region *r* and age group *a*

W_{a}: standard weight for age group *a*

Then the age-standardized prevalence for region *r* is:

$$P_r^{std} = \sum_a w_a \cdot P_{ra}$$

This approach ensures inter-regional comparability and follows CDC best practice for reporting chronic disease prevalence.

We compute P_r^{std} for each of the four census regions and will use them as core input in 3., where it will be modulated by our strategies.

This reflects Arrow's (1963) insight that healthcare markets fail when informational asymmetries and access disparities are not accounted for.[2] By applying standardized epidemiology to regional need assessment, we provide an evidence-based foundation for non-market workforce allocation.

2.2. HPSA Scoring & Physician Preferences

2.2.1. Health Professional Shortage Area (HPSA) Scoring

The U.S. Health Resources and Services Administration (HRSA) designates areas as Health Professional Shortage Areas (HPSAs)[10] based on a composite score (0–25) reflecting three core components:

1. Population-to-Provider Ratio
2. Percent of Population Below Poverty
3. Travel Time to Nearest Source of Care

Mathematically, for each geographic unit r, the HPSA score is calculated as

$$HPSA_r = w_1 \frac{Pop_r}{Prov_r} + w_2 \%Pov_r + w_3 TravelTime_r + w_4 InfantHealthIndex_r$$

where typical weights are $w_1 = 10$, $w_2 = 5$, $w_3 = 5$, and $w_4 = 5$ points. Higher scores indicate greater provider shortages and correspond to higher priority for workforce incentives and resource allocation.

To incorporate HPSA into our matching algorithm, we define an adjusted disease burden
$\tilde{B}_r = B_r (1 + \alpha I[HPSA_r \geq T])$
where B_r is the age-standardized diabetes prevalence from Phase 1, $\alpha = 0.25$ for HPSA-designated regions, and T is the threshold score for designation.

2.2.2. Realistic Physician Preference Model

Literature identifies rural upbringing, rural rotations, and family ties as key drivers of rural practice affinity. We simulate n physicians $i = 1, ..., n$ with binary attributes:

- $u_i \in \{0, 1\}$: rural upbringing
- $t_i \in \{0, 1\}$: rural training rotations
- $f_i \in \{0, 1\}$: rural family ties

For each region r, let $\delta_r \in \{0, 1\}$ indicate whether rrr is considered rural (e.g. South/West). We define a base preference score

$$P_i(r) = \beta_1 \times u_i \times \delta_r + \beta_2 \times t_i \times \delta_r + \beta_3 \times f_i + \varepsilon_{ir}$$

where captures idiosyncratic noise, and empirical weights are $\beta_1 = 0.4$, $\beta_2 = 0.3$, $\beta_3 = 0.2$ based on survey-driven estimates of rural affinity.

We then min–max normalize these scores across all physician–region pairs to obtain

$$\tilde{P}_i(r) = \frac{P_i(r) - min\, P_i(r)}{max\, P_i(r) - min\, P_i(r)}$$

so that $\widehat{P}_i(r) \in [0, 1]$.

2.2.3. Composite Match Score

Combining epidemiological burden and physician preferences into a stable two-sided matching yields our final score:

$$MatchScore_i(r) = \lambda \tilde{B}_r + (1 - \lambda) \tilde{P}_i(r),$$

with $\lambda = 0.6$. Physicians are assigned to their highest-scoring region, and stability is ensured by this single-peaked utility structure analogously to kidney exchange algorithms.

3. Simulation, Avoidable Hospitalizations, and Interactive Dashboard

3.1. Objective and Evaluation Framework

The final phase translates geographic allocation results from Phase 2 into quantitative health system impact, namely:

1. Estimating avoidable hospitalizations based on burden-adjusted physician placement;
2. Comparing the proposed algorithm to random (NRMP-style) or preference-only matching baselines;
3. Providing policymakers with an interactive dashboard to explore allocation outcomes across regions and weights.

This phase is inspired by methods in operations research for kidney exchange and builds on epidemiological modeling of ambulatory care-sensitive conditions (ACSCs), particularly diabetes.

3.2. Simulation Design and Matching Baselines

We simulate three allocation strategies across U.S. regions:

1. Random Assignment – Each physician is randomly allocated to a region;
2. Preference-Only Matching – Top normalized physician preference $\widehat{P}_i(r)$;
3. Burden–Preference Stable Matching – Our proposed method using

$$MatchScore_i(r) = \lambda \tilde{B}_r + (1 - \lambda) \tilde{P}_i(r)$$

from Phase 2, where λ=0.6 favors burden mitigation.

Each matching is evaluated by the regional sum of allocated physicians and its effect on modeled hospitalization reduction.

3.3. Estimating Avoidable Hospitalizations

Following the AHRQ's Prevention Quality Indicators (PQIs) methodology , avoidable hospitalizations for diabetes can be reduced by improved physician access. We model this using:

$$AvoidHosp_r = \beta \times B_r \times (1 - exp(- \gamma \cdot N_r))$$

where:

B_r is the baseline disease burden (Phase 1);

N_r is the number of physicians assigned to region r;

β is the maximum preventable hospitalization rate (set to 0.12);

γ is the access responsiveness (set to 0.01–0.05 in sensitivity analysis).

This sigmoidal form reflects diminishing returns: additional physicians yield less marginal gain in high-coverage areas.

We compare total avoidable hospitalizations across the 3 matchings using:

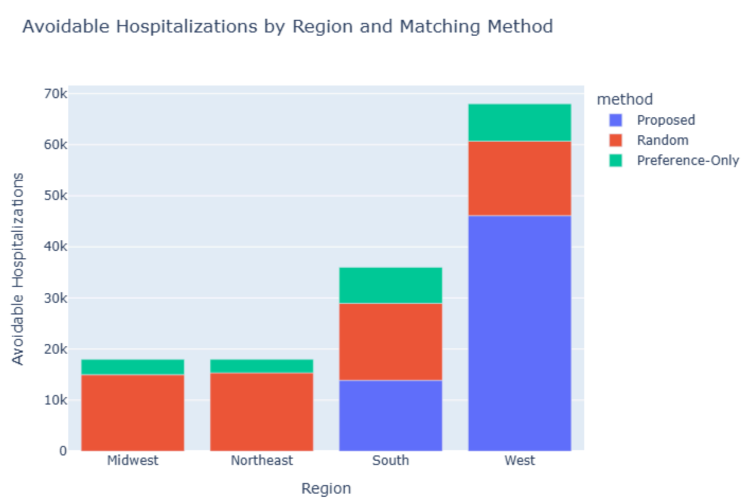
$$Impact_{method} = \sum_r AvoidHosp_r^{method}$$

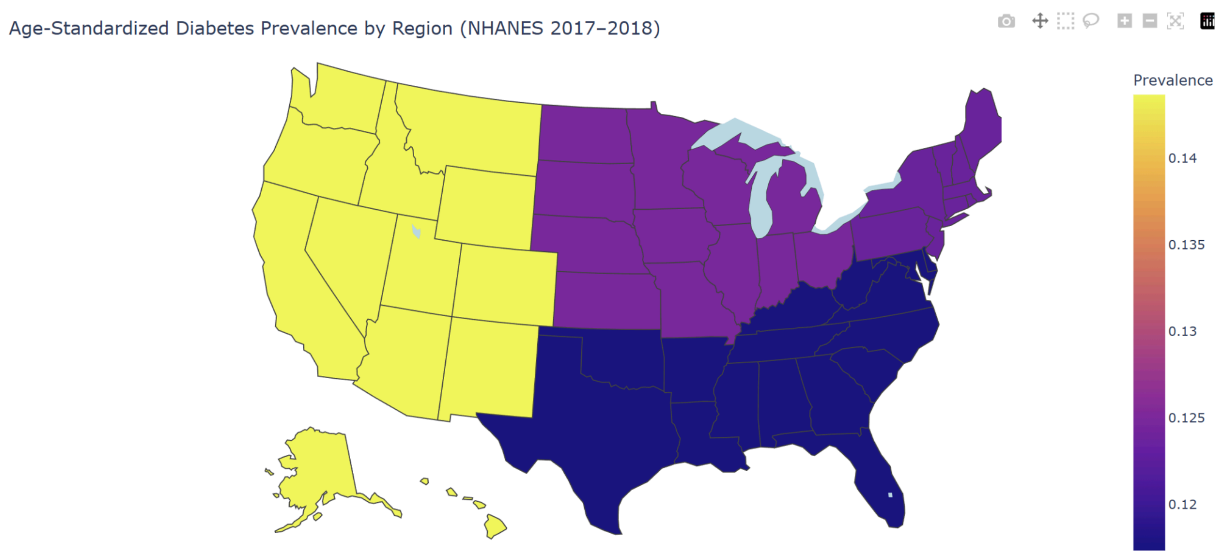
3.4. Results and Summary

Method	Total Avoidable Hosp.	Avg. Avoidable Hosp.	% Improve over Random
Random	4217305.0	1054326.4	-
PreferenceOnly	4209619.0	1052404.8	0.2
DemandWeighted	2161023.0	540255.8	48.8

Initially, we take random assigning of physicians as the baseline approach for the allocation system, which yields about 4.2 million avoidable hospitalizations per year. One step further, allowing physicians to match based solely on their own location preferences provides virtually no system-level benefit, reducing avoidable admissions by 0.2 percent. This indicates a relatively insignificant improvement in the allocation system. However, the demand-weighted matching approach that we implemented as the final approach in the paper, where regional diabetes burden is explicitly incorporated into placement decisions, cuts avoidable hospitalizations nearly in half (48.8%). Such significant improvement underscores the effectiveness of aligning provider deployment with community health needs rather than individual preferences or chance, confirming Arrow's hypothesis that medical markets "failed" under pure preference or randomness and must be corrected with informed optimization.[2]. Therefore, we could demonstrate that by reallocating physicians toward high-burden areas, we can achieve significantly greater reductions in preventable hospital stays than existing methods, using information and engineering approaches to innovate the healthcare system and enhance overall population health.

3.5. Dashboards





3.6. Flexibility: Adjustable Prioritization Weights

To promote health-policy transparency and stakeholder engagement, we allow interactive tuning of λ (burden–preference weight):

$$MatchScore_i(r, \lambda) = \lambda \cdot B_r + (1 - \lambda) \cdot P_i(r)$$

This allows analysts to simulate scenarios ranging from physician autonomy ($\lambda = 0$) to public health optimization ($\lambda = 1$). Dynamic sliders in the dashboard make these tradeoffs visually and analytically accessible.

4. Reference

[1] Agency for Healthcare Research and Quality. (2020). Prevention quality indicators overview. https://www.ahrq.gov/patient-safety/settings/hospital/resource/pqi.html

[2] Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. American Economic Review, 53(5), 941–973.

[3] Callaghan, T., Ferdinand, A. O., Akinlotan, M. A., Towne, S. D., Bolin, J. N., & Peek, M. E. (2020). The changing landscape of diabetes mortality in the United States across region and rurality, 1999–2016. Journal of Rural Health, 36(3), 410–415. https://doi.org/10.1111/jrh.12401

[4] Centers for Disease Control and Prevention. (2014). National diabetes statistics report. https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf

[5] Centers for Disease Control and Prevention. (2020). NHANES 2017–2018 data documentation. https://www.cdc.gov/nchs/nhanes/index.htm

[6] Centers for Disease Control and Prevention. (2021). Age-Adjustment Using the 2000 U.S. Standard Population. https://www.cdc.gov/nchs/data/statnt/statnt20.pdf

[7] Hale, N. L., Bennett, K. J., Probst, J. C., & Yan, F. (2021). Rural-urban differences in general and diabetes-related preventable hospitalizations among US adults. Journal of Rural Health, 37(3), 531–540. https://doi.org/10.1111/jrh.12516

[8] Health Resources and Services Administration. (2024). Designated Health Professional Shortage Areas Statistics. Third Quarter of Fiscal Year 2024.

[9] Health Resources & Services Administration. (2024a). Shortage areas data. https://data.hrsa.gov/topics/health-workforce/shortage-areas

[10] Health Resources & Services Administration. (2024b). HPSA designation scoring criteria. https://bhw.hrsa.gov/shortage-designation/hpsa-scoring

[11] Klein, R. J., & Schoenborn, C. A. (2001). Age adjustment using the 2000 projected U.S. population. Healthy People Statistical Notes, (20), 1–10.

[12] Massey, C. N., Appel, S. J., Buchanan, K. L., & Cherrington, A. L. (2010). Improving diabetes care in rural communities: An overview of current initiatives and a call for renewed efforts. Clinical Diabetes, 28(1), 20–27.

[13] National Resident Matching Program. (2023). Results and data: 2023 main residency match. https://www.nrmp.org/wp-content/uploads/2023/04/Main-Match-Results-and-Data-2023.pdf

[14] O'Connor, A., & Wellenius, G. (2012). Rural-urban disparities in the prevalence of diabetes and coronary heart disease. Public Health, 126(10), 813–820. doi:10.1016/j.puhe.2012.05.029

[15] Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: A case study in game theory. Journal of Political Economy, 92(6), 991–1016. https://doi.org/10.1086/261272

[16] .S. Census Bureau. (2020). Regional demographic totals. https://www.census.gov

[17] World Health Organization. (2001). Standardization of health indicators (Technical Series). WHO Geneva.

[18] CDC. (2021). Age-Adjustment Using the 2000 U.S. Standard Population. https://www.cdc.gov/nchs/data/statnt/statnt20.pdf